

# NEOPHILOLOGICA

35



# NEOPHILOLOGICA

*volume 35*

*Syntaxe, sémantique, lexique :  
Hommage à Gaston Gross  
(1939–2022)*

sous la rédaction de / edited by  
**Wiesław Banyś et Beata Śmigielska**

Rédacteur en chef / Editor-in-Chief: **WIESŁAW BANYŚ** (Université de Silésie, Pologne)  
Rédacteur en chef adjoint / Deputy Editor-in-Chief: **BEATA ŚMIGIELSKA** (Université de Silésie, Pologne)

COMITÉ SCIENTIFIQUE / EDITORIAL BOARD

<b>Denis APOTHÉLOZ</b>	Université Nancy 2, France
<b>Laura CALABRESE</b>	Université Libre de Bruxelles, Belgique
<b>Jean-Pierre DESCLÉS</b>	Université Paris-Sorbonne, France
<b>Francis GROSSMANN</b>	Université Grenoble Alpes, France
<b>Zlatica GUENTCHÉVA</b>	CNRS, Paris, France
<b>Anna KRZYŻANOWSKA</b>	Université Marie Curie-Skłodowska, Lublin, Pologne
<b>Katarzyna KWAPISZ-OSADNIK</b>	Université de Silésie à Katowice, Pologne
<b>Fabrice MARSAC</b>	Université de Strasbourg, France
<b>Salah MEJRI</b>	Université Sorbonne Paris Nord, France
<b>Igor MEL'ČUK</b>	Université de Montréal, Canada
<b>Teresa MURYN</b>	Université Pédagogique de Cracovie, Pologne
<b>Małgorzata NOWAKOWSKA</b>	Université Pédagogique de Cracovie, Pologne
<b>Michele PRANDI</b>	Université de Bologne, Italie
<b>Monika SUŁKOWSKA</b>	Université de Silésie à Katowice, Pologne
<b>Dan VAN RAEMDONCK</b>	Université Libre de Bruxelles, Belgique
<b>Joanna WILK-RACIĘSKA</b>	Université de Silésie à Katowice, Pologne

CORRECTION LINGUISTIQUE / LANGUAGE EDITORS

Paweł GOLDA (français), Ewa ŚMIŁEK (espagnol), Tomasz KALAGA (anglais)

SECRÉTARIAT DE RÉDACTION / EDITORIAL SECRETARY

Anna CZEKAJ      [anna.czekaj@us.edu.pl](mailto:anna.czekaj@us.edu.pl)

Institut de Linguistique  
Université de Silésie à Katowice  
ul. Grota-Roweckiego 5  
41-205 Sosnowiec  
Pologne

Accessible sous forme électronique en accès libre / Available in open access electronic form :

[www.journals.us.edu.pl/index.php/NEO/](http://www.journals.us.edu.pl/index.php/NEO/)

Central and Eastern European Online Library

[www.cceol.com](http://www.cceol.com)

Śląska Biblioteka Cyfrowa

[www.sbc.org.pl](http://www.sbc.org.pl)

## TABLE DES MATIÈRES

- Wiesław BANYŚ : Présentation
- Xavier BLANCO ESCODA, Rafael GARCÍA PÉREZ : Verbes supports intensifs et noms prédicatifs de la classe <colère> en ancien français et ancien espagnol
- Peter BLUMENTHAL, Giovanni ROVERE : Valence des verbes de perception en français et italien : perspective diachronique
- Krzysztof BOGACKI : Analyse linguistique au service du TAL
- Pierre-André BUVET : La prédication : une relation
- Wiesław BANYŚ : Inférences au pays de la prosodie
- Sibilla CANTARINI : Anglicisms in German – An Overview with Reference to the *Neologismen-wörterbuch* in the *Online-Wortschatz-Informationssystem Deutsch* (OWID)
- Françoise COLLINET : Pour une grammaire des notions et de leurs transformations rhétoriques
- Marco FASCIOLO, Qianqian ZENG : La synonymie : une relation lexicale comme les autres
- Aude GREZKA : Morfetik – mises à jour et évolutions d'une ressource en ligne
- Georges KLEIBER : Sémiotique des démonstratifs cataphoriques
- Denis LE PESANT : Gaston Gross, héritier et continuateur de Maurice Gross et de Zellig Harris
- Igor MEL'ČUK : Russian Free Phrasal Indefinite Pronouns
- Salah MEJRI : Prédicats, sens, polylexicalité et figement : un parcours heuristique
- Claude MULLER : Les pronoms démonstratifs neutres du français devant les subordonnées de type relatif
- Adriana ORLANDI : Regards croisés sur l'adjectif
- José A. PASCUAL RODRÍGUEZ : La caracterización lingüística de los textos. El criptoaragonesismo *servar* como pretexto
- Magdalena PERZ : Sur la synonymie dans le lexique adjectival – questions de sémantique
- Michele PRANDI : Les relations transphrastiques à la frontière entre la phrase et le texte
- Mirosław TRYBISZ : La actualización del nombre *COVID-19* (y sus sinónimos) en los portales de información de lengua española
- Beata ŚMIGIELSKA : Traduction automatique et désambiguïsation des sens des mots. Le cas du verbe français *louer*
- Zygmunt VETULANI, Grażyna VETULANI : Towards Lexicon-Grammar Verbnets Through Lexical Ontologies

## CONTENTS

- Wiesław BANYŚ: Presentation
- Xavier BLANCO ESCODA, Rafael GARCÍA PÉREZ: Intensive Support Verbs and Predicative Nouns of the class <colère> (Anger) in Old French and Old Spanish
- Peter BLUMENTHAL, Giovanni ROVERE: The valency of perception verbs in French and Italian: A diachronic perspective
- Krzysztof BOGACKI: Linguistic analysis at the service of NLP
- Pierre-André BUVET: Predication: a relationship
- Wiesław BANYŚ: Inferences in the land of prosody
- Sibilla CANTARINI: Anglicisms in German – An Overview with Reference to the *Neologismen-wörterbuch* in the *Online-Wortschatz-Informationssystem Deutsch* (OWID)
- Françoise COLLINET: For a grammar of concepts and their rhetorical transformations
- Marco FASCIOLO, Qianqian ZENG: Synonymy: a lexical relation like any other
- Aude GREZKA: Morfetik – updates and upgrades to an online resource
- Georges KLEIBER: Semiotics of cataphoric demonstratives
- Denis LE PESANT: Gaston Gross, heir and successor of Maurice Gross and Zellig Harris
- Igor MEL'ČUK: Russian Free Phrasal Indefinite Pronouns
- Salah MEJRI: Predicates, Sense, Polylexicality, and Freezing: a heuristic path
- Claude MULLER: French neutral demonstrative pronouns before relative-type subordinate clauses
- Adriana ORLANDI: Adjectives at the intersection of perspectives
- José A. PASCUAL RODRÍGUEZ: Linguistic characterization of texts. “Servar”, a crypto-aragoneseism, as a pretext
- Magdalena PERZ: On synonymy in the adjectives – object classes and semantic issues
- Michele PRANDI: Interclausal relations at the boundary between sentence and text
- Mirosław TRYBISZ: The actualization of the name *COVID-19* (and its synonyms) in the Spanish language information portals
- Beata ŚMIGIELSKA: Automatic translation and disambiguation of word meanings. The case of the French verb *louer*
- Zygmunt VETULANI, Grażyna VETULANI: Towards Lexicon-Grammar Verbnet Through Lexical Ontologies



## Présentation

Il y a un an, le 13 octobre 2022, nous avons perdu un pilier de la linguistique, le professeur Gaston Gross. Sa disparition a laissé un grand vide dans le monde de la linguistique française.

En sa mémoire, nous lui rendons hommage dans ce numéro spécial de notre revue.

Gaston Gross a largement contribué à *Neophilologica*, dont il a été membre du comité scientifique dès sa création et président d'honneur depuis quelques années.

Gaston Gross était une figure emblématique de notre discipline. Ses travaux ont jeté les bases de nombreuses recherches actuelles et ont inspiré des générations de linguistes. Gaston était plus qu'un professeur ; il était un érudit, un mentor, et un pionnier qui a façonné le domaine de la linguistique avec une passion inégalée et une perspicacité exceptionnelle.

Il y a peu de personnes dans le milieu académique qui possèdent autant de volonté, de compétence, et de capacité d'établir et d'organiser la coopération scientifique que Gaston Gross.

J'avais eu la chance de rencontrer Gaston Gross pour la première fois à la fin des années 70, lorsqu'il était venu à Katowice à l'invitation du directeur de l'Institut de Philologie Romane de l'époque, mon maître en linguistique, l'un des plus éminents linguistes polonais du XX<sup>e</sup> siècle, le professeur Stanisław Karolak. À cette époque, Gaston Gross était déjà un chercheur renommé dans le milieu linguistique français, développant des méthodes de description linguistique proposées par Zellig Harris et Maurice Gross au célèbre LADL – Laboratoire d'automatique documentaire et linguistique de l'Université Paris 7, qui est devenu la Mecque des linguistes du monde entier souhaitant appliquer des méthodes structurelles avancées à la description du langage, qui pourraient ensuite être implémentées dans l'apprentissage machine et le traitement informatique des

données linguistiques. Nos premières réunions étaient étincelantes de discussions brillantes et d'arguments en faveur de leurs théories : Gaston Gross – alors définitivement formaliste, structurel et distributionnel, et Stanisław Karolak – le créateur de l'école de description linguistique basée sur la logique des prédicats et la sémantique, qui était naturellement étrangère aux représentants de l'approche distributionnelle. Comme c'est le cas lors de la rencontre de grands esprits et d'exceptionnels savants, ces discussions et polémiques ont conduit à l'amélioration et à l'adaptation, et parfois à la modification de leurs conceptions. Ces rencontres et discussions ont duré plus de 25 ans, avec des conférences annuelles organisées conjointement dans notre université et des déplacements également à Paris pour des séminaires et des conférences dans un cercle encore plus large de savants, pas seulement français. Et pourtant, le nombre de savants marquants que Gaston Gross a réussi à inciter à venir à l'Université de Silésie a toujours été énorme, et ce qui est encore plus important, c'est l'unicité des chercheurs qui venaient, qui étaient déjà des figures remarquables à l'échelle mondiale. Pour n'en mentionner que quelques-uns, tels que Georges Kleiber, Jean Perrot, Jean-Claude Anscombe, Robert Vivès, Michel Mathieu-Colas, Jacqueline Giry-Schneider, Danielle Leeman, Pierre Lerat et bien d'autres.

Je peux dire avec grand plaisir que tout comme notre façon de voir la description linguistique a changé grâce à la coopération avec Gaston Gross, Gaston Gross a également modifié son approche de la description linguistique pour une approche plus sémantique, ce qui l'a finalement incité à créer une école scientifique originale et à fonder dans son université, l'Université Paris-XIII, le Laboratoire de Linguistique Informatique de renommée mondiale. C'est aussi grâce à ces techniques avancées de description informatique du langage naturel pour les différentes tâches associées à son traitement en vue de la synthèse et de l'analyse du langage, du résumé des documents et de la traduction automatique, qu'une équipe internationale, dirigée en ce qui concerne la partie française par Gaston Gross et l'équipe polonaise, composée principalement des linguistes de notre Université, a notamment reçu en 2000 le projet européen Matchpad (Machine Translation Systems for use of Hungarian and Polish Administrations) dans le cadre du Programme Cadre V de la Commission Européenne et plusieurs autres projets de recherche, obtenus que ce soit directement de la Commission Européenne ou dans le cadre de projets de coopération polono-française « Polonium » ou de l'Agence Nationale de la Recherche.

L'activité scientifique du professeur Gaston Gross est si riche qu'une description plus détaillée nécessiterait essentiellement l'écriture d'un livre sur les principales réalisations scientifiques de ce distingué savant. Il convient également de mentionner que, quel que soit le livre, tout comme dans le cas de centaines



d'articles écrits par Gaston Gross, le contenu de chacun d'eux est fondamental pour la linguistique générale et romane. Pour ne mentionner que les livres les plus importants sans parler d'un très grand nombre d'articles : *Les constructions converses du français* (1989), *Les expressions figées en français : noms composés et autres locutions* (1998), *La finalité : fondements conceptuels et genèse linguistique* (avec Michele Prandi, 2004), *Sémantique de la cause* (avec Ramona Pauna et Freiderikos Valetopoulos, 2009), et une synthèse des concepts et des méthodes utilisées par Gaston Gross présentée dans un excellent *Manuel d'analyse linguistique* (2012). Et la meilleure façon d'entamer cette entreprise est d'indiquer les principaux thèmes que Gaston Gross a abordés dans son travail monumental, résumant son activité scientifique, qu'il a écrit, cela vaut la peine de le souligner, dans le but de sensibiliser pas seulement les jeunes collègues linguistes ou étudiants, à la nécessité d'avoir une théorie appropriée et une méthode convenable pour analyser les faits linguistiques, parce que, comme on peut le constater facilement, ce n'est pas toujours le cas (le numéro 33 (paru en 2021) de *Neophilologica – Perspectives pour la linguistique et autres études* était justement motivé par ces questions). Il est donc nécessaire de diffuser de telles approches générales, et des approches précises, empiriques, qui, d'une part, comme Gaston Gross aimait à le dire et l'a exprimé avec force dans, p. ex., son texte : *Des perspectives rigoureuses pour la linguistique* paru dans le numéro cité de *Neophilologica*, font de la linguistique une véritable science (empirique), et qui, d'autre part, permettent diverses façons de traiter le langage naturel pour divers besoins. Gaston Gross a considéré ce livre – conformément au titre qu'il lui a donné – comme un véritable manuel d'analyse linguistique. Ce manuel de près de 400 pages s'est classé 10<sup>e</sup> dans le classement Google Books pour le mot-clé Linguistique.

Ce travail nous permet de comprendre la méthode créée et développée par Gaston Gross au fil des années, qui repose entre autres sur le concept de classe d'objets. L'idée de départ était relativement simple : on ne peut comprendre la signification d'une phrase uniquement sur la base de la signification des éléments-objets/actants/arguments qui apparaissent en surface avec un prédicat/verbe donné. Pour utiliser un exemple cité par Gaston Gross : le verbe/prédicat *défendre* dans les trois utilisations ci-dessous fonctionne de manière différente : *défendre un point de vue/défendre une position forte – défendre à qqn de sortir*. Déjà ici, la traduction en polonais et d'autres langues montre que les prédicats/verbes dans les exemples ci-dessus se comportent différemment, ne permettant pas les mêmes opérations syntaxiques et transformations dans toutes les utilisations indiquées. Si nous prenions des exemples encore plus variés, nous pourrions citer les utilisations françaises p. ex. du prédicat/verbe, classique dans ce type d'analyse, *prendre : prendre un café vs prendre un bain*. Ces constructions

sont particulières encore d'un autre point de vue, parce que le prédicat, l'élément le plus significatif, l'élément pivot, central, dans cette phrase, est le nom *bain* et non pas le verbe, comme les grammairiens classiques pourraient le penser. La simple constatation d'une telle situation n'est pas radicalement nouvelle, en principe les bons dictionnaires ont essayé d'organiser les descriptions des entrées pour rendre compte de telles différences dans l'utilisation des mots. Mais dans le cas de Gaston Gross, cette constatation a conduit à des travaux fondamentaux extrêmement riches en matériel et en descriptions, basés sur le concept de verbe support, mettant en lumière toutes les constructions de ce type en français, dont une grande partie, comme celles mentionnées ci-dessus, ont un caractère idiomatique, donc pas nécessairement reproductible et traduisible de manière canonique dans une autre langue.

Ainsi, ce qui est nouveau dans les travaux de Gaston Gross, c'est la spécification et la systématisation des connaissances sur le sujet et la création de descriptions de nombreuses catégories de ce type composées de dizaines de milliers d'enregistrements. L'accent a donc été mis sur une description précise, rigoureuse, soignée et exhaustive des systèmes de prédicats, qu'ils soient verbaux, adjectivaux ou nominaux. Cette idée se reflète dans la structure du travail dont nous parlons : d'abord, il est question de schémas de phrases généraux : prédicats-arguments, définissant le point de départ essentiel des analyses, puis Gaston Gross passe à l'utilisation des prédicats, définissant un ensemble de propriétés qui les caractérisent (surtout en ce qui concerne le type d'argument, l'actualisation, la modification), pour passer à une description systématique des prédicats verbaux, nominaux et adjectivaux. Cela permet de passer au concept de classes d'objets et de présenter la méthode utilisée pour les constituer, et c'est précisément le concept de classes d'objets, intrinsèquement lié à la description des prédicats nominaux et des verbes supports, qui est la clé pour assurer une traduction automatique correcte, comme l'a démontré Gaston Gross dans de nombreux travaux.

Gaston Gross consacre une discussion détaillée séparée au phénomène de figement, rejetant l'analyse dans le cadre de la logique bivalente, jetant un éclairage définitivement différent et novateur sur les expressions figées : de la logique classique vrai/faux, oui/non, en faveur d'une logique polyvalente, davantage oui/d'avantage non.

En conséquence, la description présentée du figement se situe davantage dans le cadre d'une approche de type prototype, mais pas prototypique, bien que ce ne soit pas le terme de Gaston Gross, en termes de constructions plus ou moins figées, +/- . Gaston Gross présente des hypothèses et des analyses détaillées de cette question dans un autre excellent livre consacré à ces questions, cité ci-dessus : *Les expressions figées en français : noms composés et autres locutions* (1998).

J'ai l'impression que l'importance des constatations absolument novatrices sur la nature des constructions figées et de leurs conséquences – comme c'est souvent le cas lors d'un changement de paradigme de pensée – n'est pas encore pleinement reconnue par les chercheurs. Cela vaut la peine de noter que nos discussions sur les phénomènes pragmatiques, tels que l'intonation et les structures thématiques et rhématiques, ont également trouvé leur place dans les analyses de Gaston Gross qui déclarait fréquemment la nécessité d'une meilleure intégration de leur description avec la description basée sur des analyses empiriques syntaxiques.

Si l'on essayait de décrire en quelques mots – bien que cela soit très difficile, voire impossible compte tenu de la richesse des contenus, des nouvelles idées et des concepts – l'importance et l'originalité des travaux de Gaston Gross, on dirait qu'ils consistent en un traitement holistique des faits linguistiques, ne traitant pas séparément et indépendamment, comme c'est souvent le cas, le lexique, la syntaxe et la sémantique, d'où le titre de ce numéro de *Neophilologica* consacré à Gaston Gross, mais en montrant l'interpénétration mutuelle de ces domaines dans la construction des énoncés, et en utilisant un appareil conceptuel très précis pour des descriptions très détaillées des événements linguistiques basées sur de très grands corpus de données linguistiques (les deux écoles de recherche : celle de Maurice Gross et celle de Gaston Gross, sont mondialement réputées pour de telles descriptions), principalement celles liées aux classes d'objets, aux verbes supports et aux expressions (+/-) figées, le tout présenté dans un format formel tel que l'implémentation informatique pour le traitement du langage naturel, en particulier pour la traduction automatique, est devenue possible.

Gaston Gross était également un enseignant dévoué. Il avait cette capacité rare à rendre les concepts complexes accessibles, inspirant ses collègues et ses étudiants à explorer les profondeurs de la linguistique avec curiosité et enthousiasme. Beaucoup de ceux qui ont eu le privilège d'assister à ses cours et conférences témoignent de l'impact durable qu'il a eu sur leur carrière académique et professionnelle.

La perte du Professeur Gross est ressentie par toute la communauté linguistique. Mais son héritage perdure dans les nombreux travaux qu'il a publiés, dans les esprits des collègues et des étudiants qu'il a formés, et dans les cœurs de tous ceux qui ont eu la chance de le connaître.

Quand nous nous souvenons du Professeur Gaston Gross, nous célébrons non seulement son immense contribution à la linguistique, mais aussi l'homme qu'il était. Sa passion pour la recherche, son dévouement envers ses collègues et ses étudiants, et son esprit généreux continuent d'inspirer ceux qui marchent sur les traces qu'il a laissées.

En dehors de la salle de conférences et du laboratoire, Gaston Gross était un collègue attentionné et un ami fidèle. Il était connu pour sa gentillesse, son sens de l'humour et sa passion pour la musique classique.

Gaston nous manque profondément...

Ce numéro de *Neophilologica*, dédié à Gaston Gross, comporte 21 contributions de ses amis, collègues, étudiants.

Le volume ouvre le texte de Xavier Blanco Escoda et Rafael García Pérez avec le titre *Verbes supports intensifs et noms prédicatifs de la classe <colère> en ancien français et ancien espagnol*. Cette étude analyse les verbes de support intensifs en ancien français et en ancien espagnol, en se concentrant sur les combinaisons lexicalement contraintes et les métaphores de base. Elle met en évidence les similitudes et les différences entre les langues, soulignant les perspectives synchroniques et diachroniques.

Peter Blumenthal et Giovanni Rovere, avec leur texte, *Valence des verbes de perception en français et italien : perspective diachronique*, examinent les verbes de perception (comme « entendre » et « voir ») d'un point de vue de la grammaire de construction, en se concentrant sur les aspects contrastifs et diachroniques. L'analyse révèle que différents verbes de cette classe ne montrent pas un comportement syntaxique homogène, avec le verbe *observer* étant unique. L'article discute de l'acceptabilité des phrases et met en évidence le rôle d'un facteur appelé *evidentia* dans la rhétorique classique.

Bohdan K. Bogacki, dans son texte *Analyse linguistique au service du TAL*, met en relief le fait qu'en utilisant la notion de prédicat dans le sens d'un élément structurant la phrase au niveau superficiel, G. Gross a abouti à une interprétation des faits linguistiques qui est valable pour ceux qui adoptent le concept de prédicat sémantique comme outil d'analyse et montre que ce mécanisme est crucial dans la traduction automatique, où il aide à choisir l'équivalent précis d'un lexème polysémique dans la langue cible.

Pierre-André Buvet, dans le texte intitulé *La prédication : une relation*, discute le concept de prédicat à la fois du point de vue métalinguistique et linguistique. Après avoir expliqué le rôle du prédicat dans le cadre théorique de référence, l'auteur analyse les mécanismes langagiers liés aux prédicats en illustrant son analyse avec la détermination des noms en français.

Wiesław Banyś, dans le texte *Inférences au pays de la prosodie*, examine le statut inférentiel de verbes appelés « verbes implicatifs » – les verbes implicatifs bidirectionnels et unidirectionnels. L'auteur souligne le rôle de la prosodie et du focus/thème dans la détermination du statut inférentiel des prédicats. L'analyse montre que la classification des inférences verbales théoriquement possibles doit

être modifiée une fois la prosodie prise en considération. Il n'y a pas 4 groupes de verbes implicatifs unidirectionnels, comme cela a été argumenté, mais 2, à savoir les groupes : [affirmé > vrai ou faux // nié > faux] du type être capable, pouvoir et [affirmé > vrai ou faux // nié > vrai] du type *hésiter à*. Les deux autres groupes, considérés comme distincts et autonomes verbes implicatifs unidirectionnels avec les caractéristiques suggérées : [affirmé > vrai // nié > vrai ou faux] du type *forcer à* et [affirmé > faux // nié > vrai ou faux] du type *refuser de*, appartiennent aux groupes canoniques des verbes implicatifs bidirectionnels. Cette différenciation du comportement différent de certains types de prédicats est cruciale pour automatiser la reconnaissance et la génération des inférences de prédicats, surtout dans les systèmes automatiques AI de la génération de la parole.

Sibilla Cantarini, dans son article *Anglicisms in German – An Overview with Reference to the Neologismenwörterbuch in the Online-Wortschatz-Informationssystem Deutsch (OWID)* met en relief le fait que les néologismes indiquent des changements linguistiques dans le lexique et montre que l'anglais exerçait une forte influence sur le vocabulaire allemand ces dernières décennies. Ces anglicismes varient en intégration, utilisation, style et diffusion au sein de la communauté linguistique. Le *Neologismenwörterbuch* offre un aperçu des néologismes dérivés de l'anglais en allemand, intégrant divers domaines du changement linguistique.

Françoise Collinet, dans le texte *Pour une grammaire des notions et de leurs transformations rhétoriques*, rappelle que dans un article de 1955, Perelman et Olbrechts-Tyteca suggèrent que leur théorie de l'argumentation pourrait fournir une base pour une étude qu'ils placent sous le signe de la sémantique : l'étude des notions. Cependant, les nombreux exemples du *Traité de l'argumentation* peuvent monopoliser l'attention, rendant des projets sémantiques comme des filigranes. L'auteure montre que l'analyse par Ducrot du mot *mais* pourrait améliorer la compréhension de la méthode Péreelmanienne.

Marco Fasciolo et Qianqian Zeng, dans leur article *La synonymie : une relation lexicale comme les autres*, traitent de la cohérence épistémologique de la synonymie. Lorsque la synonymie est définie comme égalité de sens, elle s'avère être une relation factuellement impossible. Par conséquent, les chercheurs distinguent entre synonymie totale (concept idéal) et synonymie partielle (phénomène réel). Les auteurs soutiennent que cette distinction doit être abandonnée afin de maintenir la cohérence dans la synonymie et constatent que la synonymie ne consiste pas dans l'absence de nuances sémantiques différentes entre les sens, mais plutôt dans la condition de trouver des nuances sémantiques différentes entre les sens. Cette condition est interprétée comme le partage du même « schéma prédicatif » au sens de Gaston Gross.

Aude Grezka, dans sa contribution *Mises à jour et évolutions d'une ressource historique, Morfetik* introduit une ressource linguistique morphologique, nommée Morfetik, pour le français moderne. L'article discute de l'évolution de cette ressource, de ses caractéristiques linguistiques et technologiques, et fournit un aperçu des nombreux outils qu'elle intègre. Morfetik est une plateforme en évolution constante conçue pour améliorer le traitement des données textuelles.

Georges Kleiber, dans le texte, *Sémiotique des démonstratifs cataphoriques*, traite de la recherche sur les expressions anaphoriques, suggérant que les études récentes devraient intégrer les découvertes des quatre dernières décennies avec des perspectives socio-cognitives contemporaines et d'autres approches multidisciplinaires. L'auteur se concentre sur les démonstratifs cataphoriques (CDs) et propose d'analyser leur fonction référentielle, questionnant comment les CDs nous guident vers le référent voulu et le processus interprétatif impliqué. L'analyse conclut que comparer les CDs aux démonstratifs gestuels plutôt qu'aux anaphoriques est plus perspicace, puisque les deux introduisent de nouveaux référents dans la mémoire du discours. Cette similarité conduit à une hybridité sémiotique partagée et à une opération référentielle largement identique, qui est décrite en neuf points. Toute différence est attribuée à la nature de l'index utilisé dans chaque cas.

Denis Le Pesant, dans son article *Gaston Gross, héritier et continuateur de Maurice Gross et de Zellig Harris*, vise à décrire précisément le travail de Gaston Gross et de son équipe au sein de ce qu'on peut appeler « L'École des Grammaires Distributionnelles et Transformationnelles ». Les principaux représentants de cette école sont Zellig S. Harris, Maurice Gross, Jean Dubois, Françoise Dubois-Charlier et Gaston Gross. L'auteur discute de la collaboration entre Maurice Gross et Zellig Harris dans le développement de la théorie des Grammaires Distributionnelles et Transformationnelles, ainsi que des contributions de Gaston Gross par rapport à celles de Maurice Gross et ses innovations par rapport au projet original de Harris.

Igor Melčuk, dans le texte *Russian Free Phrasal Indefinite Pronouns*, examine les pronoms indéfinis phrastiques en russe, à la fois phraséologisés (collocationnels) comme *koe-kto* ≈ 'quelqu'un', *Bog znaet kto* 'Dieu sait qui', *kto ugodno* 'qui que ce soit' ou *kto by to ni bylo* 'peu importe qui', et libres comme *{Ja vstretil} nikto iz tvoix družej ne dogadaetsja kogo* '{J'ai rencontré} aucun de tes amis ne devinera qui' (aussi connus sous le nom d'amalgames syntaxiques). L'auteur présente ensuite trois entrées lexicales pour le lexème pronominal indéfini *kto1* ≈ '-body; -one', le marqueur de type d'indéfinition prémodifiant idiomatique 'Bog znaet' 'Dieu sait', et le marqueur de type d'indéfinition postmodifiant particule -TO4 ≈ 'quelque'. Les représentations formelles des pronoms phrastiques indé-

finis collocationnels et libres à trois niveaux (sémantique, syntaxique profond et syntaxique de surface) sont fournies, ainsi que des règles pour construire la construction des deux types de pronoms.

Salah Mejri, dans la contribution *Du figement : un parcours heuristique*, rappelle que, pour Gaston Gross, une analyse approfondie de la langue implique la description de tous les usages des unités lexicales de la langue en tenant compte de la combinaison libre, en identifiant les schémas d'arguments pour les prédicats et les prédicats appropriés pour les arguments. Cette approche est complétée par une description des séquences figées avec une combinatorialité interne plus ou moins restreinte. L'auteur met en évidence la richesse de la pensée de Gaston Gross en soulignant son impact heuristique et en mettant en avant les concepts de prédicat, de figement, de polylexicalité et d'opacité sémantique pour illustrer l'importance des suggestions méthodologiques de Gaston Gross dans le développement de la linguistique.

Claude Muller, dans le texte *Les pronoms démonstratifs neutres du français devant les subordinées de type relatif*, étudie la construction des propositions relatives et des propositions clivées qui suivent un pronom démonstratif en français. L'article analyse comment ces structures acquièrent leur signification et leur référence, soit en augmentant la référence par le biais d'interactions prédicatives et de la signification lexicale, soit par la focalisation sur un référent dans la situation ou le contexte. L'auteur démontre que le pronom original cesse d'être purement démonstratif dans ces structures, et explique comment différents démonstratifs sont utilisés dans le français contemporain dans ces contextes.

Adriana Orlandi, dans sa contribution *Regards croisés sur l'adjectif*, présente les principaux sujets abordés par Gaston Gross dans son analyse des adjectifs. G. Gross visait à développer un dictionnaire électronique d'adjectifs pour le traitement automatique du langage. Ses recherches se sont concentrées sur l'identification et la description des différentes sous-classes d'adjectifs, ainsi que sur la signification adjectivale en utilisant la notion d'« emploi ». L'auteure examine certains concepts clés de la théorie des classes d'objets appliqués à l'étude de l'adjectif, y compris la notion de prédicat adjectival et de signification adjectivale (« emploi »). Un aperçu des réalisations de Gaston Gross en la matière est présenté et certaines questions critiques sont également discutées.

José A. Pascual Rodríguez, dans son article *La caracterización lingüística de los textos. El criptoaragonésismo servir como pretexto (Linguistic characterization of texts. « Servar », a crypto-aragoneseism, as a pretext)* souligne la nécessité d'une caractérisation linguistique approfondie des textes formant un corpus pour rendre possible l'analyse historique lexicale. Les obstacles rencontrés durant cette démarche sont illustrés en exposant l'analyse appliquée à une série de cinquante

documents. Ces documents servent à élucider l'origine aragonaise du terme « *servar* », un latinisme employé par le Marquis de Santillana.

Magdalena Perz, dans sa contribution *Sur la synonymie dans le lexique adjectival – questions de sémantique*, explore les difficultés liées à la relation de synonymie à travers les emplois adjectivaux. L'auteure examine différentes approches de la synonymie, offrant à la fois une dimension théorique et applicative. La notion de classe d'objets, telle que présentée par G. Gross et W. Banyś, est utilisée pour décrire la sémantique des adjectifs afin de les désambiguïser. L'approche adoptée ici pour décrire les unités synonymes considère la synonymie comme une relation entre mots en usage. L'auteure constate qu'une approche contextuelle de la synonymie pourrait pallier les lacunes des dictionnaires classiques de synonymes.

Michele Prandi, dans l'article *Les relations transphrastiques à la frontière entre la phrase et le texte*, rappelle que l'étude des relations entre les propositions est habituellement réalisée dans le contexte de la syntaxe des phrases complexes et subordonnées. Toutefois, le lien entre la structure complexe des phrases et les connexions interpropositionnelles n'est pas fondamentalement nécessaire. Ces liaisons ne découlent pas des valeurs sémantiques des subordonnées, mais plutôt des rapports conceptuels qui unissent des processus équivalents en termes de complétude. Dans cette optique, la phrase complexe représente simplement une des multiples voies d'expression de ces relations, à l'instar des enchaînements textuels. D'ailleurs, l'auteur constate qu'une analyse comparative des structures disponibles révèle que le texte, plutôt que la phrase complexe, constitue la forme d'expression privilégiée par défaut. Là où la phrase complexe superpose une architecture syntaxique hiérarchique à une structure conceptuelle équilibrée, le texte, lui, se montre isomorphe avec la structure conceptuelle des relations interpropositionnelles. Sélectionner la phrase complexe, une alternative distinctive, exige une justification fonctionnelle particulière, orientée non pas vers la nature du lien en soi, mais plutôt vers la mise en place d'une perspective communicative étagée, partageant l'information entre ce qui est mis en avant et ce qui constitue l'arrière-plan.

Beata Śmigielska, dans sa contribution, *Traduction automatique et désambiguïisation des sens des mots. Le cas du verbe français « louer »*, présente un aperçu de l'efficacité des systèmes modernes de traduction automatique, tels que ceux de Google, DeepL, ChatGPT et d'autres, qui utilisent la Traduction Automatique Neuronale (TAN) alimentée par les développements continus en intelligence artificielle. L'article évalue les capacités de traduction de ces systèmes pour le verbe français « louer » dans divers contextes. L'auteure examine ensuite l'interprétation du verbe à travers le dictionnaire Antidote 8. Finalement, l'application de l'Approche Orientée Objet à la traduction français-polonais de « louer » est discutée



visant à clarifier son sens et à établir un ensemble de règles syntactico-sémantiques pour guider une traduction automatique précise des phrases où ce verbe est utilisé.

Mirosław Trybisz, dans le texte *La actualización del nombre COVID-19 (y sus sinónimos) en los portales de información de lengua española*, remarque qu'au début de la pandémie de COVID-19, les informations sur la maladie ont été rapidement relayées par les médias, où l'emploi de la métonymie était fréquent. Son étude se penche sur l'utilisation du terme prédicatif *COVID-19* et de ses divers synonymes dans la presse en ligne espagnole. Une analyse des déterminants et des verbes supports référencés par les journalistes de langue espagnole est présentée pour affiner la compréhension de la syntaxe et des combinaisons possibles de ces termes prédicatifs. L'auteur souligne aussi que les conclusions de sa recherche pourraient différer en fonction du degré de spécialisation des textes examinés.


Le numéro termine l'article de Grażyna Vetulani et de Zygmunt Vetulani, *Towards Lexicon-Grammar Verbnet through Lexical Ontologies* où les auteurs présentent une étude s'inspirant du projet Princeton WordNet. Ces structures organisationnelles du lexique, adaptées à une langue ou à un domaine spécifique défini par les nécessités communicatives, se révèlent essentielles pour le développement de modèles de compétences linguistiques formelles, et sont par conséquent cruciales pour la création et l'implémentation de systèmes d'intelligence artificielle compétents en communication linguistique, aussi bien en réception qu'en production. Le travail présenté marque un progrès significatif dans le programme de recherche par l'introduction des ontologies lexicales avancées, nommées ici Lexicon-Grammar Verbnet. Les auteurs se réfèrent aux contributions de A. Colmerauer, Ch. Fillmore, Ch. Fellbaum, G. Gross, M. Gross, Th. R. Gruber, R. Kittredge, G. A. Miller, M. Palmer, K. Polański et P. Vossen.

*Wiesław Banyś*



***Xavier Blanco Escoda***

Université Autonome de Barcelone  
Espagne

 <https://orcid.org/0000-0001-8210-3668>

***Rafael García Pérez***

Université Carlos III de Madrid  
Espagne

 <https://orcid.org/0000-0002-1183-4700>

# Verbes supports intensifs et noms prédicatifs de la classe <colère> en ancien français et ancien espagnol

## **Intensive Support Verbs and Predicative Nouns of the class <colère> (Anger) in Old French and Old Spanish**

### **Abstract**

The intensive support verbs selected by the predicative nouns of the class <anger> in Old French and Old Spanish are presented and analyzed in this paper. The lexically constrained combinations between these verbs and nouns are described, and the basic metaphors that govern the selection of intensive structures are identified. We observe that a synchronic perspective would prefer a lexical selection-based description of collocations, whereas a diachronic perspective would prefer the use of conceptual metaphors. We highlight certain similarities between the two languages studied, but also some differences.

### **Keywords**

Diachronic Linguistics, Support Verbs, Old French, Old Spanish, Intensity

## 1. Introduction<sup>1</sup>

Il n'est nul besoin de présenter désormais la notion de verbe support, qui est bien connue de la communauté linguistique<sup>2</sup>. Nous avons traité cette notion par rapport à l'ancien français dans Blanco (2018)<sup>3</sup>. Nous y faisons déjà référence à une extension sémantique de ces verbes qui ajoutait à l'actualisation du prédicat nominal un sens intensif ('très/beaucoup'). Pour prendre un exemple simple en français contemporain : *avoir peur* est une structure à verbe support neutre, tandis que *mourir de peur* est une structure à verbe support intensif<sup>4</sup>.

Dans cet article, nous allons nous centrer sur ce type de verbes supports et sur des structures sémantiquement équivalentes (c'est-à-dire, permettant d'actualiser un nom prédicatif et, en même temps, de véhiculer un sens intensif). Nous adopterons une perspective contrastive ancien français/ancien espagnol. Dans la mesure où il ne nous est pas possible de présenter ici la très grande variété de ces formes, nous choisirons comme exemple des noms prédicatifs de la classe <colère>, dont les principaux représentants sont, pour l'ancien français et par ordre de fréquence, *ire*, *maltaient* et *courroux* et, pour l'ancien espagnol : *ira*, *saña*, *enojo*, *furor*, *furia*, *cordojo*, *corroço*.

Tout au long de l'article, nous tiendrons compte de la dimension quantitative des cooccurrences observées. Toutefois, en ce qui concerne le français, il est important de signaler que ces données sont utilisées de manière indicative car la taille du corpus sur lequel s'appuie notre travail (le sous-corpus *Ancien français* de Frantext) est restreinte<sup>5</sup>. En revanche, les corpus de l'espagnol dont nous disposons (CORDE et CDH<sup>6</sup>, de l'Académie Royale d'Espagne) peuvent être consi-

---

<sup>1</sup> La recherche qui sous-tend cet article a été financée par le Projet COLINDANTE (*Las colocaciones intensivas del francés antiguo y su traducción al francés y al español*, Proyecto I+D+i PID2019-104741GB-I00) (Financé par MCIN/AEI/10.13039/501100011033). *Ministerio de Ciencia e Innovación* (Espagne). Nous remercions Dolores Català et Julio Murillo pour leur relecture attentive de cet article.

<sup>2</sup> Pour une présentation claire et concise, on peut consulter Gross (2012 : 97), ainsi que Mel'čuk (2005).

<sup>3</sup> Le lecteur intéressé y trouvera des indications bibliographiques concernant l'étude des verbes supports en ancien français (références que nous ne reprenons pas ici pour des raisons d'espace).

<sup>4</sup> Sur la notion de verbe support intensif, on peut consulter Blanco et Buvet (2004) et Gavriilidou (2004).

<sup>5</sup> Il est constitué, en effet, d'un ensemble de 75 textes contenant un total de 2 813 747 mots.

<sup>6</sup> Le corpus CORDE contient 250 millions d'occurrences ; le CDH est composé pour sa part de plus de 53 millions d'occurrences, dont 32 appartiennent à des textes espagnols et plus de 20 millions à des œuvres américaines. Ce deuxième corpus a fait l'objet d'un travail phi-

dérés beaucoup plus complets d'un point de vue quantitatif et qualitatif, même si, comme nous l'avions déjà annoncé dans une étude précédente sur le verbe *hacer* 'faire' (García Pérez, 2007 : 19–22), tous les corpus sont toujours nécessairement incomplets, et ce de plus en plus au fur et à mesure que nous remontons dans le temps.

Les ouvrages d'où sont issus les exemples français sont cités d'après les conventions de Frantext (titre, année, n° de page). Quant aux exemples espagnols, ils sont cités comme suit : année, auteur (s'il est connu), titre, éditeur.

## 2. Noms prédicatifs de la classe <colère> en ancien français

### 2.1. *Ire*

Le substantif de la classe <colère> le plus fréquent en ancien français est *ire*, aujourd'hui disparu sauf par archaïsme. Le support neutre de *ire* est *avoir* (dans une construction à déterminant zéro qui comporte, très souvent, l'adjectif intensif non collocationnel *grant* appliqué au nom prédicatif<sup>7</sup>) :

(1) *Or ad Carles grant ire (La Chanson de Roland, 1125, p. 212)*

Il faut cependant tenir compte du fait que *ire* est, en soi même, un intensif, que l'on pourrait paraphraser comme 'colère vive' (cf. DMF s.v. *ire*). Les structures à verbe support intensif (nous incluons ici celles qui s'organisent autour d'un adjectif : 'être plein de', 'être fou de') mettent en relief donc (plus qu'elles n'introduisent) ce caractère d'émotion très forte, qui implique une certaine perte de contrôle de soi-même. Cela peut expliquer l'importance quantitative des structures à verbe support intensif. En effet, sur environ 700 occurrences de *ire*<sup>8</sup> dans le corpus d'ancien français de Frantext, environ 250 apparaissent dans une

---

logique approfondi, et permet ainsi le choix des occurrences non seulement selon la date des documents conservés, mais aussi selon la date où les textes ont été rédigés en première instance).

<sup>7</sup> Dans quelques cas exceptionnels, on peut trouver d'autres adjectifs intensifs. Kleiber (1978 : 306) donne le contexte suivant extrait de la *Chronique des ducs de Normandie : E celui heent de neire ire* (v. 18081). C'est un exemple intéressant dans la mesure où *neir* est ici un collocatif (cf. en français contemporain : *une colère noire*).

<sup>8</sup> Précisons que *ire* a aussi l'acception 'douleur' en ancien français.

structure de ce type<sup>9</sup>. Les collocations intensives semblent s'organiser autour de quelques métaphores vives (puisqu'elles donnent lieu à une diversité notable de moyens d'expression).

### 2.1.1. Autour de 'brûler d'ire'

La métaphore la plus productive est liée à la notion de 'feu' (*brûler d'ire* et similaires)<sup>10</sup>. Nous trouvons jusqu'à onze verbes différents dans environ cinquante contextes. Le verbe le plus fréquent (plus de la moitié des cas observés) est *esprendre* ('allumer'), sous sa forme participiale (*espris d'ire*) :

(2) *Mout sonz dolant et espris d'ire* (*Miracles de Notre-Dame*, t. 2, 1218, p. 242)

Dans les quelques cas où le verbe apparaît dans des formes finies, il a une valeur inchoative :

(3) *Ceste parole le roi grieve; d'ire esprent, em piez se lieve* (*Le Roman de Thèbes*, t. 2, 1150, p. 51)

Les autres verbes relevant de cette métaphore présentent uniquement quelques occurrences dans notre corpus (entre trois et quatre contextes). Il s'agit de : *ardre*, *embraser*, *enflamber* y *enflammer* :

(4) *La pucele art d'ire et de honte* (*Cligès*, 1176, p. 60b)

(5) *Et embrasés d'ardeur et d'ire* (*Miracles de Notre Dame*, t.3, 1218, p. 37)

(6) *Kar de ire esteit trop enflambé* (*Ipomédon*, 1180, p. 352)

(7) *que la reine an est antree, de grant duel et d'ire anflamee* (*Conte du Graal*, 1181, p. 364v)

En plus du feu, il est fait mention de ses effets : *bouillir*, *eschaufer*, *fondre* (ce dernier est assez fréquent, avec une douzaine de cas) :

(8) *Nous en avons si grans pesances Que d'ire en somes tout boulant* (*Miracles de Notre Dame*, t.2, 1218, p. 242)

<sup>9</sup> Rappelons qu'une bonne partie des contextes restants présentent tout de même l'intensif (non collocationnel) *grant* (plus d'une centaine d'occurrences). Il faut tenir compte également des structures consécutives à caractère intensif (*tant... que*, *telle... que*) (une vingtaine de cas).

<sup>10</sup> La linguistique cognitive a beaucoup insisté sur la métaphore en tant que mécanisme articulatoire de notions, cf. Lakoff et Johnson (1980) et Lakoff (1987).

- (9) *Ensi trestuit d'ire eschaufé* (*Yvain ou Le Chevalier au Lion*, 1177, p. 83e)  
 (10) *La dame l'ot, a pou d'ire ne font* (*La prise d'Orange*, 1200, p. 94)

Il est fréquent de trouver une accumulation de supports et de noms prédicatifs dans un seul contexte (accumulation para-synonymique chère à la langue médiévale) :

- (11) *De ires trestrembla e ruvi*<sup>11</sup>, *Echaufe e art e de ire esprent E de hunte e de mal talent* (*Ipomédon*, 1180, p. 434)

Notons, également, que des intensifieurs adverbiaux (que ce soit des adverbes en *-ment* ou des structures comparatives) peuvent venir s'ajouter aux supports :

- (12) *D'ire et de maltalent fu durement esprise* (*Buevon de Conmarchis*, 1271, p. 156)<sup>12</sup>  
 (13) *Vint d'ire plus ardanz que breise* (*Yvain ou Le Chevalier au Lion*, 1177, p. 82d)

La variété de verbes, ainsi que l'emploi fréquent de la mitigation (*a poi que 'c'est à peine si'*), nous indiquent que la métaphore n'est pas complètement fossilisée, et qu'elle est encore perçue en tant que telle.

- (14) *Elidus oï que cil dist, A poi que dire nen esprist* (*Lais de Marie de France*, 1180, p. 608)

### 2.1.2. Autour de 'remplir d'ire'

Cette émotion peut être perçue également comme une sorte de fluide ou substance<sup>13</sup> qui s'empare de quelqu'un en le « remplissant ». Nous comptons ainsi une quarantaine d'exemples de 'plein d'ire'. Il y a lieu de signaler que souvent *ire* est coordonné avec un autre nom d'affect : *angoisse, destorbier, dolor, duel, maltalent, rage* :

- (15) *Et plaine de dolor et d'ire: Des iaux pleure* (*Deuxième continuation de Perceval*, 1210, p. 155)

<sup>11</sup> *ruvir* 'devenir rouge', cf. AND s.v. *ruvir*.

<sup>12</sup> Cet ouvrage contient le seul exemple que nous avons trouvé du verbe *alumer* comme support d'ire : *De ce a si vos fieus le cuer d'ire alumé* (*Buevon de Conmarchis*, 1271, p. 104).

<sup>13</sup> Rappelons que le tempérament colérique était mis en rapport avec un excès de bile jaune (colère du grec *χολή*, 'bile') qui « remplissait » le sujet.

- (16) *D'ire e de rage furent plein Hoel e sun cusin Walwein (Le Roman de Brut, 1155, p. 668)*

En général, 'plein d'ire' s'applique directement au sujet humain qui est en proie à cet affect, mais dans un petit nombre de cas, il est question du 'cœur' comme siège de l'émotion :

- (17) *Lu cuer ot gros e d'ire plein (Le Roman de Brut, 1155, p. 758)*

La seule variante lexicale observée pour 'plein' est *farsir* (et ce dans un seul contexte) :

- (18) *Li dui baron furent d'ire farsiz (Aspremont, 1190, p. 106)*

### 2.1.3. Autour de 'devenir fou d'ire'

Nous avons une trentaine d'exemples de supports appliqués au sens 'devenir fou d'ire'. La combinaison la plus fréquente (deux tiers des exemples observés) est *enrager d'ire* où *enrager* correspond à 'perdre la raison, devenir fou (de douleur physique ou morale)' (cf. DMF s.v. *enrager 2.b.*). Cette expression, qui frôle le pléonasmisme (la rage étant aussi, par extension de sens, un état affectif, comme l'ire), est perçue comme hyperbolique, comme le prouve le fait qu'elle est souvent accompagnée d'un atténuateur (*a poi, par poi...*) :

- (19) *Ot le Guillelmes, a pou d'ire n'enrage (La prise d'Orange, 1200, p. 107)*  
 (20) *Par pou Hersent n'enrage d'ire (Roman de Renart, Branche 7, 1175, p. 17)*

Nous trouvons ensuite (en petit nombre, cinq exemples pour chacune de ces suites) des expressions comme *forcené d'ire* et *desver d'ire* :

- (21) *mais tut forsénéz de ire vint en Samarie (Li Quatre Livre des Reis, 1175, p. 165)*  
 (22) *Quant il virent ce, si furent tuit desvé d'ire et de mautalent (La Queste del Saint Graal, 1220, p. 121)*

Et l'on peut relever quelques combinaisons encore moins fréquentes (une ou deux occurrences), comme *afoler d'ire*, *faunoyer d'ire*, *forvoyer d'ire*, *issir* (ou *perdre*) *del sens d'ire*. Pour ne pas alourdir la présentation, nous donnons juste un exemple :

- (23) *por . i . petit que il n'issi del sens , de maltalent et d'ire (Vengeance Raguidel, 1200, p. 10)*

#### 2.1.4. Autour de 'rougir d'ire'

Notre corpus contient une vingtaine de cas où le verbe support intensif est sémantiquement lié avec la notion de 'couleur'. La combinaison la plus fréquente est 'teint d'ire' (une dizaine de cas), suivie de 'rougir d'ire' (cinq cas). Il semble raisonnable d'interpréter, en général, 'teint' comme 'coloré de rouge', dans la mesure où l'ire se manifeste surtout dans le visage. Être *teint* est être 'teint de sang'. Par ailleurs, *teint* entre, dans certains contextes, en contraste avec 'noir' et avec 'perse', deux autres tonalités associées à l'émotion qui nous occupe :

- (24) *Tans a douleur et tant a d'ire Et tant est perse, noire et tainte (Miracles de Notre-Dame, t. 4, 1218, p. 57)*

Dans l'exemple suivant, qui semble sortir du même moule, *pâle*<sup>14</sup> est contrastif par rapport à *teint* et à *pers* :

- (25) *Li reis, sun uncle, i est venu, De ire esteit pale e teint e pers (Ipomédon, 1180, p. 343)*

*Pers d'ire* apparaît trois fois dans notre corpus, mais toujours coordonné avec une autre couleur. Le qualificatif serait à interpréter comme un ton bleuâtre semblable à celui qui est provoqué par un coup.

Concernant *rougir* comme intensif d'*ire*, le verbe est parfois, à son tour, intensifié par des comparaisons<sup>15</sup> et apparaît une fois coordonné avec *étinceler* :

- (26) *D'ire et de maltalent rougi comme charbons (Buevon de Conmarchis, 1271, p. 119)*

- (27) *De maltalent et d'ire rougist et estincele (Roman d'Alexandre, branche 3, 1180, p. 145)*

'Noircir d'ire' est moins fréquent (nous n'en trouvons que deux exemples). Et nous avons trois exemples 'changer de couleur' que l'on pourrait rapprocher de deux autres exemples avec 'décolorer' (cf. aussi, ci-dessus, 'pâle') :

<sup>14</sup> La combinaison 'pâle d'ire' est rare. Nous trouvons cependant une occurrence de 'pâlir d'ire' : *De mautalent et d'ire est tous empaleis (Roman d'Alexandre, branche 3, 1180, p. 269)*.

<sup>15</sup> Sur les comparaisons intensives appliquées à 'rouge', cf. Blanco (à paraître).



- (28) *D'ire nercis trestoute* (*Roman de Renart*, branche I, 1180, p. 66)  
 (29) *D'ire et de mautalant, color En a Meleaganz changiee* (*Lancelot ou Le Chevalier à la Charrette*, 1177, p. 39a)  
 (30) *Marsilies fut esculurez de l'ire* (*La Chanson de Roland*, 1125, p. 58)

### 2.1.5. Autour de 'mourir d'ire'

Dans à peine une douzaine de cas, 'ire' est intensifié par un verbe support qui signifie (ou implique) 'mourir'. Le verbe *mourir* lui-même n'apparaît que dans trois contextes. Trois autres exemples font intervenir le verbe *s'occire* ('se tuer')<sup>16</sup>, quatre ont recours à l'euphémisme *partir* et un dernier au familier *crever*<sup>17</sup>. À noter que sept des onze exemples trouvés proviennent des *Miracles de Notre-Dame*, où différents verbes apparaissent dans des contextes fort semblables. Voici un exemple pour chacun des intensifs mentionnés :

- (31) *Preudom, il muert de duel et d'ire* (*Le jeu de saint Nicolas*, 1200, p. 136)  
 (32) *Nule novele n'en seit dire Pour un petit ne s'ocit d'ire* (*Miracles de Notre-Dame*, t. 4, 1218, p. 51)  
 (33) *Tel duel en a et tel contraire Por un petit qu'il ne part d'ire* (*Miracles de Notre-Dame*, t. 3, 1218, p. 113)  
 (34) *S'en ai tel duel et tele envie Por un petit d'ire ne crief* (*Miracles de Notre-Dame*, t. 1, 1218, p. 62)

### 2.1.6. Autres structures intensives appliquées à ire

Outre les supports signalés, nous avons encore quelques autres formes, comportant cependant beaucoup moins d'exemples. Il s'agit, d'une part, de diverses manifestations physiques, qui véhiculent une idée d'intensité : une douzaine de contextes de 'frémir d'ire' (*frémir* est le verbe le plus fréquent, mais nous trouvons aussi *trembler*, *tressauter* et *trepeiller* 'trépigner'), une dizaine de contextes de 'suer d'ire' (en général avec le verbe *tressuer*), ainsi que quelques occurrences où l'ire se manifeste par la perte de la parole.

- (35) *L'enpereres d'ire tressue Qant la parole a entandue* (*Cliqès*, 1176, p. 78e)

<sup>16</sup> La forme *occire*, archaïque aujourd'hui, était usuelle en ancien français.

<sup>17</sup> Le contexte suivant est intéressant dans la mesure où *trépasser* se combine avec *mourir*, mais non pas comme verbe support intensif, mais comme terminatif : *Dunc cumença Artur a rire, Kar dunc fu trespassee s'ire* (*Le Roman de Brut*, 1155, p. 602).

- (36) *Ismeine le veit, mut se dute, De ire tressaut e fremist tute (Ipomédon, 1180, p. 430)*
- (37) *tous li sans li remue, De maltalent et d'ire la parole a perdue (Roman d'Alexandre, branche 3, 1180, p. 146)*

Nous avons aussi d'autres intensifs avec des constructions de verbe support d'occurrence (où le nom prédicatif est en position de sujet), comme *acorer* ('percer le cœur': *l'ire acore qqn*), *dépecer*, *commouvoir*, *confondre*, ou *embattre*. Mais nous n'avons qu'un seul contexte pour chacun de ces verbes. Nous donnons uniquement comme exemples des cas qui sont clairement intensifs, les autres pouvant être considérés comme des variantes du verbe support neutre *avoir*, comme le sont *émouvoir* (*être esmu de*: *Il fu esmeus d'ire et de maltalent, La suite du Roman de Merlin, 1235, p. 159*), *porter* ou *être en*<sup>18</sup> (et encore d'autres aspectuels comme *durer*, *monter*, *prendre* ou *tourner*).

- (38) *ne croi pas que ge peüsse longuement vivre; car duels et ire m'àcoreroit (La mort le roi Artu, 1230, p. 73)*
- (39) *Dous pansers vient a chief de piece Qui l'ire et la dolor depiece (Le Roman de la Rose, 1230, p. 170)*

D'autres contextes rares, et qui ne sont pas vraiment intensifs, seraient 'rire'<sup>19</sup> d'ire' ou l'expression *froncir le grenon* 'plisser les lèvres (en signe de mécontentement)' (cf. DMF s.v. *grenon*):

- (40) *Quant li sire ot quë ele dist, D'ire e de maltalent en rist (Lais de Marie de France, 1180, p. 462)*
- (41) *De maltalent et d'ire a froncié le grenon (Buevon de Conmarchis, 1271, p. 46)*

## 2.2. Maltalent, courroux et rage

*Maltalent* et sa variante *mautalent* (dans l'acception qui correspond à 'colère') présentent 196 occurrences dans le sous-corpus *Ancien français* de Frantext.

<sup>18</sup> Parfois *être en Adj ire*: *Forment fu li rois de grant ire (Roman de Thèbes, t. 2., 1150, p. 54)*. Faisons remarquer les intensifs *forment*, appliqué au verbe support, et *grant*, appliqué au nom prédicatif.

<sup>19</sup> L'exemple avec 'rire' est singulier, mais nous avons aussi un cas avec 'sourire': *Sovent m'avient que je sorri De mautalent, d'ardeur et d'ire Quant j'oi a aucun buisnart dire (Miracles de Notre-Dame, t. 1, 1218, p. 5)*.

Environ 40 % de ces exemples contiennent une structure à verbe support intensif. Les intensifs les plus communs sont ceux qui prennent appui sur la métaphore ‘feu/chaleur’, en particulier *de maltalent espris* et *de maltalent ard*, mais il y a aussi quelques cas des verbes ‘embraser’, ‘enflammer’, ‘échauffer’, ‘bouillir’, et deux occurrences d’un verbe intensif que nous n’avons pas trouvé combiné avec *ire* : *frire* (plus un cas de *défrire*). Notons, cependant, que ces derniers se concentrent dans un seul ouvrage :

(42) *De mautalent frit toz et art* (*Miracles de Notre Dame*, t. 3, 1218, p. 113)

(43) *De mautalent prend à défrire* (*Miracles de Notre Dame*, t. 3, 1218, p. 218)

Nous avons une quinzaine de cas qui s’appuient sur la métaphore du changement de couleur, en général ‘rougir de maltalent’, mais aussi quelques cas épars de ‘devenir vermeil’, ‘teindre’, ‘noircir’, ‘pâler’, ‘devenir pers’, ‘muer de couleur’. Cette émotion, comme l’ire, peut donc faire devenir rouge, noir, pâle ou bleuâtre. Viennent ensuite une série d’autres métaphores (avec quatre ou cinq contextes chacune) que nous avons trouvées également pour *ire*<sup>20</sup> :

– ‘fou de maltalent’, avec différentes solutions lexicales, comme *devenir desvee*, *se desmesurer*, *changier le sens* :

(44) *De mautalent cuida le sens changier*, (*Aspremont*, 1190, p. 346).

– ‘plein de maltalent’, dans tous les cas avec l’adjectif *plein* :

(45) *Touz en est plainz de mautalent si li respont mout fierement* (*Le Roman de Thèbes* t. 1, 1150, p. 45).

– ‘mourir (ou se tuer) de maltalent’ :

(46) *Senescaus, a poi je nèsrabe, Et muir de mautalent et d’ire* (*Le jeu de saint Nicolas*, 1200, p. 71).

– différentes manifestations physiques (mis à part la couleur que nous avons déjà évoquée) : ‘trembler’, ‘suer’, ‘perdre la parole’, ‘serrer les dents’, ‘se pâmer’ :

(47) *de mautalent fremist et tremble* (*Le roman d’Eneas*, 1160, p. 240).

Pour ce qui est de *corroz* ‘courroux’ et ses variantes (*courous*, *corroux* et *corus*) nous les trouvons dans 49 contextes de notre corpus, dont uniquement deux comportent une structure à support intensif. Il s’agit du sémantisme ‘mourir de courroux’, représenté par une occurrence de *mourir* et une de *crever* :

<sup>20</sup> Mis à part d’autres supports où l’intensité est moins marquée ou moins claire et qui pourraient être considérés comme des variantes spécifiques du support neutre : ‘jurer de maltalent’, ‘rire de maltalent’ ou ‘souponner de maltalent’.

- (48) *Perceval remaint tant dolenz que il cuide bien morir de corroz* (*La Queste del Saint Graal*, 1220, p. 89)
- (49) *Cele parole tant greva Qué par . i . po qu'il ne creva de mautalant et de corroz* (*Perceval*, 1181, p. 365f)

Quant à *rage*<sup>21</sup>, nous en avons repéré 99 occurrences, mais seul un petit nombre correspondrait au sens 'colère' ; le sens prédominant serait plutôt celui de 'souffrance morale'. Il est, par ailleurs, souvent difficile de faire le tri entre ces deux sens. Le contexte *plain de rage* est la tournure intensive préférée, p. ex. : *sire Gonbert tot plain de rage, / ses poinz detort, ses poils arache* (Pierre de Saint-Cloud, *Roman de Renart*. Branche 7, 1175, p. 52) ou *La s'ert uns grans lyons tapis / Qui tant ert fiers et plains de rage / De genz faisoit mout grant damage* (*Miracles de Notre-Dame*, 1218, t. 4, p. 383).

Il y a lieu de remarquer que *rage* se combine avec l'adjectif intensif *mortel* (deux occurrences dans la *Chanson de Roland* et deux occurrences dans le *Roman d'Enéas*, cf. p. ex. : *El cors vos est entree mortel rage* dans *La Chanson de Roland*, 1125, p. 74).

D'autres noms prédicatifs (p. ex. *forsen*, qui, mis à part le sens 'folie', présente aussi le sens 'rage, iré' : *que ne me prent forsen ou rage*, *Le roman de Thèbes*, 1150, t. 1, p. 80) pourraient également être retenus. Mais il sont nettement moins fréquents.

### 3. Les noms prédicatifs de la classe <colère> en ancien espagnol

Comme en français, le substantif de la classe <colère> le plus fréquent est *ira*, que nous trouvons déjà dans les premiers textes conservés (Menéndez Pidal, 2004, s.v. *ira*). Il était suivi, dans cet ordre, par *saña*, *enojo*, *furor*, *rabia*, *furia*, *cordojo* et *corroço*<sup>22</sup>. De la même manière, le support neutre le plus habituel était *tener* 'avoir' (et son synonyme *haber*) qui permettait d'actualiser une structure

<sup>21</sup> Rappelons que le verbe *enrager* est un des supports du nom prédicatif *ire* (cf. ex. : 19, 20).

<sup>22</sup> Mais il ne faut pas oublier la polysémie d'une bonne partie de ces substantifs (ce qui rend l'analyse des corpus particulièrement délicate). *Saña* avait pris le sens 'violence', notamment précédé de la préposition *con* 'avec'. *Enojar* signifiait dans un premier temps 'horrifier, dégoûter' (DECH), et donc 'causer des désagréments, des dommages', sens que nous trouvons souvent dans les textes et qui a été transmis au dérivé *enojo*, surtout dans la construction à verbe support *hacer enojo*. Quant à *furor* et *furia*, ils étaient très employés dans un contexte guerrier avec le sens étymologique 'délire

syntaxique biactantielle (N0 : hum ; N1 : *a, de, contra* hum), comme nous pouvons le constater dans les exemples ci-dessous. Il est à noter que ces substantifs, une fois actualisés, sélectionnaient couramment aussi le prédicat adjectival *gran(de)*<sup>23</sup> et éventuellement *fuerte* ‘fort’ :

- (50) *Et sabet / señor que non es menester que uos tengades tan grant ira contra tristan* (c1313–c1410, *Cuento de don Tristán de Leonís*, ed. Ivy A. Corfis)
- (51) *E con la muy grand sanna que tenie porque ueye que aquel fecho serie a danno de so sennor ponpeyo & de si. dio contra los sos grandes uozes* (c1270, Alfonso X, *Estoria de Espanna*, ed. Pedro Sánchez Prieto)
- (52) *E los del Consejo vinieron, e con gran enojo que él tenía les afeó porque avían dexado el combate* (1406–1411, *Crónica de Juan II de Castilla*, ed. Juan de Mata Carriazo y Arroquia).
- (53) *Entonces licomedes fue perdiendo ya quanto el furor que tenia* (a1490, *Crónica Troyana*, ed. Dawn Prince)
- (54) *Y entre los unos e los otros había tan gran confusión y escándalo, que no había lugar para lo pacificar, porque la furia que a la hora tenían les privaba el entendimiento para obedecer a la Reyna como debían* (p1480–1484, Hernando del Pulgar, *Crónica de los Reyes Católicos*, ed. Juan de Mata Carriazo).

L'existence de la variante synonymique *haber* ('avoir') a conduit, pendant toute la période médiévale, à une certaine distribution complémentaire des fonctions pour certains de ces substantifs. Si le verbe *tener*, comme il ressort des exemples précédents, était toujours le support neutre « actif »<sup>24</sup>, le verbe *haber*, bien qu'il ait été largement employé dans ce même sens (55), est devenu le support neutre converse par excellence. En tant que tel, il était particulièrement fréquent avec le nom *ira* (56) :

---

furieux' (DECH). L'idée de colère constitue un glissement sémantique qui n'est pas toujours facile à décerner dans beaucoup de contextes.

<sup>23</sup> Malheureusement, tous les noms de la classe de prédicats <colère> n'ont pas laissé de traces de constructions à verbe support (comme *corroço*) ou elle restent très limitées (comme *cordojo*, qui apparaît uniquement avec le verbe *haber* 'avoir').

<sup>24</sup> Étant donné que ce travail porte sur les verbes supports intensifs, nous ne rentrerons pas dans les détails concernant la sélection de ces verbes neutres. Il suffira d'indiquer ici que cette sélection n'est pas identique – ni chronologiquement ni quantitativement – pour tous les noms prédictifs. *Tener*, par exemple, était largement sélectionné par *ira* ou *saña* depuis le XIII<sup>ème</sup> siècle, mais il faudra attendre le XIV<sup>ème</sup> siècle pour qu'il soit sélectionné par *enojo* et le XV<sup>ème</sup> siècle – toujours dans une moindre mesure – dans le cas de *rabia*, *furor* et *furia*.

- (55) *Los griegos otrosy eran muy despagados de todo aqueste fecho, e desaueniense vnos con otros e auian muy grand saña e muy grand pesar...* (c1270, *Historia troyana en prosa y verso*, ed. Ramón Menéndez Pidal).
- (56) *Ningun omme qui esta uendida, con so robra quisie peciar, aya la yra de Dios, e in coto del rey de la tierra peche L 5 morabedis* (1236, Anónimo, « Ortiz vende en Centollinos, a don Ramiro, cuantas heredades le pertenecían de su suegro Ruy Díaz », *Colección diplomática de San Salvador de Oña*).

À la différence du français, les structures à verbe support intensif ne sont pas les plus nombreuses, au moins par rapport aux verbes supports neutres cités ci-dessus, mais elles sont très intéressantes pour comprendre la conception que les locuteurs se faisaient de cette sous-classe de noms prédicatifs de sentiment.

### 3.1. Les verbes supports intensifs et la métaphore

Nous savons maintenant que les termes employés pour désigner les émotions sont associés aux sensations physiques. La sélection des verbes supports qui actualisent les substantifs prédicatifs de la classe <colère> en espagnol ne constituent pas une collection aléatoire, mais, au contraire, il s'agit d'une série structurée sur la base d'un modèle cognitif implicite dans la sémantique de la langue (Lakoff & Johnson 1980 ; Kövecses, 1990) et construit à partir de deux métaphores de base, partagées par de nombreuses autres langues – et parmi elles, le français, comme nous l'avons vu précédemment<sup>25</sup> – : LA COLÈRE EST DU FEU et LE CORPS EST UN RÉCEPTACLE DES ÉMOTIONS.

#### 3.1.1. Autour de la métaphore du feu

La métaphore conceptuelle LA COLÈRE EST DU FEU explique, précisément, le transfert lexical direct entre les principaux verbes distributionnels qui sélectionnent l'objet « feu » et les verbes supports qui actualisent les émotions. Avec un aspect inchoatif, l'un des plus courants et nettement le plus ancien était *encender* 'allumer' (57, 58), qui a été sélectionné par tous les noms prédicatifs de la classe <colère>, mais pas à la même date. S'il apparaît avec *ira* et *saña* très tôt (depuis le XIII<sup>ème</sup> siècle) – et abondamment –, il faudra attendre jusqu'au XIV<sup>ème</sup> siècle pour qu'il s'intègre au paradigme du nom *rabia* et jusqu'au XV<sup>ème</sup> siècle dans le cas des

<sup>25</sup> Et non seulement des langues indoeuropéennes, mais aussi des langues aussi éloignées que le chinois, par exemple, cf. Yu (1995).

noms *enojo* et *furia*. Quant à *furor*, il existe, certes, un exemple au XIII<sup>ème</sup> siècle (dans le *Libro de Alexandre*), mais les autres occurrences n'apparaissent qu'aux XIV<sup>ème</sup> et XV<sup>ème</sup> siècles. Il est fort probable que cette distribution chronologique ait un rapport avec l'extension dans la langue médiévale de chacun de ces substantifs. Le verbe *encender*, en tout cas, était très souvent employé sous sa forme participiale à valeur résultative (59).

- (57) ...*assí es el alma, quando se enciende de la ira, que non puede llegar a ella ninguna predicación nin ge la puede amatar* (a1250, *Bocados de oro*, ed. Mechthild Crombach).
- (58) *La qual cosa commo lo oyo el rrey, doliose del amigo perdido, e ençendiose mas en saña contra los monges* (c1400, *Barlaam e Josafat*, ed. J.E. Keller/R.W.Linker).
- (59) *Et ellos trebejando fue el mj njeto todo ençendido en yra* (a1284 Alfonso X, *General Estoria*, ed. Pedro Sánchez Prieto).

Son synonyme *inflamar*, mot savant du XIII<sup>ème</sup> siècle, fera partie des supports des éléments nominaux de cette classe à partir du XIV<sup>ème</sup> siècle, très souvent aussi sous sa forme participiale à valeur résultative (60), à l'exception des substantifs *enojo* et *rabia*, qui n'ont laissé aucune trace de cette sélection<sup>26</sup>. C'est surtout le nom prédicatif *ira* qui sélectionnait amplement ce verbe support, probablement entraîné par l'extension de plus en plus large d'*encender* :

- (60) *el establida del lugar en periglo lelio mauilio Emperador de los Romanos inflamado por grant yra leuo una huest enta aquella part* (1385 Juan Fernández de Heredia, *Gran crónica de España*, ed. Regina af Geijerstam)

Le verbe intensif à aspect duratif *arder* 'brûler' était particulièrement approprié au nom *ira*, qui confirme ainsi son caractère de mot vedette de la classe <colère>. Le nombre d'exemples où il apparaît accompagné de ce substantif dépasse largement le nombre d'exemples où il est sélectionné par *saña*, *rabia* ou *furor*. Et d'ailleurs, en ce qui concerne ce dernier nom prédicatif, il convient de signaler que toutes ses occurrences portent la signification étymologique 'délire furieux' dans un contexte purement guerrier, ce qui nous empêche de le prendre vraiment en considération. Si nous tenons compte, par surcroît, du fait que les autres substantifs de la classe <colère> n'ont laissé aucune trace de sélection de

<sup>26</sup> Nous faisons référence ici aux noms prédicatifs les plus courants, car nous avons déjà spécifié les limitations générales que présentent les noms les moins habituels *cordojo* et *corroço*.

ce verbe métaphorique, la pertinence du nom *ira* pour l'évolution du paradigme des verbes supports devient plus flagrante. Malgré son ancienneté en tant que forme lexicale (car nous le trouvons aux origines de la langue, cf. Menéndez Pidal, 2004, s.v. *ardere*), les premiers témoignages de son rôle comme verbe support remontent au XIV<sup>ème</sup> siècle (61, 62), quand la métaphore conceptuelle s'était installée dans l'usage collectif à un plus haut degré.

- (61) *Nestor la hora, entendidas las paraulas de Anthenor et ardiendo todo en ira, se torno descolorido...* (1376–1396 Juan Fernández de Heredia, *Historia troyana*, ed. Juan Manuel Cacho Blecua).
- (62) *Tolelde lo que ha, ardera bivo en sanna. E entonçe fue el peccado a las bestias del rrico omne e matol una grand pieça dellas* (1313–1469, *La estoria de Merlin*, ed. Karl Pietsch).

### 3.1.2. Autour de la métaphore du corps en tant que réceptacle des émotions

Nous avons vu qu'une deuxième métaphore est basée sur le schéma du réceptacle (LE CORPS EST UN RÉCEPTACLE DES ÉMOTIONS), ce qui a un rapport avec les humeurs diverses du corps humain. La colère, comme nous l'avons déjà rappelé pour le français au point 2.1.2, peut être perçue comme une sorte de fluide ou substance qui s'empare de quelqu'un en le « remplissant ». Cela explique, donc, qu'en espagnol aussi le verbe *llenar* 'remplir' soit employé depuis le XIV<sup>ème</sup> siècle comme verbe support de cette classe de prédicats, bien qu'exclusivement sous sa forme participiale *lleno (de)* 'rempli de'. Cette forme, isolée, prenait une valeur adjectivale résultative (62 ci-dessous), mais accompagnée du verbe *ser* 'être', permettait de conjuguer au passé le verbe *llenar* (63) dans son interprétation purement ergative<sup>27</sup>. Le fait que le corpus n'ait pas retenu des exemples avec d'autres formes conjuguées ne veut pas dire qu'elles n'aient pas été utilisées au Moyen Âge, bien évidemment, mais simplement qu'elles étaient, sans doute, beaucoup moins usitées dans les textes, rédigés assez souvent au passé.

- (63) *E como Appius, lleno de yra e de saña, oviessse dichas estas cosas, la muchedunbre de la gente de sí mesma se aredró de allí, assí que la donzella fincó toda sola...* (c1400 Pero López de Ayala, *Traducción de las Décadas de Tito Livio*, ed. Curt J. Wittlin)
- (64) *E commo el rrey vio esto fue lleno de grant ira e mando que fuese alla tristan* (c1313–c1410, *Cuento de don Tristán de Leonís*. Ed. Ivy A. Corfis)

<sup>27</sup> Il s'agit d'un verbe à alternance causative/ergative.



D'autres verbes étaient beaucoup moins fréquents. Les corpus nous offrent seulement deux exemples (au XV<sup>ème</sup> siècle) de la sélection par le nom *ira* du verbe *hinchar*, dont l'origine est pourtant très ancienne. Quant à *henchir* (forme patrimoniale du XIII<sup>ème</sup> siècle), il apparaît une seule fois comme verbe support sélectionné par le syntagme *ira y furor* dans un texte de la fin du XV<sup>ème</sup> siècle : le *Vocabulario eclesiástico* de Rodrigo Fernández de Santaella (« Indignor. gnaris. tus sum. turbarse o henchirse de yra y furor »). C'était le début d'un usage linguistique beaucoup plus répandu au Siècle d'Or. Nous pouvons affirmer, par conséquent, qu'au Moyen Âge, malgré l'importance qu'a acquise le verbe *llenar* comme support intensif de la classe <colère>, la métaphore LE CORPS EST UN RÉCEPTACLE DES ÉMOTIONS se trouvait encore en processus de formation.

#### 4. Conclusions

La notion de collocation en phraséologie met en jeu une base (ou collocateur) qui sélectionne un ou plusieurs collocatifs. Dans la mesure où une collocation est un phrasème, cette sélection ne saurait être pilotée par la sémantique (puisque l'on serait alors dans une relation prédicat-argument non phraséologique). Il doit donc s'agir, par hypothèse de principe, d'une sélection de type lexical<sup>28</sup>. Ainsi, par exemple, en français contemporain les adverbes *gravement* et *grièvement* sont synonymes. Mais, si aussi bien *gravement blessé* que *grièvement blessé* sont possibles, seulement *gravement malade* semble pleinement acceptable<sup>29</sup>. Ce sont, en effet, des prédicats comportant l'idée de 'blessure' (*atteindre, brûler, mutiler...*) qui sélectionnent *grièvement*.

Or, en observant les supports intensifs en ancien français, force est de constater qu'une grande variété d'intensificateurs est possible qui s'organise sémantiquement autour d'une métaphore. Il n'y a donc pas de sélection lexicale stricte, même si certaines tendances quantitatives se dessinent. Ce sont ces métaphores qui orientent la sélection des formules intensives (p. ex., la métaphore du feu est

<sup>28</sup> C'est ce type de sélection qui permet de formaliser les collocations sous forme de fonction  $f(x) = y$  où  $x$  et  $y$  correspondent respectivement à la base et au(x) collocatif(s) (Mel'čuk & Polguère, 2021 : 7).

<sup>29</sup> Il est toutefois possible de trouver des occurrences de *grièvement malade*, mais elles sont beaucoup plus rares que celles de *gravement malade* (une seule occurrence de la première combinaison dans le sous-corpus contemporain de Frantext face à 37 de la seconde). En revanche, *gravement blessé* présente une fréquence semblable à *grièvement blessé* (22 occurrences face à 24).

très productive avec la classe <colère>, tandis que la métaphore de l'évanouissement l'est avec la classe <chagrin>).

D'ailleurs, le fait que l'ancien espagnol présente *grosso modo* les mêmes métaphores intensives de base (et cette observation pourrait, sans doute, être étendue à d'autres langues médiévales<sup>30</sup>), liées à des domaines conceptuels partagés, semble montrer qu'il existe bel et bien une combinatoire restreinte dont l'origine (sinon le point d'arrivée lui-même qui est, par nature, évolutif) est non seulement sémantique mais aussi culturel. Cependant, nous observons deux différences importantes entre le français et l'espagnol. La première a un rapport à la chronologie, car si une bonne partie de ces verbes supports intensifs sont communs aux deux langues (*arder/brûler, inflamar/enflammer, llenar/remplir...*), les exemples des textes français sont toujours plus anciens. L'explication se trouve dans l'histoire de la langue écrite elle-même. En effet, nous savons que les premiers témoignages de l'espagnol sont plus tardifs que ceux du français et que les textes rédigés avant le XIII<sup>ème</sup> siècle dans cette dernière langue sont plus nombreux. La deuxième différence a trait à la variation. Les verbes supports métaphoriques du français médiéval dépassent plus fréquemment les deux réseaux de métaphores de base (LA COLÈRE EST DU FEU et LE CORPS EST UN RÉCEPTACLE DES ÉMOTIONS) qui ont fourni les principales unités verbales caractéristiques de l'espagnol de ce temps. En ce sens, il convient de ne pas sous-estimer l'idiosyncrasie des locuteurs de chaque langue, qui peuvent créer leur propres métaphores au delà des cadres culturels partagés. À tel point que certains réseaux métaphoriques caractéristiques des textes de l'ancien français (*mourir de, rougir de*, par exemple) ne feront jamais partie de la langue espagnole. D'autres existent bel et bien en espagnol, mais nous ne les trouvons dans les textes qu'un peu plus tard et parfois même beaucoup plus tard. Il suffira de citer – sans vouloir pour autant être exhaustifs – *hervir de* 'bouillir de', que nous trouvons depuis le XVI<sup>ème</sup> siècle et qui est conçu comme une métonymie à partir de la métaphore *hervir*, N1 : *colère* N2 : *en hum*, et *temblar de* 'trembler de' dont les premiers témoignages remontent au XIX<sup>ème</sup> siècle.

Il est également important de signaler que, dans nombre de cas, il existe un vague sémantique. En effet, s'il est physiquement impossible de 'bouillir d'ire' et, dans ce cas, nous sommes bel et bien devant une expression imagée, il est pourtant parfaitement possible de 'rougir d'ire' ou de 'trembler d'ire'. Or, est-ce que

<sup>30</sup> Cf., p. ex. la métaphore 'allumer' / 'éteindre' appliquée à *furor* en catalan médiéval, même si les exemples datent déjà du XV<sup>ème</sup> siècle (exemples extraits du corpus CICA) : *la tua furor encesa* (*Curial e Güelfa*, p. 335) ; *e per esta manera los satisfeu e appagà lur furor* (*Dotzè del Crestià*, 1<sup>a</sup> part, vol. I, p. 141).

toutes les occurrences de ces combinaisons impliquent vraiment l'existence effective des manifestations physiques correspondantes ? Rien de moins certain.

On peut observer que *tressuer d'ire* est une combinaison où une certaine manifestation physique véhicule l'idée d'une intensité de l'émotion. La réitération de cette combinaison nous fait penser qu'il s'agit plutôt d'un intensificateur que d'une description factuelle. Mais, à côté d'une douzaine d'exemples de cette suite (et de sa variante *suer d'ire*), nous trouvons l'exemple suivant où la solution formelle retenue est loin d'être collocationnelle :

(65) *D'ire et de maltalent ot la chiere sullente (Buevon de Conmarchis, 1271, p. 82)*

Une autre manifestation d'une ire intense est la perte de la capacité de parler. Mais l'expression de cette manifestation est fort variée :

(66) *D'ire la parole a perdue (Roman d'Alexandre, branche 3, 1180, p. 146)*

(67) *Tant d'ire ai Que ne sai, las! que dire (Miracles de Notre Dame, t. 3, 1218, p. 251)*

(68) *D'ire ne puet parler n'un seul mot (Miracles de Notre Dame, t. 3, 1218, p. 369)*

(69) *D'ire respondre ne vos poi (Perceval, 1181, p. 364d)*

En ce qui concerne l'espagnol, il vaut la peine de mentionner un curieux hapax qui montre l'emploi du verbe *emborracharse* ('se saouler') dans le rôle de support intensif. Il s'agit, bien évidemment, d'un usage individuel et littéraire. Mais il révèle une métaphore sous-jacente, en cours de construction, qui deviendra plus visible à partir du XIX<sup>ème</sup> siècle avec le syntagme *ebrio de ira* 'ivre de rage'. Il ne dépassera pas le domaine purement littéraire, et pour cela même, il pourrait être considéré un exemple intéressant de la polyvalence de la métaphore :

(70) *...y tanto se emborracho de ira contra ella: que delos varones: no dexo solo vno / que no le matasse (1494, De las mujeres ilustres en romance, ed. Harriet Goldberg)*

Nous sommes donc face à un large ensemble de possibles solutions dont l'usage semble opérer progressivement un certain tri. Les conditions et les résultats de ce processus devront faire l'objet de recherches ultérieures.

## Références citées

- Blanco, X. à paraître. Le sang, le feu et la rose. La couleur rouge comme *tertium comparationis* en français médiéval. Dans G. Gross, F. Neveu & M. Fasciolo (éds), *Décrire une langue : objectifs et méthodes*. Classiques Garnier.
- Blanco, X. & Buvet, P.-A. (2004). Verbes supports et significations grammaticales. *Linguisticae Investigationes* 27(2), 327–342.
- Blanco, X. (2018). La traduction des verbes supports de l'ancien français. *Le Français moderne* 86(1), 3–54.
- García Pérez, R. (2007). ¿Qué hacíamos y qué hacemos? *El verbo hacer en la historia del español*. Cilengua.
- Gavriilidou, Z. (2004). Verbes supports et intensité en grec moderne. *Linguisticae Investigationes* 27(2), 295–308.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses Universitaires du Septentrion.
- Kleiber, G. (1978). *Le mot « ire » en ancien français (XII<sup>e</sup>–XIII<sup>e</sup> siècles). Essai d'analyse sémantique*. Klincksieck.
- Kövecses, Z. (1990). *Emotion Concepts*. Springer.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press.
- Lakoff, G. & Kövecses, Z. (1987). The cognitive model of anger inherent in American English. Dans D. Holland & N. Quinn (éds), *Cultural Models in Language and Thought*. New York (195–221). Cambridge University Press.
- Melčuk, I. & Polguère, A. (2021). Les fonctions lexicales dernier cri. Dans S. Maren-go (éds), *La Théorie Sens-Texte. Concepts-clés et applications (75–155)*. L'Harmattan.
- Melčuk, I. (2005). Verbes supports sans peine. *Linguisticae Investigationes* 27(2), 203–217.
- Yu, N. (1995). Metaphorical Expressions of Anger and Happiness in English and Chinese. *Metaphor and symbol* 10(2), 59–92.

## Bases textuelles et dictionnaires

AND : *The Anglo-Norman Dictionary*, AND2 Online edition, <http://www.anglo-norman.net>, consulté le 25.04.2023.

CICA : *Corpus informatitzat del català antic*, Joan Torruella, Manel Pérez Saldanya, Josep Martines (éds), (2010), <http://www.cica.cat>, consulté le 16.05.2023.

DMF: *Dictionnaire du Moyen Français (1330–1500)*, ATILF – CNRS & Université de Lorraine, <http://www.atilf.fr/dmf>, consulté le 25.04.2023.

Frantext: *Frantext*, ATILF – CNRS & Université de Lorraine, <https://www.frantext.fr>, consulté le 25.04.2023.


Menéndez Pidal, R. (2003). *Léxico hispánico primitivo (siglos VIII al XII)*. Espasa-Calpe. Real Academia Española. *Banco de datos (CORDE) [en línea]. Corpus diacrónico del español*, <http://www.rae.es>, consulté le 02.05.2023.

Real Academia Española (2013). *Corpus del Diccionario histórico de la lengua española (CDH)*, <https://apps.rae.es/CNDHE>, consulté le 02.05.2023.



**Peter Blumenthal**

Universität de Cologne  
Allemagne

 <https://orcid.org/0000-0002-4196-3677>

**Giovanni Rovere**

Universität de Heidelberg  
Allemagne

 <https://orcid.org/0009-0006-5942-8409>

# Valence des verbes de perception en français et italien : perspective diachronique

**The valency of perception verbs in French and Italian: A diachronic perspective**

## Abstract

This paper explores perception verbs (like ‘hear’ and ‘see’) from a construction grammar perspective, with a focus on both contrastive and diachronic aspects. Surprisingly, our corpus analysis suggests that different verbs from this semantic groups do not display a homogeneous syntactic behaviour. For example the verb ‘observer’ stands out from other verbs in this group. We attempt to account for the special status of ‘observer’ and discuss the extent to which the acceptability of sentences featuring ‘observer’ can be improved by acting on certain parameters in the sentence. Our research highlights the role of a factor referred to as *evidentia* in classical rhetoric.

## Keywords

Dependency grammar, perception verbs, deixis, evidence, history of French, history of Italian, contrastive linguistics

## 1. Petit tour d’horizon méthodologique

Nous avons pour objectif de décrire, d’expliquer en diachronie et de comparer, dans l’esprit contrastif du modèle de grammaire qui remonte à L. Tesnière, quelques constructions basées sur des propositions infinitives

(exemple : *Je le vois travailler*). Pour caractériser en une phrase cette théorie, qui renoue avec la logique du début du 20<sup>e</sup> siècle (G. Frege, E. Husserl, cf. G. Gross, 2012 : 14) et dont les notions-clés sont la valence, la saturation et la dépendance, on retiendra surtout les prémisses du modèle : au centre structurel de la proposition se trouve le verbe avec sa valence, accompagné de compléments qui se distinguent selon leur degré de nécessité (« actants » ou « circonstants »). Pour parer à l'une des faiblesses évidentes du modèle, l'insuffisance de la réflexion sémantique sur la conception des actants, la linguistique (surtout) américaine a développé, à partir des années 1970, la théorie des « rôles thématiques » (fonctions sémantiques des actants ; cf. Evans & Green, 2006 : 666–706) ; ces derniers, intégrés plus tard aux grammaires de construction, se superposent aux relations entre les arguments (actants) d'un prédicat (verbe), p. ex. celui de l'agent, du patient, de l'expérimenteur, du but. D'autres facteurs encore déterminent l'organisation des mots composant la phrase, à savoir les liens créés par des combinaisons consacrées par l'usage. C'est à Gaston Gross (1996) et à quelques autres chercheurs que nous devons une classification cohérente de ces unités souvent figées. À partir de cet état de la recherche nous voudrions mettre en relief l'intérêt de nouvelles approches des problèmes de la combinatoire des mots ; problèmes rencontrés en général au cours de nos travaux lexicographiques<sup>1</sup> sur l'analyse de 2100 verbes italiens, leurs environ 6000 schémas syntaxiques et plus de 90000 attestations (cf. WiV, 2023). La solution des difficultés relevées rend nécessaire la définition de nouveaux critères d'analyse, rappelant parfois des concepts de la grammaire de construction (ci-dessous 3.4, 5.1, 5.2, 6.)

Dans notre approche comparative centrée sur la notion de valence, nous nous intéresserons à deux langues romanes, le français et l'italien, et incidemment à l'allemand, pour dégager entre elles certaines affinités logiques ou de fines différences sémantiques. Deux exigences méthodologiques particulières se feront sentir : quelques-uns des problèmes synchroniques traités requièrent, pour une analyse satisfaisante, un cadre diachronique ; en outre, il s'avèrera indispensable d'adosser à des considérations paradigmatiques les études faites dans le prolongement des analyses combinatoires. Ces nécessités nous amèneront à renouer avec des concepts fondamentaux du structuralisme classique, vue dans une perspective moderne.

---

<sup>1</sup> Surtout lors de la préparation de l'édition suivante.

## 2. Étude de cas : *observer*, *osservare*

Afin de délimiter clairement notre champ d'investigation, nous retiendrons, à titre d'exemple et en guise de poisson-pilote, une construction du verbe *observer* et ses équivalents dans les autres langues (italien *osservare*, allemand *beobachten*). Certains des « schémas syntaxiques »<sup>2</sup> de ces verbes se ressemblent d'une langue à l'autre, cf. :

*osservo che/come...;*

*j'observe que/comment...;*

*ich beobachte, dass/wie...*

Dans d'autres contextes, le comportement valenciens des verbes diverge selon les langues, p. ex. face à la construction infinitive. Ainsi, il n'existe pas de doute sur l'agrammaticalité, dans tous les registres de la langue allemande, de la construction \**Ich beobachte Paul arbeiten* (« J'observe Paul travailler. »)<sup>3</sup>, alors que la version française proposée, par ailleurs fréquente sur Internet (consulté le 02.05.2023), semble susciter moins de critiques auprès des locuteurs natifs consultés. On obtient une formulation parfaitement conforme au bon usage en remplaçant *observer/beobachten* par *voir/sehen* : *Je le vois travailler. / Ich sehe ihn arbeiten*. Quand on remplace le verbe à l'infinitif (*travailler/arbeiten*) par un verbe d'aspect lexical ponctuel, comme « entrer », rien ne change en allemand (\**Ich beobachte ihn eintreten*). En français, ces verbes sont nettement plus rares après *observer* que ceux qui expriment une activité continue (comme *travailler, manger, dormir*), selon les données d'Internet. Et pourtant, dans notre corpus littéraire, la combinaison de *observer* avec un verbe ponctuel est bien attestée :

(1) *Je l'observe entrer.* (F. Anouk, 2020, Frantext).

En italien, le problème se pose, du moins à première vue, dans des termes semblables au français. Cela concerne d'abord, dans notre domaine, la grammaticalité

<sup>2</sup> *Satzbaupläne* en allemand. G. Gross (2012 : 11) parle de « schémas prédicatifs ». À vérifier à l'aide des dictionnaires des valences verbales (p. ex. Busse & Dubost, 1977 ; WiV, Engel & Schumacher, 1978) ainsi que des tables du lexique-grammaire (cf. M. Gross (1975).

<sup>3</sup> Cf. DUDEN (1998 : 701, 776) ; Heringer (1996 : 229) ; Engel & Schumacher (1978, s.v. *beobachten*).



parfaite de *Lo vedo lavorare/entrare*, ensuite, le caractère quelque peu étrange de *Lòsservo entrare*. Mais dès que l'on affine l'analyse en se basant sur l'emploi réel, et non plus sur des exemples forgés pour les besoins de la cause, les particularités de chacune des deux langues romanes se dessinent avec davantage de netteté. Et pourtant, dès l'entrée de jeu, il importe de souligner les limites de cette analyse, dues à l'influence de facteurs « stylistiques » au sens large du mot. En effet, les emplois de constructions infinitives en tant que parties de la valence des verbes régissants dépendent fortement, et de manière différente selon les langues, du type de texte, du niveau stylistique et des données diachroniques. Dans ces conditions, l'étude contrastive ne saurait aboutir à des conclusions simples et facilement généralisables. Il s'agira de saisir les nuances qui renvoient aux différences entre les normes en vigueur dans les langues comparées, tout en tenant compte des choix individuels des auteurs et de divers types de variétés. Cette grande prudence de la part des auteurs de ces lignes se justifie par l'état de la recherche, mais plus précisément aussi par le fait que, à part WiV 2023, les constructions en question ne sont mentionnées par aucun des dictionnaires consultés (TLFi, Busse & Dubost, 1977 ; Riegel *et al.*, 2009 ; Petit Robert, etc. ; GDLI, GRADIT, Sabatini/Coletti, etc.)<sup>4</sup>.

### 3. Construction infinitive en italien

Il n'est pas étonnant que dans une telle situation linguistique, des voix discordantes se manifestent. Ainsi, Skytte et Salvi (1991 : 512), dans le chapitre sur la proposition infinitive, présentent l'exemple suivant (construit) sans aborder la question de son acceptabilité :

(2) *Osservai Piero uscire.*

C'est l'ensemble des réactions à la fois peu favorables<sup>5</sup> et étrangement contradictoires à (2) de la part de locuteurs natifs de l'italien qui nous a amenés à pour-

<sup>4</sup> Cimaglia (2011 : 1087) mentionne sans commentaire *osservare* dans son énumération des verbes de perception ("Tra i più comuni verbi di percezione sono: *vedere, guardare, scorgere, notare, osservare, mirare, sentire, ascoltare, udire, intendere, avvertire, percepire, ecc.*").

<sup>5</sup> Malaise largement dû à l'absence d'un contexte adéquat, susceptible de valoriser l'objet de l'observation (à présenter comme complexe). À ce propos, la réflexion judicieuse faite par l'une des personnes interviewées en dit long ; elle nous a fait remarquer qu'avec une suite du genre « dall'ac-

suivre la recherche sur une autre base empirique, par l'exploitation de corpus, surtout journalistiques, utilisés déjà pour l'élaboration du WiV (dans son édition sur papier de 1998). Les résultats de cette enquête sont clairs : oui, la construction infinitive de type « sujet – verbe de perception – complément d'objet direct – infinitif » existe en italien même après *osservare* ; dans la grande majorité des cas repérés dans le corpus, *osservare* s'emploie alors à l'imparfait et désigne un processus non accompli, duratif comme l'action d'observer. Plus rarement, le présent assume le même rôle aspectuel. Selon ces critères, la structure de (2) est « atypique. » – notion qu'il convient de préciser immédiatement pour orienter l'attente des lecteurs. Serait typique un contexte ou une structure qui réaliserait l'*evidentia* au sens de la rhétorique classique, non pas celle, abstraite, de la vérité, mais « l'évidence du monde » (Grossmann, 2017), concrète et devant les yeux<sup>6</sup>. Nous concevons cette évidence comme apte à constituer une sorte de seuil grammatical : au-dessous d'un certain minimum d'évidence (et/ou de complexité, ci-dessous 3.1) présent dans la phrase, la construction avec *osservare* paraît étrange.

Dans la suite, nous réunirons quelques informations sur la genèse et les contextes syntaxiques de *observer/osservare* et nous nous demanderons dans quelle mesure les résultats de l'enquête pourraient encourager à réviser certains aspects du modèle de base valenciel.

### 3.1. Rétroaction

En plus de la double expression de la durée (ci-dessus 3.), la construction discutée ici contient souvent une autre caractéristique encore, moins évidente : la proposition à l'infinitif désigne une réalité dont émane une sorte de rétroaction sur le référent du sujet de *osservare*, lequel se trouve affecté dans sa vie psychique ou son activité sociale, comme le montrent les citations suivantes :

- (3) *Stoccolma osserva preoccupata l'Occidente non imparare le lezioni dell'Ucraina e raccontarsi favole sul regime putiniano. Troppo vicina all'Orso per cedere all'ideologia, gli legge in faccia gli imperativi strategici.*  
(Limes)

---

qua come Venere dal mare », la phrase ne ferait pas de problème. En conclusion, les causes du statut de (2) se situent principalement au niveau de la « parole ». Toutefois, une cause systémique pourrait tenir à l'existence d'une formule plus courante (*Osservai Piero che usciva...*) ; cf. Jeppesen Kragh & Strudsholm (2018 :110) sur la « relativa deittica ».

<sup>6</sup> Donc, notionnellement proche de la deixis. Cf. un exemple dans la note précédente.

- (4) *“L’ho costretto ad uscire allo scoperto”, diceva venerdì pomeriggio ai suoi ministri che, ammirati, osservavano il premier disegnare con le mani un futuro certo, e tratteggiarlo fin nei dettagli più minuscoli.* (Repubblica)

Les données observées devenant rétroactives sont, dans (3), la naïveté politique des Occidentaux et, dans (4), les capacités rhétoriques du premier ministre. Dans les deux exemples, l’indication des effets psychologiques (*preoccupata, ammirati*) est faite avant que leurs causes (les données rétroactives) soient précisées. L’observation conduit, dans (3), « Stockholm » à assumer un rôle actif: *legge in faccia...* Le sujet de cette proposition correspond à celui de la principale (« Stockholm »). L’exemple (3) montre clairement comment l’observateur, avec son rôle dynamique, devient agent, également dynamique, et par conséquent facteur de la complexité de la phrase.

### 3.2. Valence thématique

Nous appellerons « structure thématique » d’une phrase ou d’une suite de propositions ce jeu des rôles sémantiques successifs mis en musique par la « valence thématique » du verbe, conformément à ce que nous avons montré à propos de (3) et (4). Selon notre hypothèse, la valence thématique d’un verbe, dont l’analyse nécessite toujours la présence de contexte, fait partie de sa « valence » au sens large, laquelle comprend en outre sa valence syntaxique. Insistons dès maintenant sur le fait que les verbes du même champ sémantique et dotés de la même valence syntaxique (p. ex. *osservare* et *sentire*) peuvent correspondre à différentes structures thématiques. On peut illustrer ce principe par les verbes de perception en construction infinitive. C’est que la structure thématique de (2) avec un sujet agent diverge de celle de (5) avec un sujet expérimenteur :

- (5) *Maria l’ha sentito suonare il pianoforte.*

Skytte et Salvi (1991 : 510) la paraphrasent avec des termes qui rendent explicites les rôles sémantiques de ses constituants : « Maria è stata sottoposta per un certo tempo a una determinata sensazione (il suono del pianoforte causato da Piero) ». Selon la grammaire citée, la suite textuelle convenant à (5) pourrait être une phrase du type suivant :

- (6) *e le è piaciuto molto il suo stile.*

*Maria*, reprise dans (6) par le, se maintient dans le rôle d' « expérienceur » plus ou moins passif. La différence entre les structures thématiques de (2) et de (5) s'explique donc par le décalage entre les puissances agentives des deux verbes *sentire* et *osservare*. Notons en marge de cette réflexion que l'impact de cette puissance, dû à l'effort d'attention impliqué par le verbe, est tel que des objets d'observation comme la sortie de Piero dans (2), relativement banals, ne paraissent pas forcément à la hauteur de l'intensité qu'exprime la valence thématique. Cette considération pourrait expliquer le statut précaire de (2).

### 3.3. « *Osservare* + proposition infinitive » vs « *osservare* + complétive »

Les notions complémentaires de valence syntaxique et de valence thématique permettent non seulement de contraster systématiquement des verbes synonymes, comme *sentire/vedere* et *osservare*, mais aussi d'évaluer plusieurs constructions du même verbe, p. ex. *osservare* + construction infinitive et *osservare* + *che* (complétive). Sous cet angle, Skytte et Salvi comparent

(7) *Maria ha sentito Piero suonare il pianoforte.*

à la phrase suivante :

(8) *Maria ha sentito che Piero suonava il pianoforte.*

Les divergences se ressentent, entre autres, au niveau des suites textuelles possibles et probables. Selon Skytte et Salvi (1991 : 510), (7) peut conduire à (6), ci-dessus, et (8) à la phrase suivante :

(9) ... *ed è scesa ad ascoltarlo.*

Des constructions diverses, éventuellement liées à certaines acceptions du même verbe, peuvent donc déboucher sur différentes structurations thématiques du contexte (cf. Gaatone, 1987).

### 3.4. *Osservare* + construction « accusative »

Pour illustrer les « interdépendances » (cf. Stefanowitsch & Gries, 2003) entre valence et significations (du verbe et des différentes constructions), rien de plus

instructif que d'étendre l'analyse de *osservare* à une acception que nous n'avons pas encore discutée ici, celle d'un verbe de déclaration ('faire remarquer que'). Dans un style archaïsant ou formel, surtout juridique, *osservare* déclaratif peut s'utiliser dans une autre construction infinitive que celles étudiées ci-dessus ; exemple :

- (10) *Osservava quest'obbligo essere di tutto rigore in quanto che il 6. Articolo normale dispone [...].* (L'Eco dei tribunali. Giornale di Giurisprudenza Civile IV/229 (1855), p. 161)

Il est clair que cet « infinito con soggetto espresso » (Skytte & Salvi, 1991 : 527)<sup>7</sup>, d'inspiration latine savante, se distingue de l'autre construction infinitive, où celle-ci peut passer pour une expansion du complément direct. Ainsi, dans (4), « *osservano il premier* » constitue le noyau du prédicat complet. Et pourtant, on hésite parfois entre les deux types de constructions infinitives, dont chacun est lié à une certaine interprétation de *osservare*, verbe de déclaration ou de perception. C'est le cas de (3), où ce verbe peut se traduire par « constater » ou par « voir » – chaque interprétation étant liée à un certain type d'infinitive.

### 3.5. Problèmes de diachronie

Notre analyse soulève incidemment des problèmes de diachronie, entre autres à propos de l'âge des phénomènes discutés ci-dessus. Limitons-nous ici à un détail intéressant : les débuts de la construction infinitive avec *osservare*. Quelques sondages nous donnent à penser qu'elle doit être bien rare encore au 18<sup>e</sup> siècle. Pour cette période, nous avons trouvé l'attestation suivante, à vrai dire un peu particulière en raison de la forme du groupe sujet, cf. :

- (11) *Vero è, che si osservava il Re Sardo accrescere le sue truppe, e far altri preparamenti di guerra; ma il tutto veniva supposto tendere alla difesa propria, e dello Stato di Milano, caso mai che i Franzesi pensassero a qualche tentativo contro l'Italia.* (L. A. Muratori, *Annali d'Italia*. T. XII/I, Roma, 1754, p. 247)

<sup>7</sup> Cf. en ce sens Egerland (2011: 658): « Nella lingua antica e letteraria, la costruzione percettiva [= verbo di percezione seguito dall'infinito] talvolta si fonde con l'accusativo con l'infinito: Se adunque si considererà tutti e' progressi del duca, si vedrà lui aversi fatti gran fondamenti alla futura potenza (Machiavelli, *Il Principe* VII, p. 32) ».

La structure thématique de (11) ne correspond qu'imparfaitement au modèle sous-jacent à (4). L'une des différences tient à la faiblesse agentive du sujet de la principale (*si*). Toujours est-il que la deuxième phrase pourrait servir d'allusion à une « rétroaction » (*difesa propria*).

### 3.6. Résumé

Ces quelques réflexions sur certains emplois de *osservare* permettent déjà d'entrevoir les étapes par lesquelles passe la description *valencielle* complète d'un verbe. Il s'agit pour l'essentiel des tâches suivantes :

1. description sémantique du verbe et distinction de ses acceptions (à reprendre ci-dessous, note 5) ;
2. description morphosyntaxique des différents schémas syntaxiques (constructions) correspondant aux acceptions ;
3. identification des schémas thématiques (rôles thématiques) attribuables à chaque construction ;
4. hypothèses sur la continuation textuelle la plus probable après chaque construction.

## 4. Brève histoire de l'infinitive après *observer* en français

Grâce aux possibilités de fouille de données qu'offre la base Frantext, il paraît plus facile en français qu'en italien de faire des hypothèses contrôlables sur l'aspect diachronique de notre problème<sup>8</sup>, et cela notamment pour l'histoire récente. Grand inconvénient : il s'agit de textes pour la plupart littéraires, que l'on ne saurait prendre, sans consultation d'autres corpus, pour représentatifs de la langue de leur époque. Il apparaît toutefois hautement probable que la possibilité d'emploi d'une infinitive après *osservare/observer* est plus récente en français qu'en italien. Voici la première attestation trouvée dans Frantext :

---

<sup>8</sup> L'italien serait-il, sur ce point aussi, plus conservateur que le français ? Nous n'avons pas encore de réponse à cette question. Toujours est-il que, en ce qui concerne la construction infinitive après les verbes de perception, Egerland & Cenammo (2010 : 846–856) relèvent peu de différences entre l'ancienne langue et l'italien moderne ; cf. Cimaglia (2011 : 1087s).

- (12) *Mais toute considération d'avenir n'aboutissait qu'à me faire dépenser davantage. A ! Qu'aurais-je besoin de tant, une fois seul ! ... pensais-je et j'observais, plein d'angoisse et d'attente, diminuer, plus vite encore que ma fortune, la frêle vie de Marceline* (A. Gide, 1902, Frantext).

La citation contient deux caractéristiques (liées à « l'évidence ») dont la fréquence s'avèrera significative dans les constructions de ce type : la gradualité<sup>9</sup> du phénomène observé (« diminuer ») et la rétroaction (« plein d'angoisse »), à la même place qu'en (3). De là jusqu'aux années 1980, Frantext renferme à peine une dizaine d'exemples de *observer* + proposition infinitive, dont deux phrases de Raymond Queneau (1933, 1965), auteur réputé stylistiquement innovant. Puis, à la fin du 20<sup>e</sup> siècle, on assiste à une véritable explosion fréquentielle de la construction avec *observer*, qui se perpétue jusqu'en 2022. Au nombre des auteurs qui participent à cette vague, on trouve D. Cherigui, C. Laurens, S. Houine, I. Seyman, G. Bouillier, H. Guibert, J.-M. Lovay, A. Boudard, A. Djoulaït, B. Schreiber, F. Anouk. Une fois que l'intuition du locuteur (*Sprachgefühl*) s'est accommodée du changement de la combinatoire de *observer*, tout se passe comme si les vannes s'ouvraient et que les règles de l'emploi du verbe développaient leur propre dynamique. Concrètement, la combinaison « *observer* + proposition infinitive » a connu une phase d'expansion, où elle s'utilise de plus en plus en dehors des tendances dues à la structure thématique induite par *observer*. L'histoire de *observer* depuis une quarantaine d'années, du moins en langage littéraire, correspond largement à celle de l'effacement progressif de ses contraintes combinatoires – de facteurs donc qui relèvent davantage du système de la « langue » que de la « parole », comme en italien (ci-dessus note 3). Ainsi, dans la citation suivante, le contenu de l'observation n'est probablement pas incrémentiel, mais itératif (« baiser toute la journée ») :

- (13) *Sébastien fixe encore ses pieds, et son tee-shirt mal agencé. Il marche dans les pas du vieux jusqu'au comptoir, y pose son billet, ramasse la monnaie et repart avec la désagréable impression que le vieil épicier vient d'entrer dans sa vie, comme s'il l'avait **observé baiser** toute la sainte journée.* (Anouk, 2020, Frantext)

<sup>9</sup> Elle entraîne une interprétation non ponctuelle (ci-dessus 3., ci-dessous 5.) de *observer* qui fait attribuer cet emploi (Petit Robert, s.v., II.1.) aux verbes de perception (comme *voir*, admettant la même construction) et l'éloigne de l'acception ponctuelle du même verbe (Petit Robert, s.v., II.4.), synonyme de *constater*, verbe mental. Cette réflexion montre l'importance de l'aspect lexical de l'acception d'un verbe pour l'attribution de celui-ci à une classe lexicale de verbes et, par conséquent, pour ses possibilités de construction.

Cependant, d'autres éléments de la structure thématique décrite, comme p. ex. l'effet de la rétroaction (« la désagréable impression ») se maintiennent dans (13). La phase finale de ce changement linguistique consistant dans le fait que *observer* s'émancipe, dans un certain langage littéraire, des contraintes combinatoires antérieures se manifeste par des constructions du type *Je l'observe entrer* (v. (1)), qui semblent pourtant rester difficilement acceptables pour de très nombreux Français de la vieille génération, même dans la langue actuelle.

En résumé, *observer* a surmonté les obstacles qui l'empêchaient jusqu'à la fin du 19<sup>e</sup> siècle de régir une proposition infinitive. Le non-respect de cette règle (cf. (12)) avait probablement, à l'origine, un intérêt rhétorique en procurant à la phrase une plus-value sémantique et textuelle. Cette innovation est allée d'abord de pair avec une certaine structure thématique qui s'est effacée au cours des dernières décennies. Actuellement, l'emploi de *observer* suivi d'une proposition infinitive se banalise.

## 5. Pour une extension du domaine de la grammaire de dépendance

Nous nous sommes penchés sur deux (ou trois?) problèmes concrets : une construction (la proposition infinitive), un verbe (« observer ») – et, bien sûr, l'histoire de la rencontre entre les deux phénomènes. Le tout articulé sur la théorie de la valence, dont le fondateur, L. Tesnière, s'était laissé guider par une métaphore, le rapprochement entre chimie (d'où la notion de valence) et linguistique.

### 5.1. Dimension syntagmatique

Dans le sillage de notre poisson-pilote, nous resterons fidèles à la recherche d'une analogie, qu'au-delà de la chimie nous débusquerons du côté de l'écologie. L'aspect chimique : nous avons vu qu'une fois la réaction combinatoire obtenue entre telle construction (ex. : proposition infinitive) et tel verbe (ex. : *observer*), il s'ensuivra une nouvelle unité (cf. (1)), promise à une certaine usure de son impact rhétorique (ci-dessus 4.). Aspect écologique : le nouveau composé peut avoir des effets imprévus sur l'« environnement », c.-à-d. le contexte (cf. la « rétroaction », ci-dessus 3.1). Nous proposons d'intégrer à la grammaire



de dépendance l'étude de ce dernier aspect, qui augmenterait les capacités du modèle à tenir compte des interdépendances, et cela non seulement au niveau de l'axe syntagmatique.

## 5.2. Dimension paradigmatique

À côté de ces phénomènes d'ordre syntagmatique, une nouvelle grammaire de dépendance à dominante lexicale s'intéressera aussi aux structures paradigmatiques, donc aux choix auxquels doit procéder le locuteur. Les deux paradigmes<sup>10</sup> pertinents pour nous sont celui des acceptions des verbes de perception régissant et celui des types de propositions régies. Nous plaiderons ici pour un renforcement de la composante paradigmatique du modèle en partant des mots et des constructions discutés plus haut. Le locuteur peut avoir le choix, après certains verbes de perception, entre une proposition infinitive et une subordonnée complétive introduite par *que* (cf. ci-dessus 3.3), exemples :

(14) *Je vois Pierre arriver.*

(15) *Je vois que Pierre arrive.*

Intuitivement, on dira que (14) exprime toujours une perception, alors que (15), renvoie ou bien à une perception ou bien à un constat, acte mental (ci-dessus note 5). La différence porterait virtuellement sur le degré d'abstraction des deux messages. Dans ces conditions, *observer que* désigne le constat d'un « fait »<sup>11</sup>, acte ponctuel et en général conclusif ; cf. cet exemple dans Busse & Dubost, s.v. :

(16) *J'ai observé que les arbres commencent à fleurir.*

Le dictionnaire ne mentionne pas encore pour *observer* la construction infinitive comme dans (13), très rare à l'époque de la publication du livre. Cependant, la citation (3) ci-dessus présente une certaine réalité (un « état des choses ») en tant qu'objet de la perception, comme le prouve l'emploi de l'imparfait (non télique). Les deux langues romanes semblent donc appliquer le même principe, selon

<sup>10</sup> Les choix faits dans les deux paradigmes s'avèrent souvent interdépendants. Exemple : la sélection d'un type de proposition infinitive implique la sélection d'une acception verbale – et inversement.

<sup>11</sup> Selon Lalande (1972), sous FAIT « jugement d'affirmation sur la réalité extérieure ».

lequel, dans les constructions en question, la proposition infinitive tend vers l'expression d'un contenu concret et duratif, alors que la complétive va vers l'abstrait et le ponctuel, voire l'atemporel, comme le montrent si clairement les citations suivantes<sup>12</sup> :

(17) *Sentiva la lama penetrargli nella carne.* [concret]

(18) *Io vi stimo, perché sento che sapete cos'è la pittura.* [abstrait]

Toutefois, l'emploi de la proposition infinitive est lié à des contraintes lexicales, qui concernent non seulement les verbes de la principale (seule une douzaine de verbes fréquents comme *vedere* ou *sentire* s'utilise effectivement), mais aussi ceux à l'infinitif, qui désignent obligatoirement une activité concrète (comme *chanter*), alors que *sapere* serait exclu (\**vi sento sapere*, cf. (18)). Face à ces contraintes s'offre au locuteur désireux d'éviter les effets d'une abstraction jugée trop grande de la complétive un outil extrêmement pratique, consistant à rétablir le flux du temps dans la subordonnée : il s'agit du remplacement de la conjonction *che* par *come*, qui est alors « più o meno equivalente a *che* » (Wandruszka, 1991 : 471). Selon cette grammaire, « *che* può essere sostituito da *come* per rendere possibile una maniera di raccontare più icastica, meno distanziata, che faccia svolgere l'evento di cui riferisce, in un certo senso, sotto gli occhi del destinatario »<sup>13</sup> (*ibid.*). Cf. l'exemple suivant :

(19) *Hai notato come stia/sta aumentando il traffico anche sulle strade secondarie?* (WiV)

Après certains verbes, *che* présente le contenu de la subordonnée comme un « fait » objectif, *come* met l'accent sur sa dimension temporelle, évolutive et subjective – ce que l'on pourrait rapprocher de la « deixis à l'imaginaire » de K. Bühler (Verdier, 2009 : 153) ou de « l'évidence » en rhétorique. Cette opposition est-elle pertinente dans le cadre d'une grammaire valencielle de l'italien ? Oui, car elle permet de caractériser les verbes par leurs valences. En effet, il existe des verbes qui admettent *come*, mais pas *che* (p. ex. *descrivere*) et d'autres verbes qui prennent seulement *che* (*sapere che*, mais \**sapere come*<sup>14</sup>). On notera que *descrivere come* (WiV s.v.) n'équivaut pas structurellement à *descrivere comment* (Busse

<sup>12</sup> WiV, s.v. Cf. les constructions du même type pour *ascoltare*, *distinguere*, *vedere* (WiV, s.vv.).

<sup>13</sup> À comparer à la distinction que fait Kant entre intuition et concept (*Anschauung* et *Begriff*) ; cf. Lalande (1972, sous INTUITION).

<sup>14</sup> Sauf cas de *come* interrogatif (WiV, s.v.).

& Dubost, s.v. *décrire*), expression d'une interrogation indirecte). En revanche, l'allemand *wie* constitue souvent un excellent équivalent des constructions françaises et italiennes discutées plus haut. Cela vaut en particulier pour les infinitives après *osservare* (cf. (3)) et les verbes de perception proprement dits (cf. (17)) ainsi que pour *come* rival de *che* (cf. (19)).

## 6. Schémas prédicatifs et la possibilité d'une construction

Par cette contribution, nous espérons avoir montré que l'évolution des schémas prédicatifs et des valences verbales se prêtent à des études approfondies en matière de grammaire historique et comparée. Pour nous, il s'agissait ici plus particulièrement de voir dans quelle mesure et par quelles voies les verbes *observer*, *osservare* et *beobachten* ont élargi (ou non) leur valence grâce à l'annexion d'une certaine construction. Restées inchangées en allemand, les règles à la base de ce phénomène ont évolué dans les deux langues voisines, mais selon des principes et dans le cadre de variétés qui n'étaient pas toujours les mêmes d'une langue à l'autre<sup>15</sup>.

D'un point de vue méthodologique, certains des faits analysés ici donnent à penser que le modèle de Tesnière (dans sa phase actuelle) gagnerait à privilégier le principe d'interdépendance plutôt que celui de dépendance (cf. Stefanowitsch & Gries, 2003 : 236). L'étude du comportement de *observer/osservare*, verbe sémantiquement complexe, nécessitant des voisinages du même genre, nous a permis de thématiser plusieurs aspects, synchroniques ou diachroniques, de cette interdépendance.

## Références citées

- Busse, W. & Dubost, J.-P. (1977). *Französisches Verblexikon*. Klett-Cotta.  
 Cimaglia, R. (2011). Verbi di percezione. Dans R. Simone (éd.), *Enciclopedia dell'italiano* (1086–1088). Treccani.  
 DUDEN. (1998). *Die Grammatik*. Bibliographisches Institut.

<sup>15</sup> Mais comme le soulignent – sans s'en étonner – Jeppesen Kragh & Strudsholm (2018 : 121), les ressemblances l'emportent globalement.

- Egerland, V. (2011). Frasi infinitive. Dans R. Simone (éd.), *Enciclopedia dell'italiano* (656–660). Treccani.
- Egerland, V., & Cenammo, M. (2010). Frasi subordinate all'infinito. Dans G. Salvi & L. Renzi (éds), *Grammatica dell'italiano antico* (815–859). Il Mulino.
- Engel, U., & Schumacher, H. (1978). *Kleines Valenzwörterbuch deutscher Verben*. Gunter Narr.
- Evans, W., & Green, M. (2006). *Cognitive Linguistics. An Introduction*. Edinburgh University Press.
- Frantext. Nancy: ATILF, <http://www.frantext.fr>, consulté le 02.05.2023.
- Gaatoone, D. (1987). Rien de moins évident que le sens de *évident* : réflexions sur l'opposition entre infinitives et complétives et les rapports entre syntaxe et sémantique. *Revue québécoise de linguistique* 16(2), 291–303.
- GDLI = Battaglia, S. (éd.) (1961–2002) *Grande Dizionario della Lingua Italiana*. Utet.
- GRADIT = De Mauro, T. (éd.) (1999–2007). *Grande dizionario italiano dell'uso*. Utet.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Septentrion.
- Gross, M. (1975). *Méthodes en syntaxe. Le régime des constructions complétives*. Hermann.
- Grossmann, F. (2017). La notion d'évidence et son expression linguistique dans la rhétorique scientifique. *Travaux neuchâtelois de linguistique* 65, 37–52.
- Heringer, H. J. (1996). *Deutsche Syntax. Dependentiell*. Stauffenburg.
- Jeppesen Kragh, K. A. & Strudsholm, E. (2018). I verbi di percezione e le loro modalità di costruzione in una prospettiva diacronica ed interlinguistica francese-italiano. Dans P. Greco, C. Vecchia & R. Sornicola (éds), *Strutture e dinamismi della variazione e del cambiamento linguistico. Atti del Convegno DIA III, Napoli, 24–27 novembre 2014* (107–124). Giannini Editore.
- Lalande, A. (1972). *Vocabulaire technique et critique de la philosophie*. PUF.
- Neveu, F. (2011). *Dictionnaire des sciences du langage*. Armand Colin.
- Petit Robert = Rey, A. (éd.) (1993). *Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Le Robert.
- Riegel, M., Pellat, J.-C., & Rioul, R. (2009). *Grammaire méthodique du français*, PUF.
- Sabatini/Coletti = Sabatini, F. & Coletti, V. (2008). *Dizionario di italiano*. Sansoni.
- Skytte, G. & Salvi, G. (1991). Frasi subordinate all'infinito. Dans L. Renzi & G. Salvi (éds), *Grande grammatica italiana di consultazione, vol. II*, (483–569). Il Mulino.
- Stefanowitsch, A., & Gries, Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–243.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.
- Verdier, M. (2009). Deixis ad oculos et deixis à l'imaginaire dans les interactions médiatisées par ordinateur en malgache. *Verbum* 31, 153–176.
- Wandruszka, U. (1991). Frasi subordinate al congiuntivo. Dans L. Renzi & G. Salvi (éds), *Grande grammatica italiana di consultazione, vol. II*, (415–481). Il Mulino.

WiV = Blumenthal, P. & Rovere, G. (1998). *Wörterbuch der italienischen Verben – Dizionario dei verbi italiani*. Klett PONS; (avec la collab. de M. Mayer) Acolada, 2023, <https://acolada.de/woerterbuecher/woerterbuch-der-italienischen-verben/>, consulté le 09.05.2023.



**Krzysztof Bogacki**

Professeur émérite, Université de Varsovie  
Pologne

 <https://orcid.org/0000-0003-2755-4276>

# Analyse linguistique au service du TAL

## Linguistic analysis at the service of NLP

### Abstract

Using the notion of predicate in the sense of an element structuring the sentence at the surface level, G. Gross arrives at an interpretation of linguistic facts that is valid for those who adopt the concept of semantic predicate as a tool for analysis. It belongs to the abstract level of meaning and is considered to be an essential part of the mechanism for disambiguation and selecting the meaning of lexemes or grammatical structures. This mechanism is particularly important in the context of machine translation where it is necessary to choose, in the bilingual transfer dictionary, the accurate equivalent of a polysemic lexeme in the target language.

### Keywords

Ambiguity of predicate, semantic predicate, linguistic analysis, disambiguation

## 1. Introduction

Il y a une dizaine d'années, P. Vaillant (2014), hésitant, se demandait si « la question la plus fondamentale que l'on se pose, en fin de compte, à la lecture du *Manuel d'analyse linguistique* de G. Gross, est de savoir s'il s'agit d'un livre de linguistique ou d'un livre de TAL ». À notre avis, toute la carrière professionnelle de G. Gross est une bonne illustration d'une combinaison intelligente des recherches en linguistique théorique avec leur application pratique en traitement automatique de la langue.

En effet, l'ouvrage de G. Gross contient un exposé d'une méthode d'analyse linguistique développée depuis de nombreuses années à l'Université de Paris 13. Dans l'Introduction on lit : « Ce manuel a un double objectif : d'une part, celui de décrire les mécanismes du français contemporain et, d'autre part, de proposer une méthode d'analyse de la langue destinée au traitement automatique, vu du côté de la linguistique ». Elle met au centre des préoccupations la phrase considérée comme unité minimale d'analyse et réserve beaucoup de place à une série de concepts tels que verbes supports, classes d'objets, prédicats, arguments, actualisation, opérateurs appropriés décrits à l'aide de classes en extension et non de mots isolés testés dans différents projets financés par des sources nationales françaises ou internationales, européennes avant tout. Le *Manuel* se termine par une présentation détaillée de la structure d'un dictionnaire électronique, fondé sur la notion d'emploi qui devrait permettre de reconnaître les structures qui caractérisent les phrases des textes.

Ceux qui avaient suivi le cheminement du chercheur et sa formation intellectuelle au sein du LADL de M. Gross ont été témoins de ses contacts avec des théoriciens polonais. Les premières discussions ont permis de constater que les romanisants de Silésie et de Varsovie rejoints rapidement par les chercheurs de Poznań tout en utilisant les mêmes termes que leurs collègues parisiens leur donnaient un sens différent.

À titre d'exemple, prenons le terme « prédicat », qui est l'un des concepts centraux du livre de G. Gross<sup>1</sup>. En dépit de la différence de sens donné à ce terme par l'équipe de G. Gross et par les représentants du courant sémantique polonais, les conclusions auxquelles arrivent les uns et les autres sont identiques. Ils ont tous utilisé ce terme, en lui donnant un sens différent. Les Parisiens ont poursuivi l'idée de Z. Harris axée sur les phénomènes directement observables en surface, tandis que les Polonais étaient résolument proches des hypothèses fondamentales des linguistes russes tels que J. Apresjan, A. K. Zolkovsky, L. N. Iordanskaya, E. V. Padučeva, dont les idées ont été développées par A. Wierzbicka et S. Karolak<sup>2</sup>.

---

<sup>1</sup> Disposant d'une version électronique du *Manuel*, nous avons pu examiner son contenu grâce au logiciel NooJ v. 6 de M. Silberstein. Sur 126223 occurrences de mots, nous avons pu trouver 1160 occurrences de *prédicat* ce qui représente les formes substantivales (sg. et pl.) et adjectivales (sg. et pl.).

<sup>2</sup> Cf. Une des conséquences de cet état de choses a été la publication en 1983 en Pologne d'un *Dictionnaire sémantique et syntaxique des verbes français* (K. Bogacki & H. Lewicka, 1983). Fruit d'un travail d'équipe entrepris dès 1975, remis à l'éditeur en 1978.

## 2. Contenu sémantique vs. élément structurant de surface

Ce qui était caractéristique pour la démarche des linguistes russes et polonais c'était la recherche d'éléments abstraits de contenu les plus simples possibles auxquels on réserve le terme de prédicats sémantiques qui, en dernière analyse, peuvent être réduits à des primitives sémantiques. On se tourne donc résolument vers le contenu et on souscrit à la thèse que la description sémantique des lexèmes doit constituer une paraphrase complète, équivalente et non-tautologique du sens expliqué. Cela n'empêche pas de considérer le sens global du signe linguistique décrit comme une structure composée d'unités abstraites de sens reliées entre elles par des relations spécifiques obéissant à des contraintes de combinabilité formant donc une sorte de syntaxe. De façon très générale on peut dire que ce qui est typique pour G. Gross c'est au contraire de rester très près des formes de structures observables. En cela il reste dans la tradition de son directeur de thèse, M. Gross et de l'équipe du LADL, qui comme Z. Harris demeure attaché à la surface plutôt que d'opter pour un examen de la structure profonde transformée en une structure de surface à l'aide d'opérations de transformations complexes.

Voici comment G. Gross interprète la métonymie :

Si nous examinons la relation entre un prédicat et ses arguments appropriés en termes de classes d'objets, nous observons que la métonymie se caractérise par :

- une violation de la distribution ordinaire du prédicat : *lire* ne sélectionne pas normalement un <humain> en position d'objet, mais un <texte> ou un <support de texte> ;
- un raccourci dans le schéma d'arguments du prédicat : *garer* est normalement un verbe transitif direct avec un troisième argument désignant un <lieu de stationnement> : *J'ai garé ma voiture dans ce parking* ; *Je me suis garé dans ce parking* ;
- une structure sous-jacente inférée : *Lire Sartre* infère *Sartre a écrit un livre* et non pas *Sartre possède un livre* ni même *Sartre a traduit un livre*.

(G. Gross, 2012 : 83)

La clé de l'énigme se trouve dans la formule « violation de la distribution ordinaire du prédicat ». Elle est complétée dans l'alinéa suivant qui parle d'« un raccourci dans le schéma d'arguments du prédicat ». Le même mécanisme sera évoqué par ceux qui utiliseront le terme de prédicat sémantique comme concept central de leur appareil analytique relevant du niveau abstrait du sens.



### 3. Prédicats des mathématiciens

Parmi plusieurs interprétations du terme de prédicat nous allons retenir, en la modifiant, celle qui est utilisée par les mathématiciens pour formaliser le langage des mathématiques proposée à la fin du XIX<sup>e</sup> siècle par G. Frege. Empruntée par les linguistes, elle avait trouvé sa place tout d'abord en syntaxe et ensuite en sémantique où elle fait partie du modèle de grammaire à base sémantique développé par K. Bogacki, S. Karolak (1991). Les mathématiciens posent 4 éléments : les prédicats reliant des variables (arguments), les connecteurs logiques (de coordination, d'implication, de disjonction, de négation) et les quantificateurs. La grammaire à base sémantique se distingue du modèle mathématique sur plusieurs points. Tout d'abord, en ce qui concerne l'arité (la valence) des prédicats. Elle est limitée à 2 arguments connotés par un prédicat dans le modèle de grammaire tandis qu'elle n'a pas de borne supérieure en mathématiques. Cette restriction est levée si, évoquant le parallélisme entre le niveau abstrait des prédicats sémantiques et celui de surface, on se place au niveau des schémas syntaxiques générés par les lexèmes verbaux. Là, on relève à côté des verbes mono- et bivalents, des prédicats verbaux trivalents et même tétravalents.

Dans son manuel de syntaxe du polonais, S. Karolak (2002) donne des exemples :

- des prédicats monovalents avec un argument objet : *s'endormir, se taire, sourire,*
- des prédicats à deux arguments objets : *se marier, visiter, flâner, être père,*
- des prédicats à trois arguments objets : *présenter qqn. à qqn., déposer de l'argent dans une banque,*
- des prédicats à quatre arguments objets : *échanger qqch. contre qqch. d'autre avec qqn.,*
- des prédicats bivalents à deux arguments propositionnels : *causer, entraîner, provoquer,*
- des prédicats trivalents à un argument objet et deux arguments propositionnels : *décourager, obliger,*
- des prédicats trivalents à deux arguments objets et un argument propositionnel : *demander, persuader, inviter,*
- des prédicats tétravalents à trois arguments objets et un argument propositionnel : *dénigrer qqn devant qqn de P, louer qqn. devant qqn de P.*

Le modèle mathématique des prédicats est très puissant et permet différentes extensions. Citons-en deux. La première concerne la nature des arguments. On peut traiter comme arguments les indices temporels qui contrairement aux ar-

guments référant à des objets dans une grammaire à base sémantique n'ont pas de contenu sémantique et qui, tels des gestes, fonctionnent comme des éléments déictiques. Les indices temporels n'ont pas de traits sémantiques associés<sup>3</sup>. Ils relèvent d'un mécanisme de fonctionnement différent de celui des arguments de type temporel tels que *période*, *Moyen Age*, *époque*, *mois*, *semaine* qui renvoient par leur contenu sémantique à des points ou à des segments sur la ligne du temps et qui sont caractérisés avec traits sémantiques [+Temps]. Ils s'avèrent utiles affectant chaque contenu de type prédicatif: qu'il soit prédicat autonome ou prédicat en position d'argument. Ils permettent non seulement de rendre compte des oppositions aspectuo-temporelles<sup>4</sup> mais aussi de décrire les relations dans le temps p. ex. entre l'antériorité de la cause par rapport à l'effet (dans les phrases construites sur *causer* ou *provoquer*) ou entre les notions exprimées par les prépositions *avant*, *après* ou le concept de simultanéité. Ainsi la combinaison de deux indices temporels différents employés comme arguments d'un prédicat dyadique permet de construire la notion d'antériorité et de postériorité. L'adjonction de la négation débouche sur la simultanéité :

$t_i$  AVANT  $t_j$  = antériorité du moment  $t_i$  par rapport au moment  $t_j$

$t_j$  APRES  $t_i$  = antériorité du moment  $t_i$  par rapport au moment  $t_j$

NON ( $t_i$  AVANT  $t_j$ ) & ( $t_j$  AVANT  $t_i$ ) = simultanéité des deux moments  $t_i$  et  $t_j$ .

Une autre extension du modèle mathématique entrevue par A. Wierzbicka (2000) consiste à considérer comme arguments d'un prédicat des expressions faciales analysées à l'aide des configurations des paramètres tels que *sourcils* (qui peuvent être froncés ou levés), *yeux* (grands ouverts ou non), *bouche* (ouverte ou fermée), *coins de la bouche* (abaissés ou levés), *lèvres* (serrées l'une contre l'autre; la lèvre supérieure et le nez «surélevés»)<sup>5</sup> avec des valeurs spécifiques quoique inhabituelles en grammaire traditionnelle permet de sortir du cadre strict des études linguistiques. Les expressions faciales décrites en termes de ces paramètres seraient universellement interprétables et véhiculeraient des significations comparables aux significations des énoncés verbaux. Elles pourraient donc être analysées avec des outils empruntés à la sémantique linguistique. La voie est ainsi ouverte vers la sémiotique.

<sup>3</sup> Cf. D. Dobre (2007).

<sup>4</sup> Cf. A. Kopecka (2009).

<sup>5</sup> Cf. A. Wierzbicka (2000).

## 4. Grammaire à base sémantique

La langue permet à ses usagers de représenter la réalité extralinguistique par des assemblages de signes linguistiques aptes à exprimer tout ce que l'homme veut dire. La fonction représentative des langues naturelles est ainsi étroitement liée à la prédication qui n'est possible que si les éléments sont associés les uns aux autres selon des règles sémantiques et grammaticales précises. Ces critères permettent certains emplois porteurs de sens et en interdisent d'autres. Le sens d'une unité, pertinent pour la communication, résulte d'une part de la place qu'une unité occupe dans le système, d'autre part de son référent dans l'assemblage des objets du monde. Le prédicat sémantique, quand il est actualisé dans le discours, organise sa structure phrastique.

Contrairement à A. Wierzbicka, les partisans de cette méthode n'ont jamais prétendu que les prédicats qu'elle mettait en place étaient inanalysables, qu'ils avaient le statut d'universaux ni que tel inventaire de prédicats suffisait pour construire les représentations sémantiques de toutes les unités lexicales d'une langue donnée.

Comme chez A. Wierzbicka, on trouve à la base la notion de prédicat sémantique qui est présenté d'entrée de jeu sous le terme de concept/notion/contenu sémantique évoquant un sens abstrait (une fonction). Plusieurs distinctions sont à faire. La première concerne l'opposition entre prédicat sémantique et son/ses argument(s) saturant une/des position(s) impliquée(s) par le prédicat. Ces arguments sont sémantiquement obligatoires et se distinguent d'éléments adjoints<sup>6</sup>. D'un autre côté on oppose le niveau conceptuel et le niveau formel et on parle de la syntaxe à chaque niveau (syntaxe des concepts vs. syntaxe des expressions). Les arguments soit réfèrent à des individus auquel cas ils peuvent être caractérisés par des traits sémantiques tels que [Humain], [Animé], [Concret], [Abstrait], [Liquide] etc. qui divisent virtuellement le lexique en très grands groupes (Une structure hiérarchique est supposée : [+Concret] = [-Abstrait]/[+Humain]/[+Animé]) – et ce qui est important – susceptibles de subdivisions ultérieures – soit ils sont de nature prédicative saturant une position impliquée par un prédicat indépendant. Cette double possibilité permet de rendre compte du fait qu'au niveau de surface un verbe correspondant à un prédicat régit plusieurs arguments. Les prédicats utilisés sont introduits avec leurs arguments dont on précise le type (individu ou prédicat en position d'argument), auquel on ajoute le trait séman-

---

<sup>6</sup> Cf. B. Śmigielska (2022) pour une analyse détaillée de la valence des prédicats et l'opposition entre arguments et arguments adjoints.

tique, p. ex. pour le prédicat de perception PERCEIVE (x=[Animé], y=[Concret]), pour le prédicat épistémique KNOW (x=[Humain], R [prédicat en position d'argument]), pour le prédicat de localisation LOC (x=[Concret], y=[Concret]) etc.

La combinaison de plusieurs concepts peut parfois correspondre à un lexème ayant une représentation sémantique fort complexe. Passé un certain seuil, on franchit la limite entre le lexique pour aller vers le syntagme et bien au-delà, vers la phrase. Le mécanisme se prête donc à l'analyse des textes alors qu'à l'autre bout on peut descendre vers celle des affixes (cf. K. Bogacki, 1983). On utilise les prédicats sémantiques, qui dans l'idéal devraient être sémantiquement simples, pour les combiner afin de rendre compte du sens de lexèmes sémantiquement complexes. Les représentations sémantiques qui en résultent peuvent rendre compte aussi bien des lexèmes simples que des expressions développées. On a donc à sa disposition des représentations sémantiques associées à des mots.

Le contenu sémantique propre à un prédicat peut occuper différentes positions dans la représentation sémantique globale du lexème. Considérons le prédicat de perception visuelle (PERCEIVE). S'il domine, l'intuition est celle de lexème de perception visuelle : *Max voit Victor*. Mais il peut occuper une place marginale. Considérons *avoir l'air* tel qu'il est employé dans la phrase : *Victor avait l'air ivre*. Son sens pourrait être rendu par '*J'ai vu Victor et je crois (d'après son apparence) qu'il était ivre*'. Le contenu de cette phrase consiste en un certain énoncé (conclusion – *Victor était ivre*) et admettrait l'hypothèse qu'il y a eu un acte de perception visuelle qui a servi de base à la formulation de cet énoncé (conclusion). Il y aurait des éléments perceptifs (présupposés : *j'ai vu une situation R avec Victor comme un de ses arguments*). C'était la cause de mon calcul qui a conduit à la conclusion ouvertement assertée : *Victor était ivre*). La perception serait profondément enfouie dans une représentation sémantique et ne le dominerait pas comme dans *Max a vu Victor*.

Les traits sémantiques sont largement employés dans K. Polański (1980). En effet, ils y sont mentionnés de façon systématique pour chaque argument. Au départ, leur liste comporte 16 traits très généraux : [+Abstr], [-Abstr]<sup>7</sup>, [+Anim], [-Anim], [+Hum], [-Hum], [Coll] (=collectif), [Elm] (=élément), [Fl] (=végétal), [Inf] (information), [Instr], [Liqu] (=liquide), [Mach] (=machine), [Mat] (=matière), [Pars] (=partie). Ils peuvent se combiner entre eux. Ainsi [-Abstr -Anim] est utilisé pour les concrets inanimés, p. ex. *Pierre, verre*, [+Hum Pars] pour les parties du corps humain (*main, dent, tête*) etc. À la lecture du dictionnaire on s'aperçoit que le mécanisme des traits sémantiques très généraux s'avère très vite insuffisant pour traiter les lexèmes. [Anim] pour le sujet de *galopować*

<sup>7</sup> Il équivaut à [+Concret], inutilisé dans le système.

'galoper' semble trop large parce qu'il autorise en cette position *cigogne, souris, renard* etc. tandis que [Hum] en position de sujet de *bankrutować* 'faire faillite' exclut les phrases telles que *Cette (banque + société) a fait faillite* etc. Très vite, de nouveaux traits, plus précis, apparaissent<sup>8</sup>, p. ex. [Substance chim] (avec le verbe *analizować* 'analyser'), [Jugement, Décision, Sentence] (avec *apelować*, 'adresser un recours à'), [Œuvre musicale] (*aranżować* 'arranger'), [Teinture] (*barwić* 'colorier'), [Sommeil], [Nature], [Etat d'inconscience] (tous les trois avec le verbe *budzić się* 's'éveiller'), [Tissu] (avec *doszywać* 'coudre'), [Source de chaleur] (avec *grzać* 'chauffer'), [Oiseau], [Insecte] et [Avion] (avec *latać, lecieć* 'voler'), [Circuit] (*galwanizować* 'galvaniser'), [Argent] (*dewaluować* 'dévaluer'), [Ciel] (*chmurzyć* 'couvrir de nuages') etc. Le nombre total des traits utilisés tout au long du dictionnaire dépasse plusieurs centaines.

Les exemples que nous venons d'évoquer montrent que les traits des arguments impliqués par un prédicat ne sont pas nécessairement sémantiquement simples mais, dans certains cas, peuvent constituer des amalgames de prédicats avec leurs arguments impliqués tant et si bien que lorsqu'on les isole du contexte, ils réfèrent à des objets ou à des situations et figurent à ce titre dans un dictionnaire.

On se rend compte de l'utilité des traits si on fait appel à la traduction automatique. Dans ce domaine on voit leur rôle de sélecteur d'équivalents d'un mot dans la langue cible. M. Hrabia (2011) mentionne 17 équivalents polonais de l'adjectif français *farouche*. Ils sont sélectionnés par le trait du substantif approprié qu'ils qualifient et dans un seul cas par le registre (archaïque) de la langue :

[Animal], [Disposition Psychique], [Collectif Animé], [Paysage Naturel], [Adversaire Ou Partisan], [Guerrier], [Humain Au Pouvoir], [Phénomène Météorologique Négatif], [Entêtement], [Son], [Mœurs], [Lieu: Obstacle], [Désir], [Fille]

*płochliwy, dziki, zagorzały, okrutny, wrogi, przeraźliwy, groźny, niepohamowany, niewzruszony, zaciekły, burzliwy, pierwotny, nieprzystępny, nietowarzyski, nieokrzesany, nieprzyjazny, wyuzdany*

Ainsi l'adjectif qualifiant le substantif *biche* affecté du trait [Animal] sera traduit par *płochliwy* et non par *okrutny* qui sera choisi dans le contexte *guerrier farouche*. *Zagorzały* sera sélectionné si le substantif est [Adversaire ou partisan] tandis que *pas farouche* qualifiant *une fille* exigera *wyuzdany*.

<sup>8</sup> Dans la traduction en français pour les besoins de cet article.

Les contraintes sur les arguments des prédicats sémantiques circonscrivent des limites strictes qui expliquent le mécanisme de la sélection de sens. Leur transgression se répercute au niveau du sens. Trois cas sont alors à distinguer :

- Elle peut conduire à l'agrammaticalité (*\*D'innombrables idées vertes...*),
- à une figure rhétorique ou (*la ville respire la fraîcheur de vivre*);
- dans le champ polysémique, elle peut relever d'un autre sens tout aussi valide. (*Ce candidat a bien vu le sujet* : transgression de la contrainte [Concret] sur l'argument objet qui est [Abstrait] entraînant le sens épistémique de *voir*).

S'il est vrai que le nom de G. Gross est associé à une longue liste de concepts qu'il a introduits dans la linguistique on doit reconnaître que certains phénomènes que nous venons d'évoquer ont été entrevus et discutés par des représentants d'autres courants. Ainsi dans la sémantique componentielle l'exemple classique d'analyse en termes de traits sémantiques est celui proposé par B. Pottier (1965) qui pour l'examen des noms de sièges s'est servi de 6 traits : [Pour s'asseoir](=s1), [Sur pied](=s2), [Pour une personne](=s3), [Avec dossier] (=s4), [Avec bras](=s5), [Matériau rigide] (=s6).

## 5. En guise de conclusion

Le parcours professionnel de G. Gross est inséparable de projets du vaste domaine du TAL financés souvent par des sources européennes et qui réunissaient linguistes théoriciens et informaticiens. Parmi différents types d'applications en traitement automatique de la langue, les plus spectaculaires sont sans doute les systèmes de traduction des textes. Cela a permis aux uns et aux autres de se rendre compte du fait que les progrès de la traduction automatique dépendent du perfectionnement de la description de la morphologie, de la syntaxe et de la sémantique d'une langue. Il est évident qu'on a besoin de modules performants que seule la linguistique théorique peut fournir et sans lesquels l'informatique rencontrera vite des obstacles insurmontables.

L'observation de l'évolution des systèmes de la traduction informatique conduit le plus souvent à distinguer cinq étapes. Si on néglige le tout premier système qui s'appuie sur deux listes de mots anglais et russes et génère des résultats médiocres on distinguera 4 approches :

- Traduction automatique basée sur des règles linguistiques et grammaticales associées à des dictionnaires ;

- Traduction automatique statistique utilisant des modèles statistiques obtenus à partir des bases de données linguistiques formées de textes parallèles ;
- Traduction automatique neuronale utilisant des algorithmes neuronaux. Ce moteur a la capacité de relier les phrases au contexte. Il s'autocorrige en revenant sur les parties du texte traduit ;
- Traduction automatique hybride. C'est un système qui comprend au moins une base de données, un module de traduction à base de règles en y ajoutant la traduction humaine.

La publication du livre de G. Gross intervient au moment où on abandonne petit à petit la traduction à base de règles et on assiste à l'avènement des systèmes statistiques. À chaque étape on note une nette amélioration des résultats, mais aussi un enrichissement des mécanismes mis en œuvre pour analyser le texte source et générer le texte traduit en même temps qu'une extraordinaire complication au niveau des dictionnaires. Ils ne sont plus des listes des mots simples avec codes morphologiques mais renferment différents types de figements tenant compte des possibilités d'insertions diverses.<sup>9</sup> Étant donné les progrès accomplis en traduction automatique, on ne s'étonne pas que G. Gross (2012 : 297) termine son livre sur un ton optimiste : « il est possible d'envisager de générer toutes les phrases qui définissent un emploi » espérant minimiser les pièges structurels, morphologiques et syntaxiques qui apparaissent à tout bout de champ en traduction automatique. Mais il reste d'autres difficultés à surmonter....

Mentionnons, pour terminer, une des plus importantes. Elle concerne la traduction des phraséologismes que peuvent contenir les textes à traduire. Ils sont plus ou moins figés et se caractérisent par la discontinuité et l'arbitraire des éléments constitutifs<sup>10</sup> mais ce qui est important par les éléments culturels, l'évocation de connotations variées et impossibles à deviner à partir du sens propre des lexèmes utilisés. Si le programme peut être amélioré quant à la couche « superficielle »<sup>11</sup>, les culturèmes sont beaucoup plus réfractaires à un traitement satisfaisant. M. Saad Ali (2016) compare deux traductions du Coran en français : l'une faite par un traducteur non-musulman, l'autre par un traducteur arabe musulman. La comparaison des deux textes conduit à la conclusion que celle-ci est « plus efficace ». La recommandation est donc celle-ci :

---

<sup>9</sup> Cf. Bień J. & Szafran K., 2001 ; Constant M., 2003 ; Courtois B & Silberztein M., 1990 ; Piot M., 1988 ; Silberztein M., 1989 ; Silberztein M., 1990 ; Szafran K., 1993 ; Vetulani Z., Martinek J., Obrębski T. & Vetulani G., 1998 ; Vetulani Z. *et al.*, 1998 ; Wołosz R., 2000.

<sup>10</sup> Cf. M. Sułkowska (2020).

<sup>11</sup> Cf. M. Constant (2003), J. Senellart (1998).

Si la langue réceptrice ne dispose pas d'équivalence directe, le traducteur a deux choix : soit traduire littéralement l'expression figée pour transmettre la couleur locale aux destinataires, tout en accompagnant cette traduction d'une phrase explicative pour expliciter le vouloir-dire du texte source, soit l'inverse, c'est-à-dire inclure la version explicative au texte et, dans une note, proposer une traduction littérale. Par ce biais, il nous semble, d'une part, que le traducteur sera plus fidèle au texte source et, d'autre part, que le lecteur recevra le maximum d'informations contenues dans ce texte sacré et pourra ainsi partager en grande partie l'effet resenti par les destinataires du texte coranique original

(M. Saad Ali, 2016 : 123).

En effet, ce que l'on recherche c'est la transmission d'un message dans ses relations complexes avec l'environnement culturel et social et moins les contraintes formelles au niveau des signes linguistiques arrangés d'une certaine façon en surface.

C'est dire qu'il y a du pain sur la planche...

## Références citées

- Apresjan, J. D. (2000/1980). *Semantyka leksykalna – synonimiczne środki języka*. Ossolineum.
- Bień, J. & Szafran, K. (2001). Analiza morfologiczna języka polskiego w praktyce. *Biuletyn Polskiego Towarzystwa Językoznawczego* 67, 3–19.
- Bogacki, K. (1983) Représentations sémantiques et suffixes d'adjectif. *Linguisticae Investigationes* 7(1), 1–9.
- Bogacki, K. & Karolak, S. (1992). Założenia gramatyki o podstawach semantycznych. *Język a Kultura* 8, 157–187.
- Bogacki, K. & Lewicka, H. (1983). *Dictionnaire sémantique et syntaxique des verbes français*. PWN.
- Constant, M. (2003). *Grammaires locales pour l'analyse automatique de textes : méthodes de construction et outils de gestion*. Thèse de doctorat, Université Paris-Est.
- Courtois, B. & Silberztein, M. (1990). Dictionnaires électroniques du français. *Langue française* 87, 3–4.
- Dobre, D. (2007). *Mécanismes déictiques dans le discours de presse : le quotidien*. Editura Universităţii din Bucureşti.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses universitaires du Septentrion.



- Hrabia, M. (2011). Désambiguïsation des sens du prédicat adjectival *farouche* dans le cadre d'une approche orientée objets. *Neophilologica* 23, 150–167.
- Iordanskaya, L. N. (1970). Popytka leksikografičeskogo tolkovanija gruppy russkich slov so značenem čustva. *Mašinnyj Perevod i Prokladnaja Lingvistika* 13, 275–289.
- Karolak, S. (2002). *Podstawowe struktury składniowe języka polskiego*. Warszawa.
- Kopecka, A. (2009). L'expression du déplacement en français : l'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial. *Langages* 173, 54–75.
- Mathieu-Colas, M. (1987). *Variations graphiques de mots composés, rapport n° 4 du Programme de recherches coordonnées « informatique linguistique »*. CNRS.
- Pađučeva, E. V. (1996). *Semantičeskije issledovanija. Semantika vremeni i vida v ruskom jazyke. Semantika narrativa*. Škola « Jazyki ruskoj kul'tury ».
- Piot, M. (1988). Conjonctions de subordination et figement. *Langages* 90, 39–56.
- Polański, K. (éd.) (1980). *Słownik syntaktyczno-generatywny czasowników polskich*. Ossolineum.
- Pottier, B. (1965). La définition sémantique dans les dictionnaires. *TraLiLi* (3)1, 33–39.
- Saad Ali, M. (2018). La traduction des expressions figées : langue et culture. *Traduire* 235, 103–123.
- Senellart, J., (1998). Reconnaissance automatique des entrées du lexique–grammaire des phrases figées. Dans B. Lamiroy (éd.), *Le lexique-grammaire* (109–127). Rijksuniversiteit van Gent.
- Silberztein, M. (1989). *Dictionnaires électroniques et reconnaissance lexicale automatique*. Thèse de doctorat, Université Paris VII.
- Silberztein, M. (1990). Le dictionnaire électronique des mots composés. *Langue française* 87, 71–83.
- Sułkowska, M. (2020). Frazeologia w przekładzie, czyli kilka uwag o frazeotranslacji. *Rocznik Przekładoznawczy* 15, 335–350.
- Szafran, K. (1993). *Automatyczna analiza fleksyjna tekstu polskiego (na podstawie Schematycznego indeksu a tergo Jana Tokarskiego)*. Thèse de doctorat, Université de Varsovie.
- Śmigielska, B. (2022). Il ne faut pas se laisser tromper par la langue : entre syntaxe et sémantique. *Neophilologica* 34, 1–19.
- Vaillant, P. (2014). Note de lecture de *Manuel d'analyse linguistique* (de Gaston Gross). *Texto! — Textes et Cultures* 19(1), 1–8.
- Vetulani, Z., Martinek, J., Obrębski, T. & Vetulani, G. (1998). *Dictionary Based Methods and Tools for Language Engineering*. Wydawnictwo Naukowe UAM.
- Vetulani, Z. et al. (1998). *Unambiguous coding of the inflection of Polish nouns and its application in electronic dictionaries – format POLEX*. Wydawnictwo Naukowe UAM.


Wierzbicka, A. (2000). The semantics of human facial expressions. *Pragmatics & Cognition* 8(1), 147–183.

Wołosz, R. (2000). *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*.  
Thèse de doctorat, Université de Varsovie.



**Pierre-André Buvet**

Université Sorbonne Paris Nord  
France

 <https://orcid.org/0000-0001-7657-1270>

# La prédication : une relation

## Predication : a relationship

### Abstract

We discuss the concept of predicate from both metalinguistic and linguistic perspectives. After having explained the role of predicate in the theoretical framework of reference, we specify the language mechanisms related to predicates. We then illustrate our purpose with the determination of nouns in French.

### Keywords

predication, modelization, determination

## 1. Introduction

Le terme *prédication* est employé avec une valeur différente de son acception religieuse<sup>1</sup> dans les domaines de la logique et de la linguistique. Dans l'un et l'autre domaines, la prédication se rapporte à la caractérisation d'une proposition<sup>2</sup> : « seul le caractère prédicatif de la proposition peut donc valoir comme critère » (Benveniste, 1974). Le concept de prédication présuppose celui de prédicat : « il paraît dès lors souhaitable de ne pas détacher la notion de prédicat de

---

<sup>1</sup> « Prédication 1 : 'action de prêcher' » selon le *Grand Robert de la Langue Française*.

<sup>2</sup> La proposition est définie ici comme une assertion irréductible à d'autres. L'existence de propositions complexes, c'est-à-dire de propositions constituées d'au moins une autre proposition n'est pas en contradiction avec cette définition car ce qu'asserte fondamentalement la proposition enchâssante est indépendant de la proposition enchâssée, ce qui lui confère un caractère unique.

celle de prédication et de voir alors dans celui-là le support formel de celle-ci » (Gatone, 2008).

La définition de prédicat diffère selon qu'il relève d'une logique aristotélienne ou bien d'une logique non-aristotélienne. Dans le premier cas de figure, la proposition est conçue comme l'application d'un prédicat à un sujet. De ce point de vue, il s'agit de l'assertion d'une propriété, dite rhème, relative à une entité, dite thème. Autrement dit, c'est une structure bipartite telle que le prédicat fonctionne comme un attribut du sujet : « l'attribution d'une certaine propriété à un certain objet [telle qu'il y a] deux parties : un sujet [dit] logique désignant l'objet dont quelque chose est affirmé, et un prédicat indiquant la propriété affirmée » (Ducrot & Todorov, 1972). Dans le domaine de la linguistique, cette représentation de la proposition est exploitée, entre autres, par la grammaire de Port-Royal et, plus récemment, par la grammaire générative (cf. Chomsky, 1971). Cette dernière s'appuie sur la structure logique bipartite pour définir syntaxiquement la phrase élémentaire avec la formule  $P \Rightarrow SN + SV$  (cf. Ruwet, 1967).

La logique non-aristotélienne s'est mise en place au XIX<sup>e</sup> siècle, notamment avec les travaux de Gottlob Frege sur le calcul des prédicats et ceux de Giuseppe Peano sur la logique mathématique (cf. Blanche, 1970). Elle a abouti à une autre représentation d'un contenu propositionnel que la formule suivante synthétise : **Proposition = Prédicat (Argument)**. Il s'agit d'une représentation fonctionnelle, au sens algébrique du terme, telle que les prédicats équivalent à des fonctions dont les variables sont les arguments et les propositions la projection des arguments sur les prédicats : « Le prédicat est alors proche de la fonction mathématique du terme qui nécessite l'attribution de valeurs à des arguments qu'il introduit dans sa dépendance » (Muller, 2013)<sup>3</sup>. La représentation fonctionnelle des

<sup>3</sup> Comme pour un prédicat, le concept de relation est central pour une fonction puisqu'il s'agit d'associer des constantes à des variables pour produire des images. Le parallélisme entre fonction et prédicat est d'autant plus vrai si la fonction est discontinue (par exemple  $y = 1/x + 1$ ) car bon nombre de prédicats ont des contraintes sur leurs arguments. Le concept de spectre distributionnel rend compte de cela. L'arité d'une fonction algébrique désigne le nombre de variables qu'elle met en relation. De même, un prédicat, au sens logique du terme, est distingué selon le nombre d'arguments qu'ils mettent en relation. Le tableau ci-dessous rend compte du rapport analogique entre le concept de fonction algébrique et celui de prédicat logique.

**Tableau 1**

*Correspondance entre fonction algébrique et prédicat logique*

fonction algébrique	prédicat logique
$f(x)=y$	proposition = prédicat (argument)
fonction d'arité 1	prédicat monadique
fonction d'arité 2	prédicat dyadique

propositions sous-tend, entre autres, les études sur la notion de valence verbale (cf. Tesnière, 1959), et sur la notion d'opérateur phrastique (cf. Harris, 1976).

Les travaux présentés ici et leur soubassement théorique s'inscrivant dans le cadre de la logique non-aristotélicienne, ils reprennent à leur compte la représentation fonctionnelle de la proposition. Il s'agit ici de montrer que l'essence relationnelle de la prédication est fondamentale pour expliquer les faits de langue. Nous l'établissons à propos des faits de détermination. Après avoir présenté la théorie de référence, nous détaillons les principaux aspects du modèle de données du point de vue de la prédication et nous établissons ensuite comment la modélisation s'applique aux déterminants.

## 2. Modélisation

Les notions de prédicat, d'argument et de modalisateur sont souvent présentées comme des catégories métalinguistiques distinctes alors que les instances langagières de ces trois notions ont un fonctionnement interdépendant au sein d'un énoncé (cf. Buvet, 2016). Leur interdépendance est symbolisée par la schématisation suivante (cf. Martin, 2016) :

$$E \rightarrow M(P(A))^4$$

La Théorie des Trois Fonctions Primaires est la théorie de référence dans les analyses présentées ici. Elle s'inscrit dans la lignée des travaux de Zellig Z. Harris, *op. cit.*, de Maurice Gross (1981) et de Gaston Gross (2012)<sup>5</sup>. Cette théorie fait

Tableau 1 (Continuation)

fonction algébrique	prédicat logique
fonction d'arité 3	prédicat triadique
...	...

Le concept de prédicat logique est opératoire aussi au niveau supralinguistique qu'au niveau (méta) linguistique

<sup>4</sup> A est une variable d'une fonction P, telle que la relation entre A et P est elle-même la variable d'une autre fonction M. Le symbole E signifie énoncé, le symbole M signifie modalisateur, le symbole P prédicat et le symbole A argument.

<sup>5</sup> Les notions de prédicat et d'argument, telles qu'elles sont mises en avant ici, sont exploitées dans d'autres cadres théoriques, notamment la Théorie Sens Texte (cf. Mel'čuk, 1993) et la Théorie des Cadres (Baker, 2009). Ces deux théories ont également en commun avec les autres théories d'accorder une place centrale au lexique (cf. Blanco & Buvet, 2009).

sienne la schématisation de l'énoncé présentée ci-dessus puisqu'elle vise à modéliser les faits de langue et à expliquer leur mode de fonctionnement à partir de trois fonctions primaires, à savoir la fonction argumentale, la fonction modalisatrice et la fonction prédicative (cf. Buvet, 2013 ; Mejri, 2016, 2017).

La fonction prédicative a deux aspects fondamentaux : d'une part, elle a une dimension cognitive ; d'autre part, elle est de nature métalinguistique. Le premier aspect a trait à l'essence relationnelle de la prédication<sup>6</sup>. C'est un principe explicatif crucial pour la mise en place de la modélisation car il s'applique à différents niveaux d'analyse et contribue à expliquer le langage du point de vue de sa genèse et de celui de son mode de structuration. Ainsi, ce principe intervient dans l'analyse du rôle sémantique des trois fonctions primaires. De même, il entre en ligne de compte dans les études syntaxiques à la fois au niveau macrosyntaxique au niveau microsyntaxique, et au niveau syntagmatique. Il est également fondamental pour les études phraséologiques ainsi que dans les études morphologiques.

La fonction prédicative ressortit à la prédication. Celle-ci est définie comme une relation qui, par l'intermédiaire d'un prédicat, met en contact d'autres constituants langagiers. De plus, c'est une relation orientée : la relation constitutive d'un prédicat a un point de départ et un point d'arrivée qui ne sont pas interchangeables. Autrement dit, la mise en contact des constituants n'est pas symétrique : le fait qu'un prédicat (Pi) relie un premier constituant (C1) à un second constituant (C2) n'implique qu'il relie de la même façon le second au premier. Cette propriété de la prédication est résumée dans le schéma ci-dessous qui fait état de la représentation fonctionnelle d'un prédicat :

$$Pi(C1,C2) \neq \Rightarrow Pi(C2, C1)$$

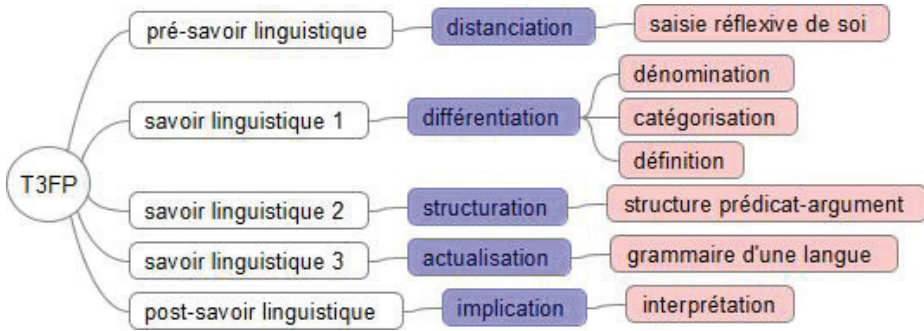
Pour expliquer son mode opératoire, la prédication est rapportée à deux niveaux : d'une part, le niveau supralinguistique ; d'autre part, le niveau (méta)linguistique<sup>7</sup>. Au niveau supralinguistique, la prédication se conçoit comme une opération de pensée, c'est-à-dire un outil cognitif qui est constitutif de la langue en tant que système. Au niveau (méta)linguistique, la prédication se conçoit comme un opérateur énonciatif relatif notamment à des faits de langue d'ordre sémantique, morphologique et syntaxique.

<sup>6</sup> Le premier aspect explique pourquoi la fonction prédicative transcende la fonction argumentale et la fonction modalisatrice, pourquoi elle leur est hiérarchiquement supérieure.

<sup>7</sup> Deux autres niveaux sont distingués : le niveau logico-sémantique, qui concerne les catégories 1) et 2), et le niveau énonciatif, qui concerne uniquement la catégorie 3).

Figure 1

Structuration du langage du point de la prédication



Le schéma ci-dessus permet d'expliquer l'articulation entre le niveau supra-linguistique et le niveau (méta)linguistique du point de vue de la prédication. Cette modélisation du fonctionnement langagier présuppose trois catégories de savoirs : 1) le pré-savoir linguistique ; 2) le savoir linguistique ; 3) le post-savoir linguistique. Les trois catégories se déclinent de façon orientée : de la première à la dernière.

La catégorie 1) est un préalable aux catégories 2) et 3) : l'existence des autres savoirs est conditionnée à celle de ce premier savoir. La catégorie 2) comporte 3 sous-catégories : savoir linguistique 1 ; savoir linguistique 2 ; savoir linguistique 3. Ces trois sous-catégories sont hiérarchisées : la différenciation est première par rapport à la structuration et l'actualisation implique la différenciation et la structuration, cf. *infra*. La catégorie 3) concerne la finalité des précédents savoirs : l'acte de communiquer. La catégorie 2) a trait à la problématique de l'encodage de l'information et la catégorie 3) à la problématique du décodage de l'information (il n'en est pas tenu compte dans cet article). Il y a imbrication entre les catégories 2) et 3) car l'une et l'autre intègre le locuteur et son interlocuteur d'un point de vue systémique (cf. Kerbrat-Orecchioni, 1980). Ces catégories 2) et 3) ne sont pas hermétiques entre elles, car elles présupposent également la maîtrise du code dans les deux cas de figure.

Les concepts de distanciation, de différenciation, de structuration, d'actualisation et d'implication représentent des prédications du niveau supralinguistique. Les autres concepts donnent lieu à des prédications de niveau (méta)linguistique.

Rappelons qu'une « théorie est une manière de concevoir et de percevoir les faits et d'organiser leur représentation [... Qu'elle] sert à conceptualiser et à expliquer un ensemble d'observations systématiques relatives à des phénomènes et à des comportements complexes [... et qu'elle] sert aussi à découvrir un fait caché » (Willett, 1996). Comme toute théorie, scientifique, une théorie

linguistique est fondée sur des postulats. À partir de ce socle de postulats, il est possible d'émettre des hypothèses sur le fonctionnement des langues en proposant des modèles de données et il incombe aux analyses menées dans le cadre théorique de référence de confirmer ou d'infirmer ces hypothèses et, par voie de conséquence, les modèles de données.

La modélisation postule que la distanciation, la différenciation, la structuration, l'actualisation et l'implication sont cinq mécanismes langagiers nécessaires à la formulation et à la compréhension des énoncés.

### 3. Mécanismes langagiers

La saisie réflexive de la conscience de soi est centrée sur la catégorie JE qui caractérise le sujet parlant. L'existence du sujet parlant est tributaire de celle du sujet, indépendamment du trait parlant. Autrement dit, le sujet parlant est une conséquence du sujet en tant que tel. Pour devenir un sujet parlant, le sujet doit se concevoir préalablement comme une entité propre, il doit savoir qu'il est un individu à part entière relativement à ce qui constitue son environnement. La saisie réflexive de la conscience de soi est donc un savoir, dit ici pré-savoir linguistique. Celui-ci est fondé sur un mécanisme de distanciation. La distanciation est une opération de pensée qui, appliquée à un individu, lui permet de s'extraire de son milieu sur le plan cognitif. La distanciation est définie comme une prédication dans le sens où elle établit une relation entre soi et ce qui est environnant à soi. En tant qu'acte de cognition, elle est de nature supralinguistique. La saisie réflexive de soi est la conséquence de la distanciation en tant que reconnaissance de l'individualité du soi. Elle est également définie comme une prédication puisqu'elle fonctionne comme une relation qui a le soi à la fois comme point de départ et comme point d'arrivée. C'est donc une prédication de nature métalinguistique relative à la catégorie JE<sup>8</sup>.

---

<sup>8</sup> La saisie réflexive de la conscience de soi n'a pas trait uniquement à la fonction prédicative, elle se rapporte également à la fonction modalisatrice. Par définition, cette autre fonction primaire est relative à la subjectivité dans le langage dans tous ses aspects (cf. Buvet, 2018a). La définition de la fonction modalisatrice fait sienne celle de la modalité selon laquelle il s'agit de l'implication du sujet parlant dans le processus langagier qui aboutit à la formulation d'un énoncé. De ce point de vue, la fonction modalisatrice participe aussi de la saisie réflexive de la conscience de soi en stipulant que la condition pour devenir un être de parole est de se concevoir primitivement en tant qu'individu. En conséquence, la modélisation fait apparaître que dès l'origine de la construction du langage, il y a interaction entre fonction prédicative et fonction modalisatrice.



La dénomination, la catégorisation et la définition constituent un premier type de savoir linguistique et le mécanisme de la différenciation est à la source de ce savoir. C'est un mécanisme complémentaire de celui de la distanciation dans la genèse du sujet parlant. En s'extrayant de la dimension spatiotemporelle où il évolue, en s'appréhendant comme une individualité, le sujet crée les conditions pour devenir un être de parole, un sujet parlant. L'élaboration du langage, en tant que système de signes, se met alors en place par le moyen de la différenciation. Celle-ci, en tant que prédicat de nature supralinguistique, consiste à relier un faisceau de qualités suffisamment discriminantes à un élément d'un ensemble, à une partie d'un tout de telle sorte que cet élément, ou cette partie, apparaisse comme une entité propre. En se dissociant de ce qui l'entoure, le sujet procède à des distinctions au sein de son environnement. Après s'être perçu et conçu dans son entièreté et dans sa singularité, le soi perçoit et conçoit l'autre. Il distingue ainsi les entités constitutives de son environnement. De ce mécanisme de différenciation, découlent sa capacité langagière à discriminer, l'opération dite de dénomination, sa capacité à organiser, l'opération dite catégorisation et sa capacité à expliciter, l'opération dite définition. La mise en place d'un code partagé par une collectivité se situe à ce niveau de la genèse du langage.

La dénomination, la catégorisation et la définition sont trois concepts fondamentaux pour l'analyse des substantifs. Elles portent respectivement sur trois aspects constitutifs des mots, leur forme, leur relation et leur valeur. La dénomination est centrée sur le premier aspect, la catégorisation sur le second aspect et la définition sur le troisième aspect.

La dénomination est définie ainsi :

relation qu'établit un locuteur (individuel ou collectif) entre lui-même et son public en attribuant un nom à un segment de réalité, dont l'existence est plus ou moins préalablement admise.

(Mortureux, 1984)

parce qu'elle unit un nom à un concept rassembleur d'occurrences particulières, [la dénomination] permet, une fois la compétence acquise, d'utiliser le nom en question, sans nouvel apprentissage, pour tout particulier qui présente les propriétés spécifiques de la classe délimitée par le concept en question.

(Kleiber, 1984)

La dénomination consiste donc à attribuer, dans une langue donnée, une forme linguistique à un concept de telle sorte que tout objet décrit par ce concept est qualifié sur le plan langagier par cette forme. Par exemple, la forme *fourchette* se réfère

à tout objet ayant les caractéristiques physiques d'une fourchette. De ce point de vue, la dénomination est clairement un prédicat métalinguistique puisqu'elle sert à établir une relation entre une forme linguistique et un concept spécifique.

La catégorisation est une conséquence de la dénomination. L'association de concepts à des formes linguistiques nécessite un principe organisateur. En s'appuyant sur des relations hiérarchiques, typiquement des relations hyponymie/hyperonymie, et sur de relations holonymiques, typiquement des relations partie/tout, la catégorisation associe des formes linguistiques à des emboitements conceptuels. Ces derniers expriment des regroupements d'entités, préalablement dénommées, effectués sur la base de leurs caractéristiques ontologiques communes.

*Moucheron > mouche > insecte > animal > être vivant*  
*taon > mouche > insecte > animal > être vivant*

Les entités catégorisées sont de nature concrète ('animal', 'artefact', 'minéral', 'locatif', 'végétal', etc.) ou bien abstraite (par exemple les disciplines scientifiques). Dans tous les cas de figure, elles relèvent fondamentalement de la fonction argumentale. C'est pourquoi la catégorisation concerne essentiellement les substantifs correspondant à des noms catégorématiques (cf. Kleiber, 1981), c'est-à-dire des substantifs « qui présupposent un concept général catégorématique [...] parce que ce type de concepts s'applique à (ou rassemble) des occurrences individuelles qui constituent une catégorie référentielle stable » (Bartning, 1984).

La définition est la troisième prédication métalinguistique résultant de la différenciation. Elle sert à spécifier les propriétés d'une entité dénommée et catégorisée ainsi que sa finalité. Par exemple, *une table, c'est plat et cela sert à poser des choses*. La définition synthétise les deux précédentes prédications métalinguistiques, la dénomination et la catégorisation et, parallèlement, augmente la portée du processus de différenciation en l'inscrivant dans la langue avec une plus grande précision. La définition prend appui sur une dénomination, par exemple *cheval*, en faisant état d'une catégorisation du type la relation hyponymie/hyperonymie. Ainsi, le début de la définition de *cheval* est :

- 'mammifère domestique' dans le Trésor de la Langue Française Informatisé ;
- 'grand mammifère' dans le Grand Robert de la Langue Française ;
- 'mammifère herbivore' dans le Larousse en ligne.

Pour d'autres dénominations, par exemple *museau*, la définition peut également faire état d'une catégorisation du type relation holonymie/méronymie ; par exemple, le début de la définition du mot *museau* est :

- 'partie antérieure' dans le Trésor de la Langue Française Informatisé, le Grand Robert de la Langue Française et le Larousse en ligne.

Ces deux types de catégorisation correspondent en métalexigraphie au concept, d'inspiration aristotélicienne, de genre prochain (cf. Rey-Debove, 1978). La définition d'un mot ne se limite pas pour autant au genre prochain, elle intègre également la différence spécifique (cf. Imbs, 1960). Les informations faisant état de la finalité de l'entité dénommée contribuent le plus souvent à préciser la différence spécifique. Par exemple, la dénomination *musée* est catégorisée 'lieu' ou 'établissement' et a comme différence spécifique :

- 'où sont conservés, répertoriés, classés des objets, des documents, des collections d'intérêt artistique, scientifique ou technique' dans le Trésor de La Langue Française Informatisée ;
- 'dans lequel sont rassemblées et classées des collections d'objets présentant un intérêt historique, technique, scientifique, et, spécialement, artistique, en vue de leur conservation et de leur présentation au public' dans *le Grand Robert de la Langue Française* ;
- 'où sont réunies, en vue de leur conservation et de leur présentation au public, des collections d'œuvres d'art, de biens culturels, scientifiques ou techniques' dans le Larousse en ligne.

Des énoncés génériques résultent des définitions, ils constituent un savoir qui est partagé par les locuteurs d'une même communauté linguistique :

- *Dans un musée, on conserve des objets ;*
- *Dans un musée, on répertorie des objets ;*
- *Dans un musée, on présente des objets ;*
- etc.

Ce savoir linguistique est constitué de représentations langagières des entités du monde qui procèdent des opérations de dénomination, de catégorisation et de définition consécutives au mécanisme de différenciation. La charge référentielle de ces représentations explique leur mode d'organisation, qui s'inscrit dans une perspective ontologique en faisant état de relations de hiérarchisation et d'inclusion. Par ailleurs, les représentations sont porteuses de nouveaux prédicats. Ces derniers inscrivent dans la langue des propriétés des entités ou les relations qu'elles entretiennent entre elles. Il s'ensuit le rôle fondamental de la définition pour expliquer le passage entre deux prédicats supralinguistiques, la différenciation et la structuration.

La structuration est un mécanisme aboutissant à un autre mode d'organisation des représentations langagières que celui de la différenciation. Elle permet de spécifier la nature non seulement logico-sémantique mais aussi morphologique des relations entre les entités sous forme de structures prédicat-argument.

Par exemple, les entités de la catégorie ÊTRE VIVANT et celles de la catégorie ALIMENTS sont spécifiées avec la relation orientée du type ABSORPTION telle que

ÊTRE VIVANT est le point de départ de la relation et ALIMENT le point d'arrivée de la relation. Il s'agit d'une structure prédicat-argument binaire associée à la représentation fonctionnelle ABSORPTION (ÊTRE VIVANT, ALIMENT), cf. *infra*, et telle que ABSORPTION correspond au prédicat et la paire ÊTRE VIVANT/ALIMENT à ses arguments. La structuration relève du niveau supralinguistique et les structures prédicat-argument du niveau (méta)linguistique. C'est pourquoi ces dernières sont associées aux dénominations d'entités que subsument les catégories, par exemple *homme, femme, enfant, animal*, etc... pour la catégorie ÊTRE VIVANT et *fruit, viande, herbe*, etc... pour la catégorie ALIMENT ainsi qu'aux différentes dénotation de la relation ABSORPTION qui correspondent soit à des verbes, par exemple *absorber*<sup>9</sup>, *avaler, brouter, ingérer, manger* ou *se nourrir*, soit à des noms, par exemple, *absorption*, ou *ingestion*, soit à des adjectifs, par exemple *mangeable*.

Toutes sortes de relations caractérisent les entités. Par exemple, la relation orientée du type VENTE caractérise, entre autres, les entités de la catégorie ALIMENT, point d'arrivée de la relation, en la rapportant avec la catégorie ÊTRE HUMAIN, point de départ de la relation. Ce qu'exprime la représentation fonctionnelle de la structure prédicat-argument VENTE (ÊTRE HUMAIN, ALIMENT).

Les unités lexicales et la manière dont elles sont associées dans des structures prédicat-argument sont autant de préconstruits morphosémantiques<sup>10</sup>. En servant de matrices aux contenus propositionnels, les structures prédicat-argument constituent un autre savoir linguiste partagé. Les contenus propositionnels sont les images des arguments par application de prédicats ou, autrement dit, les arguments sont des antécédents sur lesquels s'appliquent des prédicats pour produire des contenus propositionnels. La modélisation montre que la prédication a un rôle prépondérant pour structurer les représentations du monde et contribuer au savoir linguistique.

La structuration, contrairement à la différenciation, implique que les relations sont prééminentes par rapport aux entités. Il s'ensuit la prévalence des prédicats

<sup>9</sup> La modélisation rend compte de l'ambiguïté lexicale, c'est-à-dire le fait qu'une même forme a des acceptions différentes. C'est le cas de par exemple du verbe *absorber* qui a deux valeurs différentes dans *L'invité a absorbé la dernière goutte de son verre* et *Le buvard absorbe l'eau*. De même, elle prend en compte, en intégrant la dimension phraséologique du langage dans ses analyses, le fait qu'un verbe simple de la classe ABSORPTION, par exemple *manger*, soit le constituant d'un verbe complexe sans rapport avec son acception initiale, par exemple *manger son chapeau*. De ce point de vue, elle traite également les combinaisons de nature collocationnelle, par la séquence adverbiale, correspondant à une intensité forte, *comme quatre* relativement au verbe *manger* (cf. Buvet, 2018b).

<sup>10</sup> La structure prédicat-argument ABSORPTION (ÊTRE VIVANT, ALIMENT) présente différentes contraintes entre les arguments dont font état les énoncés génériques issus des définitions : *Une vache, ça mange du (foin + \*poulet)* ; *Un enfant, ça mange du (\*foin + poulet)*.

sur les arguments au niveau (méta)linguistique. Cette prévalence explique la prédication complexe fondée sur des prédicats du second ordre (cf. Prandi, 1992). Ces derniers se distinguent des prédicats du premier ordre par la nature d'au moins un de leurs arguments : il s'agit d'une structure prédicat-argument. Autrement dit, la prédication complexe, contrairement à la prédication simple, est fondée sur un emboîtement prédicatif : le domaine d'arguments d'un prédicat de second ordre comporte *a minima* un autre prédicat (Harris, 1976). Par exemple, la relation ternaire DIRE associe une première entité du type ÊTRE HUMAIN, point de départ de la relation, une seconde entité du type ÊTRE HUMAIN, point intermédiaire de la relation, et un contenu propositionnel P, de nature quelconque et point d'arrivée de la relation. Cette relation au niveau (méta)linguistique correspond à une structure prédicat-argument dont la représentation fonctionnelle est DIRE (ÊTRE HUMAIN, ÊTRE HUMAIN, P) telle que P implique une autre structure prédicat-argument. La structuration est donc aussi à la source de discours plus élaborés que les énoncés faisant état des représentations langagières du monde en termes d'entité et de relation.

La distinction entre la prédication élémentaire et la prédication complexe fait écho à la distinction entre le niveau microsyntactique et le niveau macrosyntaxique. Le premier niveau « décrit l'articulation des morphèmes et des syntagmes, c'est-à-dire des unités qui sont des segments de chaîne de rang inférieur ou égal à la proposition » (Neveu, 1998). Le second niveau « décrit [...] l'organisation des grandes unités discursives [...] dont] l'enchaînement [...] définit quant à lui un niveau d'organisation de nature textuelle » (Neveu, *Ibidem*). Le rôle de la prédication complexe dans la construction du texte est illustré par les relations logiques<sup>11</sup>. L'agencement de ces relations, qu'elles soient explicites, c'est-à-dire des connecteurs du type hypotaxe syndétique ou implicites, c'est-à-dire des parataxes asyndétiques (cf. Trevisse, 2003), contribuent à l'organisation du texte : « Dans l'enchaînement linéaire du texte, les connecteurs sont des termes de liaison et de structuration : ils contribuent à la structuration du texte et du discours en marquant des relations entre les propositions ou entre les séquences qui composent le texte et en indiquant les articulations du discours » (Riegel *et al.*, 1994).

Avec la structuration, c'est donc l'articulation entre la langue et le discours, entre le préconstruit et le construit, entre le niveau logico-sémantique et le niveau énonciatif qui se met en place. Si les dénominations, les catégories et les définitions relèvent quasi-exclusivement de la langue du fait de leur nature exclusi-

---

<sup>11</sup> Les relations logiques prototypiques sont la relation de cause entre deux propositions, par exemple *Vous qui souffrez parce que vous aimez* (Victor Hugo) ou la relation de conséquence entre deux propositions, par exemple *Notre société est organisée de telle sorte qu'il vaut mieux être une chose qu'une conscience* (Eric-Emmanuel Schmitt).

vement métalinguistique, c'est moins le cas pour les structures prédicat-argument qui se positionnent, en tant que matrices de contenus propositionnels, à l'intersection du stade logico-sémantique et du stade énonciatif comme l'attestent leurs propriétés morphosémantiques.

Les structures prédicat-argument, qu'elles soient simples ou complexes, servent à produire des contenus propositionnels, insérables dans des énoncés, en fonction d'un contexte. L'actualisation est le mécanisme permettant l'insertion des premiers dans les seconds. L'énonciation est tributaire de l'intention communicative du locuteur qui se manifeste sous la forme d'un acte de discours (cf. Austin, 1962 ; Searle, 1972). Des actes de discours différents peuvent partager un même contenu propositionnel. Par exemple, les énoncés *Si seulement le maire démissionnait* et *Le maire démissionne* sont deux actes de discours distincts, le premier en rapport avec un souhait et le second en rapport avec un constat, qui partagent un même contenu propositionnel, 'démissionner ('mairie)'. Il s'ensuit les différences de construction directement observables. Inversement, des contenus propositionnels distincts peuvent se rapporter à un même acte de discours. Par exemple, les contenus propositionnels 'agréable'('météo') et 'arriver'('train') peuvent l'un et l'autre faire l'objet d'un questionnement dont fait état la forme interrogative des énoncés : *Est-ce que la météo est agréable ? ; Est-ce que le train arrive ?* Le fonctionnement des actes de discours, lorsqu'il s'agit d'actes illocutoires élémentaires est explicable en rapportant « les énoncés élémentaires [à une représentation fonctionnelle] de la forme f(p) où f est un marqueur de force illocutoire et p une clause<sup>12</sup> » (Vanderveken, 1988).

L'actualisation, en tant qu'énonciation de contenus propositionnels sous forme d'énoncés, nécessite également de prendre en compte la situation de communication et la dimension déictique du discours qui lui est consubstantielle, notamment lorsqu'il prend la forme d'interactions verbales (cf. Kerbrat-Orecchioni, 1990). « L'énonciation est cette mise en fonctionnement de la langue par un acte individuel d'utilisation » (Benveniste, 1966) de telle sorte que la situation de communication est rapportable à trois paramètres : le sujet de l'énonciation (symbolisé par l'embrayeur JE), le lieu de l'énonciation (symbolisé par l'embrayeur ICI) et le temps de l'énonciation (symbolisé par l'embrayeur MAINTENANT). Ces trois embrayeurs sont aussi dénommés respectivement déictique de personne, déictique temporel et déictique spatial dans la mesure où tout recours à la situation de communication est désigné par le terme deixis et les éléments constitutifs de

<sup>12</sup> « Dans la première moitié du XVI<sup>e</sup> siècle, Maigret, en son traité de grammaire française, appelait clause la proposition comme renfermant une pensée parfaitement terminée. » Littré. Par conséquent, une clause correspond ici à un contenu propositionnel.

ce processus par le terme déictique (cf. Ringoot & Robert-Demontrond, 2003). Les marqueurs de la deixis, notamment lorsqu'ils sont désignatifs, constituent « une classe de formes sans dénotation concrète, à référence variable qui ne peut être saisie qu'en relation étroite à la situation [de communication] ; si la situation [de communication] n'est pas connue, le référent ne peut être identifié » (Kleiber, 1986). Par exemple, *ça* et *ici* sont respectivement des déictiques désignatifs dans : *Regarde ça* et *Viens ici*.

La subjectivité est inhérente à tout acte de discours puisqu'il s'agit de la manifestation d'une intention du locuteur. Elle l'est également à la situation de communication du fait que les embrayeurs ICI et MAINTENANT dépendent de l'embrayeur JE<sup>13</sup>, autrement dit du fait que le déictique de personne prévaut sur le déictique spatiale et sur le déictique temporel. En conséquence, la modalité<sup>14</sup> est aussi constitutive de l'actualisation lorsqu'elle implique la prise en charge par le locuteur de son propos (cf. Le Querler, 1996). Les actes de discours et la situation de communication sont donc subsumables sous la modalité, conçue comme la dimension subjective du langage. Les conditions nécessaires à la production d'un énoncé, son énonciation, définissent la modalité. Elle correspond alors à l'ensemble des conventions langagières permettant d'exprimer une intention communicative, c'est-à-dire un acte de discours, en fonction d'un contexte déterminé, c'est-à-dire une situation de communication.

La modalité est conçue comme la composante sémantique de l'actualisation et la formulation comme sa composante morphosyntaxique. Cette dernière correspond à l'ensemble des règles entrant en ligne de compte pour produire des énoncés bien formés. Les règles permettent l'ancrage des contenus propositionnels dans des énoncés bien formés en faisant appel aux actualisateurs, qui sont les marqueurs linguistiques de cet ancrage (cf. Buvet, 2013)<sup>15</sup>. Relativement à une langue donnée, le mode de fonctionnement des actualisateurs, tant sur le plan sémantique que sur le plan morphosyntaxique, est un autre savoir linguistique partagé, la grammaire de cette langue.

---

<sup>13</sup> En effet, la définition du déictique spatial (lieu de l'énonciation) et celle du déictique temporel (repère temporel de l'énonciation) incorpore nécessairement la définition du déictique de personne (auteur de l'énonciation) tandis que la définition du déictique de personne ne fait pas appel aux deux autres définition. Il s'ensuit la prévalence de l'embrayeur JE par rapport aux deux autres embrayeurs.

<sup>14</sup> Les modalités sont catégorisables de la façon suivante : 1) la modalité allocutive : « le locuteur implique l'interlocuteur dans son acte d'énonciation et [lui] impose le contenu de son propos », 2) la modalité élocutive : « le locuteur situe son propos par rapport à lui-même [...de telle sorte qu'il] révèle sa propre position quant à ce qu'il dit » ; 3) la modalité délocutive : « le locuteur laisse s'imposer le propos en tant que tel » (cf. Charaudeau, 1992).

<sup>15</sup> Ce dernier point est illustré dans la section suivante.

Le dernier mécanisme du schéma ci-dessus est l'implication. Il n'est pas traité dans cet article. Signalons néanmoins qu'il concerne *a minima* un autre être de parole : l'interlocuteur, celui qui est visé par l'intention communicative du locuteur. Précisons aussi que son rôle est de permettre à l'interlocuteur le décodage et l'interprétation de cette intention (cf. Moeschler, 2018). Comme dans les autres cas de figure, le mécanisme considéré, en l'occurrence l'implication, prend sa source dans le mécanisme précédent, en l'occurrence l'actualisation. Notamment, l'interlocuteur partage avec le locuteur, d'une part, les savoirs résultant de l'actualisation, et plus largement, de la différenciation et de la structuration, c'est-à-dire des connaissances de nature linguistique et, d'autre part, la situation de communication, c'est-à-dire des connaissances de nature extralinguistique.

#### 4. Étude de cas : les déterminants

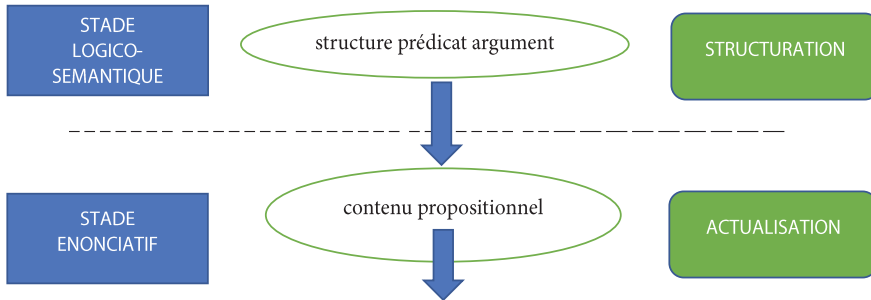
Pour discuter des faits de détermination du point de vue de la prédication, nous présentons un dispositif conçu de manière procédurale : il rend compte des mécanismes de la structuration et de l'actualisation à partir d'un nombre fini d'étapes tels que le point de départ est une structure prédicat-argument et le point d'arrivée un énoncé bien formé.

La première étape se situe à l'interface du stade logico-sémantique et du stade énonciatif. Elle met en œuvre une structure prédicat-argument, correspondant à une représentation langagière du monde et fonctionnant comme une matrice de contenus propositionnels. La seconde étape aboutit à un contenu propositionnel en spécifiant la nature exacte des arguments du prédicat, voire de son seul argument lorsque le prédicat est monadique. Ces deux étapes sont autant de préalables à la formation de l'énoncé qui se met véritablement en place avec les étapes suivantes. Le schéma ci-dessous les résume. Le trait en pointillé symbolise le passage du niveau logico-sémantique au niveau énonciatif. L'énonciation proprement dit se met en place lorsque la frontière est franchie.



**Figure 2**

Passage du stade logico-sémantique au stade énonciatif (1)

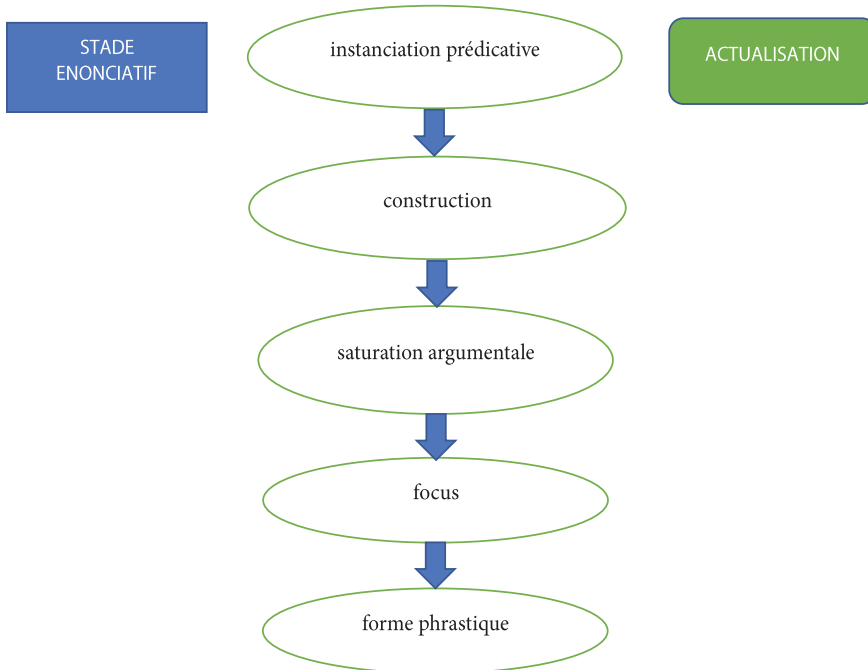


La troisième étape est le choix de la forme de l'emploi prédicatif; elle consiste à stipuler l'instanciation du prédicat sous une forme verbale, adjectivale, nominale, etc.<sup>16</sup>. Les prédicats nominaux et les prédicats adjectivaux, entre autres, nécessitent des constructions différentes de celles qui caractérisent les prédicats verbaux, il incombe à la quatrième étape de préciser la nature de la construction en fonction de la forme du prédicat, par exemple, des constructions à verbe support pour les prédicats nominaux (cf. Daladier, 1996), et des constructions avec copule pour les prédicats adjectivaux (cf. Picabia, 1978). La cinquième étape est la saturation par les arguments des positions sujet et objet d'une construction donnée. La sixième étape, dite focus, est la mise en avant de l'un des constituants majeurs d'une construction par rapport aux autres; il s'agit du prédicat ou bien de l'un de ses arguments. La septième étape porte sur la forme qu'aura l'énoncé en tant que phrase: forme neutre, forme interrogative, forme négative, forme interro-négative, forme impérative. Le schéma ci-dessous résume ces étapes.

<sup>16</sup> Dans le cadre de la théorie des trois fonctions primaires, l'instanciation des prédicats dans un énoncé ne produit pas uniquement des prédicats verbaux comme *conduire* dans *Il la conduisait au poste*. Il peut s'agir également de prédicats adjectivaux comme *surprenant* dans *Il est surprenant*, de prédicats nominaux *pas* comme dans *Il a fait un pas vers nous*, de prédicats adverbiaux comme dans *Il est au-dessous*, de prédicats prépositionnels comme dans *Il est avant elle*. Certains prédicats sont instanciables sous plus d'une forme, comme dans les exemples suivants: *Je respecte cela/J'ai du respect pour cela/Je suis respectueux de cela*.

**Figure 3**

Passage du stade logico-sémantique au stade énonciatif (2)



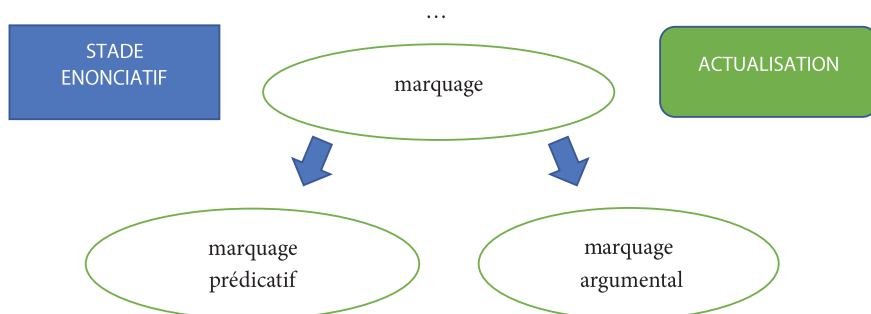
Le marquage est l'étape finale de l'actualisation. Sa représentation formelle est centrée sur un enchaînement de conditions qui caractérisent la formulation en tant que production d'un énoncé bien formé. Les différentes étapes du dispositif procèdent aussi de la modalité car elles sont toutes tributaires de l'intention communicative du locuteur et de la situation de communication. Le marquage consiste à associer le prédicat et ses arguments à des actualisateurs en rapport avec les choix lexicaux effectués par le locuteur. Il est distingué selon qu'il s'applique au prédicat, marquage prédicatif, ou bien à ses arguments, marquage argumental. Pour les prédicats, le marquage est de nature aspecto-temporelle et de nature locutive. Par exemple, dans *Il aimait vraiment cela*, la marque de conjugaison, l'imparfait, est un actualisateur aspecto-temporel et l'adverbe *vraiment* un actualisateur locutif<sup>17</sup>. Pour les arguments, le marquage varie selon que les arguments prennent la forme d'un groupe nominal, d'un pronom, d'une complétive ou d'une infinitive. Lorsqu'il s'agit de groupes nominaux., le marquage argumental porte sur la détermination des substantifs. Les valeurs de

<sup>17</sup> C'est-à-dire il rend compte du point de vue du locuteur sur ce qu'il énonce.

celle-ci concernent obligatoirement la définitude et la quantification et, d'une façon facultative, le locutif (cf. Buvet, 2013). Le schéma ci-dessous résume cette ultime étapes.

**Figure 4**

*Passage du stade logico-sémantique au stade énonciatif (3)*



Le tableau ci-dessous illustre le dispositif décrit précédemment en établissant comment sont produits des énoncés dont la source commune est une prédicat-argument.

**Tableau 2**

*Exemplification du passage du stade logico-sémantique au stade énonciatif*

Numéro étape	nom étape	Exemple
1)	structure prédicat-argument	EXAM- (HUMAIN, HUMAIN)
2)	contenu propositionnel	EXAM- (médecin, patient)
3a)	instanciation prédicative option 1	examen (médecin, patient)
3b)	instanciation prédicative option 2	examiner (médecin, patient)
4a)	construction option 1	X0 procéder à DET examen de X1
4b)	construction option 2	X0 examiner X1
5a)	saturation argumentale option 1	médecin procéder à DET examen de patient

Tableau 2 (Continuation)

Numéro étape	nom étape	Exemple
5b)	saturation argumentale option 2	médecin examiner patient
6a)	focus option 1 focus X0	médecin procéder à DET examen de patient
6a)	Focus option 2 focus X0	médecin examiner patient
7a)	Forme phrastique option 1 forme neutre	médecin procéder à DET examen de patient
7b)	construction option 2 forme neutre	X0 examiner X1
8a)	marquage prédicatif et argumental option 1	<p><i>Le médecin procède à l'examen d'un patient.</i></p> <p><i>Le médecin commence l'examen d'un patient.</i></p> <p><i>Le médecin a procédé à l'examen d'un patient.</i></p> <p><i>Le médecin a commencé l'examen d'un patient.</i></p> <p><i>Un médecin procède à l'examen d'un patient.</i></p> <p><i>Le médecin multiplie les examens de ce patient.</i></p> <p><i>Le médecin a multiplié les examens de ce patient.</i></p> <p><i>Le médecin a salopé l'examen de ce patient.</i></p> <p><i>Un médecin procède à des examens de ce patient.</i></p> <p><i>Un médecin a procédé à une amorce d'examen de ce patient.</i></p> <p><i>Un médecin commence l'examen de ces patients.</i></p> <p><i>Le médecin va procéder à l'examen d'un patient.</i></p> <p><i>Le médecin doit procéder à l'examen d'un patient.</i></p> <p><i>Le médecin peut procéder à l'examen d'un patient.</i></p> <p><i>etc.</i></p>

Tableau 2 (Continuation)

Numéro étape	nom étape	Exemple
8b)	marquage prédicatif et argumental option 2	<i>Le médecin examine un patient.</i> <i>Le médecin a examiné des patients.</i> <i>Des médecin examinent des patients.</i> <i>Le médecin examine ce patient.</i> <i>Les médecins examinent leurs patients.</i> <i>Le médecin a examiné un patient.</i> <i>Un médecin examine de nouveau le patient.</i> <i>Le médecin a examiné plusieurs fois ce patient.</i> <i>Le soi-disant médecin examine ce patient.</i> <i>Soi-disant, le médecin examine ce patient</i> <i>Le médecin a examiné à la va vite ce patient</i> <i>Le médecin va examiner ce patient.</i> <i>Le médecin doit examiner ce patient.</i> <i>Le médecin peut examiner ce patient.</i> etc.

La dernière étape atteste de la variété des énoncés produits à partir d'un même contenu propositionnel. La détermination, à l'instar de la conjugaison, apparaît uniquement à cette étape du dispositif. La variété énonciative est encore plus grande lorsqu'interviennent les étapes dites focus et formes phrastiques car elles impliquent d'autres constructions que celles qui sont standard. Les possibilités associées à ces deux étapes n'ont pas été développées ici. Parmi les actualisateurs, les déterminants et les verbes supports contribuent pleinement à la variété énonciative. On observe dans la série d'exemples indiqués ci-dessus que les premiers portent sur les arguments et les prédicats nominaux tandis que les seconds concernent uniquement les prédicats nominaux. Les semi-auxiliaires sont d'autres actualisateurs qui se combinent avec les prédicats verbaux ou bien avec les verbes supports des prédicats nominaux.

Remarquons que le lexique contribue également à la variété énonciative pour ce qui est du locutif. Par exemple, le nom à valeur péjorative *charlatan* est substituable à *soi-disant médecin* : *Le (soi-disant médecin + charlatan) examine ce patient*. De même, le verbe complexe *jeter un coup d'œil sur* est substituable à la séquence verbale *examiner à la va vite* : *Le médecin (a examiné à la va vite + a jeté un coup d'œil sur) ce patient*.

D'un point de vue méthodologique, le dispositif présenté ici est fondé sur la prédication puisque chaque transition entre deux étapes s'analyse de la façon suivante :

$$\text{étape}_{i+1} = \text{transition} (\text{étape}_i).$$

L'étape postérieure est l'image de l'étape antérieure par application de la transition, qui correspond à la mise en relation des deux étapes. Le dispositif encadre la grammaire produite par le mécanisme de l'actualisation en faisant état de manière hiérarchisée des différentes règles nécessaires à la bonne formation des énoncés.

## 5. Conclusion

Aux différents stades de la modélisation, les trois fonctions primaires contribuent à expliquer les mécanismes de nature supralinguistique et leurs conséquences de nature (méta)linguistique.

La fonction argumentale joue un rôle de premier plan au stade de la différenciation pour comprendre l'articulation entre dénomination, catégorisation et définition. Elle entre en ligne de compte au stade de la structuration car les arguments interagissent avec les prédicats pour spécifier des représentations langagières du monde. Elle intervient également au stade de l'actualisation lorsqu'il s'agit des étapes de la saturation argumentale et du marquage argumentale.

La fonction modalisatrice est consubstantielle du stade de la distanciation puisque ce mécanisme aboutit au sujet parlant, celui qui va organiser le discours, en l'occurrence le locuteur. Elle l'est également au stade de l'actualisation puisque la subjectivité est la marque de l'intention communicative et de la situation de communication qui la constituent.

La fonction prédicative est opérationnelle à tous les stades. Premièrement, les prédicats rendent compte de l'articulation entre le niveau supralinguistique et le niveau (méta)linguistique. Deuxièmement, ils expliquent la dénomination, la catégorisation et la définition. Troisièmement, ils jouent un rôle central dans la structuration. Quatrièmement, la modalité est un prédicat monadique dont l'argument est un contenu propositionnel. Cinquièmement, toujours au stade de l'actualisation les transitions entre les différentes étapes du dispositif sont des prédicats. La nature relationnelle de la prédication explique son omniprésence dans le dispositif.

La modélisation doit faire l'objet de toutes sortes d'expérimentations et d'analyses afin d'être validée. Différents travaux portant sur des aspects morphologiques, syntaxiques, sémantiques, phraséologiques, rhétoriques, textuels des faits de langue ont produit des résultats qui corroborent les hypothèses formulées dans le cadre de théorie de référence. D'autres travaux sont en cours pour confirmer

ces résultats. Notamment, pour ce qui concerne le post-savoir linguistique, le stade où entrent en ligne de compte, au niveau métalinguistique, l'implication et, au niveau (méta)linguistique, l'interprétation des énoncés qui en résultent.

## Références citées

- Austin, J. (1962). *Quand dire c'est Faire*. Seuil.
- Bartning, I. (1984). Aspects syntaxiques et sémantiques des adjectifs ethniques. *Revue Romane* 19, 177-216.
- Baker, C. (2009). La sémantique des cadres et le projet FRAMENET : une approche différente de la notion de valence. *Langages* 176, 32-49.
- Benveniste, E. (1966). *Problèmes de linguistique générale, tome 1*. Gallimard.
- Benveniste, E. (1974). *Problèmes de linguistique générale, tome 2*. Gallimard.
- Blanche, R. (1970). *La logique et son histoire d'Aristote à Russell*. Armand Colin.
- Blanco, X., & Buvet, P.-A. (2009). Présentation : les représentations des structures prédicatives. *Langage* 176, 3-6.
- Buvet, P.-A., Girardin, C., Gross, G. & Groud C. (2005). Les prédicats d'<affect>. *LIDIL* 32, 125-143.
- Buvet, P.-A. (2013). *La dimension lexicale de la détermination*. Champion.
- Buvet, P.-A. (2016). La fonction argumentale au regard des noms d'artefact. *Actes du colloque CMLF 2016*.
- Buvet, P.-A. (2018a). Collocation et modalisation : l'exemple des prédicats d'affect. *Le français moderne 2018-1*, 43-55.
- Buvet, P.-A. (2018b). Collocation et expression de l'intensité. Dans O. Soutet, S. Mejri & I. Sfar (éds), *La phraséologie : théories et applications* (121-135). Champion.
- Charaudeau, P. (1992). *Grammaire du sens et de l'expression*, Hachette.
- Chomsky, N. (1971). *Aspects de la théorie syntaxique*. Seuil.
- Daladier, A. (1996). Le rôle des verbes supports dans un système de conjugaison nominale et l'existence d'une voix nominale en français. *Langages* 121, 35-53.
- Ducrot, O. & Todorov, T. (1972). *Dictionnaire encyclopédique des sciences du langage*. Seuil.
- Gatone, D. (2008). Le prédicat : pour quoi faire ?. *LIDIL* 37, 45-60.
- Gross, G. (2012). *Manuel d'analyse linguistique - Approche sémantico-syntaxique du lexique*. Presses du septentrion.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 7-52.
- Imbs, P. (1960). Au seuil de la lexicographie. *Cahier de lexicologie* 2, 89-105.

- Harris, Z. S. (1976). *Notes du cours de syntaxe*, Le Seuil.
- Kerbrat-Orecchioni, C. (1980). *L'énonciation de la subjectivité dans le langage*. Armand Colin.
- Kerbrat-Orecchioni, C. (1990). *Les interactions verbales 1*. Armand Colin.
- Kleiber, G. (1981). *Problèmes de référence : descriptions définies et noms propres*. Klincksieck, 77-94.
- Kleiber, G. (1984). Dénomination et relations dénominales. *Langages* 76, 77-94.
- Kleiber, G. (1986). Déictiques, embrayeurs, "token-réflexives", symboles indexicaux, etc. : comment les définir ?. *L'Information Grammaticale* 30, 3-22.
- Le Querler, N. (1996). *Typologie des modalités*. Presses universitaires de Caen.
- Martin, R. (2016). *Linguistique de l'universel. Réflexions sur les universaux du langage, les concepts universels, la notion de langue universelle*. Académie des Inscriptions et Belles-Lettres.
- Mejri, S. (2016). Le prédicat et les trois fonctions primaires. Dans D. de Souza Silva Costa & D. Rosane Bençal (éds), *Nos caminhos do léxico* (313-337). Presses de l'UFMS.
- Mejri, S. (2017). Les trois fonctions primaires. Une approche systématique. De la congruence et de la fixité dans le langage. Dans C. Carvalho, M. Planellas Iváñez & E. Sandakova (éds), *De la langue à l'expression : le parcours de l'expérience discursive* (123-144). Presses de l'Université d'Alicante.
- Melčuk, I. (1993). *Cours de Morphologie Générale (théorique et descriptive) 1*. Les Presses de l'Université de Montréal/CNRS.
- Moeschler, J. (2018). L'implicite et l'interface sémantique-pragmatique : où passe la frontière ?. *Corela HS-25*.
- Muller, C. (2013) Le prédicat, entre (méta)catégorie et fonction. *Cahiers de lexicologie* 102, 51-65.
- Mortureux, M.-F. (1984). La dénomination approche sociolinguistique. *Langages* 19, 95-112.
- Muller, C. (2013). Le prédicat, entre (méta)catégorie et fonction. *Cahiers de lexicologie* 102.
- Neveu, F. (1998). Macrosyntaxe : le problème des niveaux de l'analyse syntaxique dans La Route des Flandres. *L'Information Grammaticale* 76, 38-41.
- Picabia, L. (1978). *Les Constructions adjectivales en français : systématique transformationnelle*. Droz.
- Prandi, M. (1992). *The Building Blocks of Meaning: Ideas for a Philosophical Grammar*. John Benjamins Publishing.
- Rey-Debove, J. (1978). *Le métalangage, étude linguistique du discours sur le langage*. Éditions Le Robert.
- Riegel, M., Pellat, J.-C. & Rioul, R. (1994). *Grammaire méthodique du français*. Presses Universitaires de France.
- Ringoot, R. & Robert-Demontrond, P. (2004). *L'Analyse de discours*. Éditions Apogée.



- Ruwet, N. (1967). *Introduction à la grammaire générative*. Plon.
- Searle, J. R. (1972). *Les actes de langage*. Hermann.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.
- Trevisse, A. (2003). À propos de quelques relations inter-énoncés : hypotaxe, parataxe, asyndète et construction du sens. Dans A. Celle & S. Gresset (éds), *La subordination en anglais. Une approche énonciative* (51-68). Presses Universitaires du Mirail.
- Vanderveken, D. (1988). *Les Actes de discours*. Mardaga.
- Willett, G. (1996). Paradigme, théorie, modèle, schéma : qu'est-ce donc ?. *Communication et organisation* 10.



**Wiesław Banyś**

Université de Silésie à Katowice  
Pologne

 <https://orcid.org/0000-0003-2471-6751>

## Inférences au pays de la prosodie

### Inferences in the land of prosody

#### Abstract

When it comes to language, it's not just the grammatical structure and the literal meaning of words that matter. The way a predicate imposes inferences on its propositional arguments is crucial to understanding the true meaning of a message. However, these inferences are influenced by many factors, such as prosody, world knowledge, speakers' expectations regarding language use, situational stereotypes, and other implicit or contextual elements.

In this paper, we examine the inferential status of the verbs referred to by Karttunen as “implicative verbs”. On the one hand, two-way implicative verbs, and on the other, one-way implicative verbs. The latter have been little studied from this perspective. This will lead us to highlight the fundamental role, too often forgotten, of prosody and focus/theme in this type of analysis and in determining the inferential status of predicates.

Our analyses show that, once prosody has been considered, the classification of theoretically possible verb inferences accepted until now needs to be modified.

We do not have 4 groups of one-way implicative verbs, as has been argued, but 2, namely the groups: [+ + / - // - / -] [affirmed > true or false // denied > false] of the type *être capable, pouvoir* and [+ + / - // - / +] [affirmed > true or false // denied > true] of the type *hésiter à*.

The remaining two groups, considered as distinct and autonomous one-way implicative verbs with the suggested characteristics: [+ + // - + / -] [affirmed > true // denied > true or false] of the type *forcer to and* [+ - // - + / -] [affirmed > false // denied > true or false] of the type *refuser de*, belong to the canonical groups of two-way implicative verbs, respectively: *forcer à* to the group of verbs of the type *réussir à*: [+ + // - / -] [affirmed > true // denied > false] and *refuser de* to the group of verbs of the type *oublier de*: [+ / - // - / +] [affirmed > false // denied > true].

Naturally, this classification differentiation, important as it is, only reflects the different behaviour of certain types of predicate, and this is the most important element of these analyses with a view to automating the recognition of predicate inferences.

#### Keywords

Implication, entailment, presupposition, implicature, prosody, theme, rheme, two-way implicative verbs, one-way implicative verbs, text-to-speech synthesis

## 1. Introduction

En ce qui concerne la langue, ce n'est pas seulement la structure grammaticale et le sens littéral des mots qui comptent. La manière dont un prédicat impose des inférences à ses arguments propositionnels est essentielle pour comprendre le véritable sens d'un message. Cependant, les inférences, à part les implications sémantiques nécessaires, sont influencées par de nombreux facteurs, tels que la prosodie, la connaissance du monde, les attentes des locuteurs concernant l'utilisation de la langue, les stéréotypes situationnels et d'autres éléments implicites ou contextuels. Alors que nous aspirons à créer des machines qui « comprennent » la langue comme nous le faisons, surtout à l'ère de l'intelligence artificielle (IA), il est essentiel de reconnaître que cette compréhension va bien au-delà de la simple décomposition lexicale et de l'analyse syntaxique.

Dans cet article, nous examinons le statut inférentiel des verbes désignés par Karttunen comme « verbes implicatifs ». D'une part, les verbes à implication bidirectionnelle (« two-way implicative verbs »), et d'autre part, les verbes à implication unidirectionnelle (« one-way implicative verbs »). Ces derniers ont été peu étudiés de ce point de vue. Cela nous amènera à mettre en évidence le rôle fondamental, trop souvent oublié, de la prosodie et du focus/thème dans ce type d'analyse et dans la détermination du statut inférentiel des prédicats c'est sur les relations entre la prosodie et les inférences des verbes à implication bi- et unidirectionnelle que nous nous concentrerons par la suite.

Il n'y a pas de langue humaine sans parole et il n'y a pas de parole sans prosodie. L'analyse de la langue et de la parole ne peut pas se passer de l'analyse de la prosodie. La prosodie ne se résume pas à de simples modulations musicales de la voix qui suivent les fluctuations de la pensée.

Le sens d'un message ne découle pourtant pas uniquement des mots et de leur signification directe, mais également de la façon dont ils sont prononcés. La prosodie, dans ce cadre, est essentielle pour une communication effective apportant une multitude de nuances influençant l'interprétation du discours et elle ne devrait pas être considérée comme un élément secondaire ou périphérique du langage. Il est donc vital de ne pas oublier d'intégrer ces éléments lors des analyses linguistiques.

La prosodie englobe des éléments tels que l'accent, les tonalités, le rythme, la durée et l'intonation, impliquant des aspects comme l'intensité, la durée et la hauteur sonore. Ces phénomènes, souvent qualifiés de suprasegmentaux, transcendent l'analyse phonémique traditionnelle. Les éléments de la phrase

et la phrase elle-même sont donc constitués de deux couches : une couche de phonèmes et l'autre, une couche mélodique, se superposant à la première.

Pendant longtemps, on n'a pas donné suffisamment d'importance à ces phénomènes, qui étaient perçus davantage comme de simples expressions sonores que comme des unités à part entière, à la différence des phonèmes.

C'est en fait seulement durant les dernières décennies que l'étude de la prosodie et de l'intonation de la phrase a connu un développement important. Les nouveaux logiciels d'analyse acoustique (p. ex. Signalyze, Klatt Speech Tools, R, WinPitch, Praat, Cspeech, etc.) y ont joué leur rôle et ont contribué à ce développement permettant aux chercheurs de collecter et de traiter des données de manière plus facile et plus précise. Ces données ont permis aux chercheurs de développer aussi de nouvelles théories sur l'intonation, qui ont souvent été inspirées de la phonologie générative et des descriptions autosegmentales des langues africaines.

Les travaux des chercheurs nord-américains (cf. en particulier Pierrehumbert, 1980) ont joué un rôle important dans ce développement et ont contribué à l'émergence d'une approche dominante de l'intonation, basée sur le concept de structure prosodique profonde. Cette approche a été formalisée dans le système de notation ToBI – *Tones and Breaks Indices* (cf. p. ex. Pierrehumbert & Hirschberg, 1992, qui est désormais utilisé par de nombreux chercheurs dans le monde entier, cf. p. ex. Delais-Roussarie *et al.* 2015).

Cette approche dominante a conduit à une certaine standardisation des analyses de l'intonation. Les chercheurs se concentraient désormais en particulier sur l'étude des structures prosodiques sous-jacentes aux énoncés, plutôt que sur l'analyse des intonations spécifiques utilisées dans des contextes particuliers.

Ce changement de paradigme a également eu des implications méthodologiques. Les chercheurs s'intéressaient de plus en plus à l'étude des intonations dans des contextes contrôlés, en principe dans des laboratoires.

Cette façon de procéder a permis aux chercheurs de collecter des données plus précises et de contrôler les facteurs qui peuvent influencer l'intonation. Elle a aussi permis d'accroître notre compréhension de l'intonation, mais elle a également été critiquée pour son manque de prise en compte de la variabilité de l'intonation dans la parole réelle.

Toutes ces analyses ont démontré aussi que la mélodie de la parole peut également être divisée en segments, et que les éléments prosodiques sont authentiquement des unités linguistiques (cf. p. ex.: Zwanenburg, 1965 ; Léon, Martin & Baligand 1969 ; Rossi 1999 ; Lacheret-Dujour & Beaugendre, 1999 ; Simon, 2004 ; Gussenhoven, 2004 ; Mertens, 2004 ; Ladd, 2008 ; Di Cristo, 1998, 2000, 2013 ; Frotta & Prieto, 2015 ; Avanzi, Simon & Post, 2016 ; Féry, 2017 ; Persson, 2017 ; Gardes Tamine, 2018 ; Martin, 2009, 2015, 2018 ; Dalrymple, Lowe & Mycock, 2019).

Une fois les unités des analyses prosodiques détectées et décrites, la question se posait de savoir quelles sont les significations que la prosodie, dont l'intonation, peut transmettre. L'éventail de de ces significations est très large. Néanmoins, on peut distinguer quatre types généraux de ces significations : émotions, attitude du locuteur, actes de langage et significations pragmatiques, dont les inférences et la structure informationnelle de la phrase (cf. p. ex. Ladd, 1980, 1996 ; Bolinger, 1986, 1989 ; Hirschberg, 2006 ; Hedberg & Sosa, 2008 ; Büring, 2016 ; Féry, 2017 ; Tonhauser, 2019).

La prosodie et l'intonation servent à exprimer divers états émotionnels, notamment la joie, la tristesse, la rage, la surprise, la peur. Une intonation montante à la fin d'une phrase peut p. ex. indiquer un sentiment de surprise, tandis qu'une intonation descendante peut exprimer un sentiment de tristesse.

L'attitude du locuteur à l'égard du contenu de la phrase peut être aussi exprimée par le choix de l'intonation appropriée, p. ex. une intonation montante peut être employée pour exprimer un sentiment de doute, une intonation descendante peut être employée, en revanche, pour donner un sentiment d'assurance.

La prosodie décide aussi du type d'acte de langage ou du statut pragmatème ou non des constructions. Par exemple, dans le cas des actes de langage impératifs (ordres, demandes, etc.), on peut utiliser la prosodie pour indiquer le degré de politesse ou d'urgence de l'ordre ou de la demande : la phrase canonique « Passe-moi le sel » peut être prononcée avec une intonation montante à la fin, ce qui indiquerait une demande polie, ou avec une intonation descendante, ce qui indiquerait un ordre (cf. aussi quant aux pragmatèmes p. ex. Lin, 2018a, 2018b ; Banyś, 2020).

Les fonctions pragmatiques de l'intonation englobent p. ex. le codage d'éléments tels que ce dont on parle (topic, thème, information ancienne) et ce dont on en dit (focus, rhème, information nouvelle, concentration sur un contenu particulier de la phrase, éléments « at-issue ») (cf. p. ex. Bogusławski, 1977 ; Karolak, 1984, 2007 ; Karolak & Bogacki, 1991 ; Banyś & Karolak, 1988 ; Hajičova, Partee & Sgall, 1998 ; Mertens, 2008 ; Lee, Gordon & Büring, 2008 ; Katz & Selkirk, 2011 ; Beyssade *et al.*, 2015 ; Lee, Kiefer & Krifka, 2017 ; Kügler & Calhoun, 2020).

Les désignations citées – thème, topic, etc. – ne sont pas toujours équivalentes, mais décrivent d'une manière ou d'une autre la structure informationnelle de l'énoncé. Une intonation montante ou contrastive de l'énoncé ou d'une partie de l'énoncé indique généralement que le locuteur présente l'élément marqué de cette manière comme information nouvelle, tandis qu'une intonation non marquée de cette manière indique que le locuteur transmet une information ancienne. Cette différenciation prosodique représentant une structuration différente de la

structure informationnelle de l'énoncé influence naturellement aussi le statut des contenus assertés et inférés à partir de l'énoncé, tels que présuppositions et implicatures (cf. p. ex. les références ci-dessus), et c'est sur cette relation que nous nous concentrerons en particulier par la suite.

Avec les différents types de prosodie (intonation, accent) de l'énoncé et différente structuration informationnelle de l'énoncé va de pair un ordre déterminé des éléments de l'énoncé (dans le cas des langues avec l'ordre des mots dit « libre » ou « flexible »), qui est marqué s'il est différent de la structuration prosodique par défaut de l'énoncé, ou par des procédés syntaxiques, tels que p. ex. mise en relief ou détachement à gauche ou à droite.

En général, la prosodie est un élément essentiel des caractérisations des énoncés et de leurs éléments du point de vue de la théorie générale du marquage et de ses liens avec la théorie de la cognition et les concepts du fond et de saillance (cf. p. ex. Wertheimer, 1945 ; Köhler, 1920 ; Battistella, 1990 ; Langacker, 1991).

Paraphrasant et modifiant, ou peut-être explicitant, un peu la constatation de Levinson (1995 : 109–110) affirmant que « [l]a théorie du type de sens de l'énoncé devrait être une théorie de l'interprétation par défaut », nous dirions que la théorie du type de sens de l'énoncé devrait être une théorie de l'interprétation par défaut dans laquelle la prosodie et l'ordre des mots par défaut jouent un rôle fondamental.

Dans ce qui suit, nous nous concentrerons donc sur les relations entre la prosodie et un type particulier d'inférences qu'on trouve dans les verbes à complément phrastiques mettant en jeu des relations subtiles entre prosodie, assertions, présuppositions, implicatures, structure informationnelle et ordre des mots.

Nous commencerons notre présentation par une introduction sommaire :

- de l'importance des études sur les inférences et de la combinatoire des inférences sémantiques possibles des prédicats (sec. 1),
- pour passer par la suite à un bref rappel des distinctions entre implications, présuppositions, implicatures, et des discussions autour d'elles (sec. 2),
- cela nous permettra d'analyser plus en détail le statut d'un groupe des verbes étudiés de différentes manières de ce point de vue, à savoir les verbes implicatifs, et d'un groupe de verbe – les verbes non-implicatifs – très peu étudiés de ce point de vue, ce qui nous amènera à mettre en relief le rôle fondamental de la prosodie et du focus/rhème dans les analyses (sec. 3),
- ces analyses nous amèneront, dans la partie *En guise de conclusion*, à jeter un bref coup d'œil sur au moins deux tâches à accomplir dans un avenir proche, ainsi qu'à formuler quelques commentaires sur l'état actuel de la synthèse vocale conversationnelle, notamment en ce qui concerne les liens entre la prosodie, les inférences et la structure informationnelle des énoncés.

## 2. Combinatoire des inférences sémantiques possibles des prédicats

Pour comprendre pleinement un énoncé ou un texte en langue naturelle, il est essentiel de reconnaître et d'analyser ce qui est inféré. Cette compétence est nécessaire non seulement pour les humains, mais aussi pour les systèmes automatisés qui sont basés sur l'intelligence artificielle. Elle nécessite une capacité d'interpréter, d'inférer et d'intégrer une multitude d'indices linguistiques et extra-linguistiques pour arriver à une compréhension complète.

Les recherches sur ces questions ont considérablement augmenté et de nombreux progrès ont été réalisés dans ce domaine. À titre d'exemple, des projets comme le *PASCAL Network of Excellence Recognising Textual Entailment (RTE-1)* ou le programme *Advanced Question Answering for Intelligence (AQUAINT)* illustrent les nombreuses avancées dans ce champ d'étude (cf. p. ex. Crouch, Saurí & Fowler, 2005; Dagan, Roth, Sammons & Zanzotto, 2013; Bernardy & Chatzikyriakidis, 2020; Liu, Wei, Niu, Chen & He, 2020; Elshazly, Haggag & Ehssan, 2021; Paramasivam & Nirmala, 2021).

Le débat sur les inférences sémantiques des prédicats, notamment des verbes mais aussi, par exemple, des adjectifs, lorsqu'ils sont liés à un argument propositionnel, est bien documenté, surtout en ce qui concerne la langue anglaise (cf. p. ex. les œuvres fondamentales de Kiparsky & Kiparsky, 1970; Karttunen, 1971a, 1971b, 1972; Keenan, 1971; Givón, 1972; Rudanko, 1989, 2002; Nairn, Condoravdi & Karttunen, 2006; Egan, 2008).

Contrairement à l'anglais, en français, les recherches sont principalement axées sur les questions liées aux présuppositions et implicatures, comme en témoignent les travaux de chercheurs tels que p. ex. Ducrot (1972, 1977), Anscombe et Ducrot (1983), Zuber (1972), Kleiber (2012), Fasciolo (2012), Biglari et Bonhomme (2018).

En ce qui concerne les implications sémantiques des prédicats lorsqu'ils sont appliqués à leurs arguments propositionnels en français, il n'y a en réalité que quelques études à citer (cf. p. ex. Drapeau & Gérard, 1973; Mørdrup, 1975; Falk & Martin, 2017a, 2017b; Banyś, 1987a, 1987b, 2022).

Avant de poursuivre notre analyse, il est nécessaire de mentionner la définition du mot « implication » telle qu'elle est définie par Karttunen, ce sur quoi Karttunen (1971b) lui-même a mis naturellement l'accent. Nous adoptons cette définition ici, et à ce sujet, on peut également se référer, parmi une multitude de travaux, aux discussions dans p. ex. Strawson (1966, 1971), Grice (1975), Levinson (1983) et Atlas (2005).

Pour clarifier ses propos, Karttunen utilise des exemples qu'il a fournis et qui ont été traduits en français dans la version française de son article (1973a : 62). Il note que quand il emploie le terme « implication », il veut dire que la phrase p. ex. (numérotation des exemples empruntée au texte français) :

(18)(b) *Max n'a pas eu la prévoyance de rester éloigné.*

implique

(19)(b) *Max n'est pas resté éloigné.*

Lorsque Karttunen utilise le mot « implique », il ne l'entend donc pas dans le sens de « implique logiquement » ou « entraîne ». Il souligne que la relation qu'il décrit est légèrement moins forte. Il ne l'utilise pas non plus dans le sens classique d'« implication matérielle » (cf. p. ex. Lewis, 1917 ; Ajdukiewicz, 1956 ; Nelson, 1966 ; Horn & Ward, 2006 ; cf. aussi Bogusławski, 1986 ; Banyś, 1989, 2000). La manière dont Karttunen définit le terme « implique » est la suivante :

« P implique Q ssi

à chaque fois que P est asserté,

le locuteur est obligé de croire que Q. » (Karttunen, 1973a : 62).

Karttunen met en avant que la règle du *modus tollens*, qui est valide pour le sens logique classique d'« implication », ne s'applique pas au sens plus « faible » du terme « implique » qu'il utilise. On voit bien que l'emploi de la formule « est obligé de croire que Q », sans discuter davantage la différence éventuelle entre « est obligé » et « il est nécessaire que », ne dissipe pas les doutes quant à la nature exacte de la relation décrite.

Dans leurs travaux ultérieurs, Karttunen et ses collègues ont précisé qu'ils adoptent la conception du type « Strawson entailment » (Karttunen & Cases, 2019 : 7), comme l'a appelée von Fintel (1999), et, dans cette formulation, A implique B seulement si A satisfait toutes les présuppositions de B. Ce qui entraîne cette fois-ci, à son tour, la nécessité de préciser ce qu'est la présupposition.

Par conséquent, en reprenant et en modifiant légèrement l'exemple (18) de Karttunen, nier la proposition complétive de (18) ne signifie pas nécessairement nier ou contredire sa proposition principale :

(19)(b2) *Max est resté éloigné*

n'implique pas la phrase

(18)(b) *Max a eu la prévoyance de rester éloigné,*



ce qui montre tout de suite, entre autres, que, même si les deux phrases ont les mêmes valeurs de vérité (en fin de compte dans le cas des deux phrases le fait que Max est resté éloigné est vrai), le prédicat, ici, (*avoir*) *la prévoyance* apporte une différence sémantique, et une majeure partie des discussions là-dessus concerne le statut (sémantique, pragmatique, de quelle nature, etc.) de l'élément apportant ce type de différence.

C'est pour cette raison qu'on peut affirmer que la phrase – exemple classique des discussions :

(22a) *Jean est parvenu à embrasser Marie*

implique dans le sens précisé, avec toutes ses imprécisions, du terme que

(22b) *Jean a embrassé Marie.*

On peut établir la combinatoire théorique possible des valeurs de vérité de l'argument propositionnel impliquées par le prédicat principal. Dans le tableau ci-dessous représentant une telle combinatoire les valeurs de vérité « vrai »/« affirmé » sont indiquées par « + » et « - » indique « faux »/« nié », et la possibilité des deux valeurs de vérité « vrai/faux », donc non implication seulement de l'une d'elles, on pourrait dire, à la Karttunen, « non-implication » tout court, est indiquée par « +/- » (cf. p. ex. Karttunen, 1971a, 2012 ; Banyś, 1987a, 1987b, 2022) :

**Tableau 1**

*Combinatoire théorique possible des valeurs de vérité de l'argument propositionnel impliquées par le prédicat principal*

Classe	Prédicat principal	Argument propositionnel
1	+	+
	-	-
2	+	+
	-	+
3	+	-
	-	+
4	+	+
	-	+/-
5	+	-
	-	+/-

Tableau 1 (Continuation)

Classe	Prédicat principal	Argument propositionnel
6	+	+/-
	-	-
7	+	+/-
	-	+
8	+	+/-
	-	+/-

Dans ses premières études, Karttunen utilise principalement le terme « verbe » pour désigner les prédicats appartenant à telle ou telle classe. Cependant, il est devenu apparent, comme Karttunen l'a reconnu lui-même en offrant d'autres exemples d'implications, et comme il a choisi de le faire plus exhaustivement dans ses recherches subséquentes (cf. p. ex. Karttunen, 2013), qu'il est essentiel d'examiner également les adjectifs, ou en d'autres termes, les prédicats dans leur ensemble. Issus de la tradition de l'école linguistique des analyses des structures prédicat – arguments fondée par Karolak, nous adoptons d'une manière naturelle cette approche plus globale.

De cette manière,

- les prédicats qui illustrent la combinaison des valeurs 1, formant la première classe de prédicats, ont été nommés par Karttunen « two-way implicative verbs » – verbes à implication bidirectionnelle ou « implicatifs » (p. ex. *réussir*) (cf. Karttunen, 1973, 2015),
- les prédicats de la classe 2 – « factifs » (désignation introduite par Kiparsky & Kiparsky, 1970) (p. ex. *regretter*),
- les prédicats de la classe 3 – two-way implicative verbs – verbes à implication bidirectionnelle ou « implicatifs négatifs » (p. ex. *refuser*),
- les prédicats de la classe 4 – one-way implicative verbs – verbes à implication unidirectionnelle ou « si-prédicats » (p. ex. *forcer qqn à*),
- les prédicats de la classe 5 – one-way implicative verbs – verbes à implication unidirectionnelle ou « si prédicats négatifs » (p. ex. *empêcher qqn de*),
- les prédicats de la classe 6 – one-way implicative verbs – verbes à implication unidirectionnelle ou « seulement si prédicats » (p. ex. *avoir la possibilité de*),
- les prédicats de la classe 7 – one-way implicative verbs – verbes à implication unidirectionnelle ou « seulement si prédicats négatifs » (p. ex. *hésiter à*),
- les prédicats de la classe 8 – des prédicats « trous », qui n'ont pas d'implication déterminée dans le cas de l'assertion et de la négation.

C'est pour cette raison que Karttunen n'intègre pas les prédicats du type 8 dans sa première classification. Il se concentre uniquement sur les verbes ayant des implications clairement définies dans les deux scénarios – que ce soit en affirmation ou en négation.

Karttunen nomme donc les verbes qui ont des implications soit « + » soit « - » dans les deux cas: assertion et négation verbes à implication bidirectionnelle (« two-way implicatives verbs ») et ceux qui n'en ont qu'une, soit en assertion soit en négation verbes à implication unidirectionnelle (« one-way implicatives verbs »), ne s'intéressant pas tellement aux cas où il y avait plus qu'une telle implication possible et où on avait affaire à deux possibilités: vérité ou fausseté acceptables. La combinatoire ci-dessous présente, en revanche, toutes les combinaisons d'inférences prédicatives possibles.

Seraient de ce point de vue, selon Karttunen, verbes à implication bidirectionnelle les prédicats de la classe (1):

**Tableau 2**

<b>1</b>	+	+
	-	-

p. ex. *réussir, arriver, parvenir, être capable de, aboutir, oser, daigner.*

et de la classe 3:

**Tableau 3**

<b>3</b>	+	-
	-	+

p. ex. *manquer, négliger, refuser, oublier (d'), décliner, repousser, rejeter, renvoyer, résister, dédaigner.*

En revanche, seraient verbes à implication unidirectionnelle les prédicats des classes:

**Tableau 4**

<b>4</b>	+	+
	-	+/-

p. ex. *forcer, contraindre, obliger, pousser, imposer, presser, coercer, astreindre.*

**Tableau 5**

5	+	-
	-	+/-

p. ex. selon Karttunen, *refuser de, empêcher de, éviter de, interdire de, prévenir de.*

**Tableau 6**

6	+	+/-
	-	-

p. ex. *être capable, être en état, permettre, autoriser, consentir, on y reviendra par la suite.*

**Tableau 7**

7	+	+/-
	-	+

p. ex. *hésiter, craindre, ne pas oser.*

Drapeau (1973), Mørdrup (1975) et Falk et Martin (2017a, 2017b) ont partiellement classé les verbes français selon la classification de Karttunen, mais une classification complète et une discussion approfondie des principes précis d'analyse et de classification sont encore à faire.

Il faut pourtant tout de suite noter un problème avec ce type de classification si l'on regarde de plus près les classes de verbes à implication unidirectionnelle, p. ex. la classe 5.

Le problème consiste en ceci que le fait que Karttunen (cf. Karttunen, 2012 :7 ; Karttunen & Cases, 2019 : 14, 16) donne pour exemple canonique de cette classe le prédicat *refuser*, qui a été aussi qualifié par lui et ses collaborateurs comme verbe à implication bidirectionnelle est tout à fait révélateur d'un problème important sous-jacent à cette classification. Cela revient en fait à poser la question de savoir si nous avons là affaire à deux classes d'implications (bidirectionnelles et unidirectionnelles) distinctes ou une seule et même classe d'implications avec des ajouts contextuels de sens dépendant d'autres facteurs que le sens du prédicat ou encore un autre arrangement des cases de cette classification.

De ce point de vue le statut des implications à la Karttunen de 4 classes de verbes, théoriquement possibles, appelés verbes à implication unidirectionnelles est loin d'être clair, ce qui pose la question de savoir s'il est légitime de les postuler comme classes à part, au moins sous la forme présentée.

C'est sur cette question que nous nous pencherons par la suite. Avant de l'aborder pourtant, nous rappellerons très brièvement les éléments principaux du sens qui peuvent entrer ici en jeu et nous signalerons les grandes lignes des discussions autour d'eux.

### 3. Présupposition vs implicature

Les recherches en sémantique et pragmatique ont montré, avec toutes leurs vicissitudes, que le sens d'une phrase ne se limite pas à une interprétation monolithique; il peut être segmenté en plusieurs couches. La métamorphose de p. ex. Wittgenstein dans ces travaux allant du *Tractatus Logico-Philosophicus* (1922) aux *Investigations philosophiques* (1953) est très révélateur dans ce contexte. Une de ces couches clés concerne la différence entre ce qui est pré-supposé et ce qui est asserté. Où exactement se situe cette distinction par rapport aux domaines de la sémantique et de la pragmatique est encore un sujet de débats. Tandis que certaines théories orientent les présuppositions vers la sémantique, d'autres estiment qu'elles relèvent davantage de la pragmatique (cf. p. ex. entre autres, ensemble avec les travaux cités ci-dessus, Strawson, 1966, 1971; Stalnaker, 1974; Levinson, 1995; Recanati, 2004; Saul, 2002; Horn, 2002, 2006; Moeschler, 2018; Abrusan, 2022). Néanmoins, toutes les approches intègrent des aspects à la fois sémantiques et pragmatiques, illustrant ainsi la riche interconnexion entre les présuppositions, les implicatures conversationnelles, les implicatures conventionnelles et les assertions à l'intersection de ces deux territoires.

Il n'est pas nécessaire et possible, même si c'est très tentant, d'évoquer ici toutes les subtilités des débats sur ces questions extrêmement complexes, étant donné l'ampleur de la problématique et de la littérature là-dessus existante, pour réaliser l'analyse que nous envisageons.

Nous nous limiterons par conséquent à un rappel très succinct des définitions généralement acceptées des deux notions principales qui seront mises en jeu dans ce qui suit, à savoir les présuppositions et les implicatures conversationnelles, laissant pour l'instant derrière une très riche discussion sur certaines controverses que les deux notions ont évoquées et qui donnent lieu à des constatations intrigantes à propos des présuppositions du type de celle posée par p. ex. Beaver, Geurts et Denlinger (2021: sec. 6) : *Née pour échouer, ou pas encore née? (Born to fail, or unborn?)* ou ce commentaire de Karttunen : *Le zoo des déclencheurs de présuppositions auraient dû être construit avec des cages séparées pour les différentes espèces. La recherche d'une théorie unifiée de la présupposition, pragmatique ou sémantique, a été un échec* (Karttunen, 2016: 706–707). Nous empruntons ici pour les besoins de cette analyse les propositions des définitions ci-dessous à la manière de Partee (2009: 1–4), cf. aussi Levinson (1983: 97–223), Pott (2005).

Commençons par la présupposition sémantique :

Une phrase S présuppose une proposition p si p doit être vraie pour que S ait une valeur de vérité (pour être vraie ou fausse).

Ci-dessous sont mentionnés quelques propriétés et tests des présuppositions sémantiques :

- la non-annulabilité : les présuppositions sémantiques ne sont pas annulables, car si elles ne sont pas vérifiées, la valeur sémantique de la phrase ne peut même pas être calculée, elle serait donc suspendue, d'une certaine manière, dans le vide.
- Les présuppositions sémantiques sont des propriétés des phrases.
- le caractère conventionnel : les présuppositions sémantiques sont intrinsèquement liées à la signification usuelle des mots et structures utilisées.
- la mise en arrière-plan : pour que l'énoncé soit pertinent dans une situation donnée, ses présuppositions sémantiques doivent être préalablement reconstruites comme partie de l'arrière-plan conversationnel ou du terrain commun.
- le comportement sous la négation, etc. : en présence de négation, de questions ou de conditions, les présuppositions sémantiques demeurent intactes.
- la détachabilité : des mots et des phrases équivalents en termes de conditions de vérité peuvent avoir différentes présuppositions sémantiques.
- la non indépendance de ce qui est le point principal de l'énoncé (« at-issue meaning ») : normalement, les présuppositions doivent être vraies comme précondition à la vérité (ou fausseté) ou à l'admissibilité générale ce qui est l'objet de l'énoncé.

Ce dernier point et sa formulation par Partee demandent d'être commentés pour ne pas induire en erreur : le terme être « at-issue » dans un énoncé (cf. Roberts, 1996, 2012 ; Potts, 2005) ; cf. aussi Simons *et al.* 2010 ; Koev, 2018 ; Mahler, 2019 ; Ronai & Xiang, 2021 ; Hesse *et al.* (2022) se réfère au contenu principal ou au point principal qu'une phrase véhicule, contrairement à d'autres types de sens ou d'informations qui pourraient être véhiculés implicitement ou comme un effet secondaire. Ce qui constitue ce point ou contenu principal de l'énoncé est déterminé, parce qu'il doit avoir un effet quelconque dans la structuration et la forme de l'énoncé, par la prosodie et l'ordre des mots. Cet élément/ces éléments de l'énoncé qui porte(nt) une intonation montante ou un accent d'insistance constitue(nt) (un élément du) le point, le contenu principal de l'énoncé. Ces éléments de l'énoncé prosodiquement marqués sont des éléments qui sont niés si l'énoncé était nié et sont une réponse à une question portant sur l'énoncé si elle serait posée. Cette dernière formulation équivaut à la formulation répandue ces

derniers temps dans ce type d'analyse sous forme de *Questions under Discussion* (QUD) (cf. p. ex. Beaver *et al.*, 2017 ; Benz & Jasinskaja, 2017 ; Grindrod & Borg, 2019). L'idée générale est que chaque phrase dans un discours aborde une QUD, souvent implicite, soit en y répondant, soit en soulevant une autre question qui peut aider à répondre à cette QUD.

Pour les chercheurs, comme nous, issus de la tradition d'analyse linguistique en structures prédicat – arguments et structures thème – rhème créée par Karolak (1984, 2007), Karolak et Bogacki (1991), Banyś et Karolak (1988) et Bogusławski (1977, 1986), être « at-issue » veut dire, d'une manière opérationnelle, porter une saillance prosodique et être par conséquent le rhème de la phrase (cf. aussi les discussions autour de la signification du focus/rhème p. ex. dans la tradition d'alternatives de Rooth, 1992, cf. aussi Bocci, Bianchi & Cruschina, 2020 ; Cruschina, 2022 ; Spalek & Oganian, 2019).

De notre point de vue, nous dirions que les éléments de l'énoncé à une structure prosodique et l'ordre des mots déterminés qui ne sont pas potentiellement niés ou ne répondent potentiellement pas à une question possible ne constituent pas les éléments principaux du message, ne sont donc pas « at-issue », donc ils sont indépendants, non : non-indépendants, et constituent donc la partie thématique de l'énoncé, autrement dit : sa partie présupposée. À moins que la « non-indépendance » des présuppositions sémantiques du point central de l'énoncé veuille dire que, dans une interprétation spécifique du mot, les présuppositions sémantiques, une fois un énoncé articulé, dépendent du point central de l'énoncé, dans le sens que les deux éléments de l'énoncé sont fonction l'un de l'autre : le choix de l'un comme point principal de l'énoncé entraîne le fait que l'autre devient sa présupposition et inversement. Mais, dans ce cas-ci, la formulation serait plutôt inverse : Non-Indépendance de ce qui est le point principal de l'énoncé (*at-issue meaning*) des présuppositions sémantiques : le choix de l'un détermine le statut de l'autre.

Ce dernier point des caractéristiques des présuppositions sémantiques devrait donc recevoir plutôt, toutes les remarques ci-dessous prises en compte, la formulation suivante :

Indépendance de ce qui est le point principal de l'énoncé (*at-issue meaning*) : normalement, les présuppositions doivent être vraies comme précondition à la vérité (ou fausseté) ou à l'admissibilité générale ce qui est l'objet de l'énoncé.

Quant à l'implicature conversationnelle elle pourrait être caractérisée ainsi :

Dans le contexte *c*, si l'énoncé *A* entraîne pragmatiquement *B* grâce aux maximes conversationnelles, alors *A* implique conversationnellement la proposition *p*.

Ci-dessous sont mentionnés quelques propriétés et tests pour les implicatures conversationnelles :

- la possibilité d'annulation : on peut révoquer une implicature conversationnelle en introduisant des précisions. Ceci est couramment valable pour les déductions pragmatiques.
- Les implicatures conversationnelles concernent non pas les phrases elles-mêmes, mais la manière dont elles sont énoncées dans un contexte donné.
- la non détachabilité : leurs déductions se basent sur le sens de l'énoncé plutôt que sa forme linguistique. Changer les mots ne les supprime pas si le sens demeure intact. Cela diffère des présuppositions et des implicatures conventionnelles.
- l'absence de convention : elles ne sont pas rattachées à la signification usuelle des expressions. Pour les déduire, il faut d'abord comprendre le sens conventionnel de l'énoncé.
- la calculabilité : on peut les calculer en combinant le sens littéral avec le principe de coopération et les maximes conversationnelles.
- la renforçabilité : selon Sadock (1978), les implicatures conversationnelles peuvent être renforcées sans redondance en les combinant avec des énoncés qui expriment explicitement des idées similaires ou identiques.
- la non mise en arrière-plan : les implicatures conversationnelles émergent de la formulation des énonciations ; elles ne sont pas présumées être des acquis de l'arrière-plan conversationnel ou du terrain commun.
- le comportement sous la négation, etc. imprévisible. Leur calcul dépend du contexte global de l'énoncé, et non pas seulement de la présence locale d'une négation par exemple. On ne peut donc pas considérer qu'elles se produisent systématiquement « sous » la portée sémantique d'opérateurs particuliers.

#### **4. Prosodie et statut des verbes à implication bidirectionnelle et des verbes à implication unidirectionnelle**

Nous pouvons donc passer maintenant à une analyse du statut des prédicats en question et à l'analyse de la question posée de savoir si nous avons là affaire à deux classes d'implications (4 groupes de verbes à implication unidirectionnelle et 2 groupes de verbes à implication bidirectionnelle) distinctes ou une seule et même classe d'implications avec des ajouts contextuels de sens



dépendant d'autres facteurs que le sens du prédicat ou encore un autre arrangement des subdivisions de cette classification.

Les analyses de ces deux grandes classes de verbes sont relativement rares, par rapport à p. ex. les analyses des prédicats factifs (cf. p. ex. Karttunen, 1971 ; Erdmann, 1974 ; Hazlett, 2010 ; Kastner, 2015 ; Tonhauser, 2016 ; Simons *et al.*, 2017 ; Djärv & Bacovcin, 2017 ; White & Rawlins, 2018 ; Jeong, 2020, 2021 ; Degen & Tonhauser, 2022 ; Ziembicki, 2022 ; Ziembicki, Seweryn & Wróblewska, 2022), et concernent avant toute chose la nature catégorielle de l'élément additionnel du sens des verbes implicatifs : présupposition ou implicature conversationnelle, et ce qu'elle contient. Les analyses portent avant tout sur le sens et le fonctionnement du prédicat canonique du groupe des verbes à implication bidirectionnelle, à savoir *to manage – réussir* (cf. p. ex. Nadathur, 2016 ; Pavlick & Callison-Burch, 2016 ; Włodarczyk, 2019, 2021).

Prenons comme point de départ, l'analyse originale de Karttunen de ces verbes, ce qui nous permettra de voir comment la prosodie y avait sa part.

Comme exemples des prédicats implicatifs, à implication bidirectionnelle, positifs et négatifs, Karttunen (2015 : 301) donne les phrases (traduites ici en français) du type (numérotation originale) : verbes implicatifs positifs (+ + | – –) :

**Tableau 8**

*Deux groupes de verbes à implication bidirectionnelle*

+ +   – – implicatives	+ –   – + implicatives
manage to	fail to
bother to	neglect to
remember to	forget to
see fit to	refrain from ... ing
happen to	avoid ... ing

(Karttunen, 2015 : 300)

- (8) a. *Le malfaiteur a réussi à s'enfuir.* → *Le malfaiteur s'est enfui.*  
 b. *Elle n'a pas pris la peine d'expliquer.* → *Elle n'a pas expliqué.*  
 c. *Il a jugé bon de lui demander une nouvelle chance.* → *Il lui a demandé une nouvelle chance.*  
 d. *Kim n'a pas pensé à prendre son petit-déjeuner.* → *Kim n'a pas pris de petit déjeuner.*

Et les verbes implicatifs négatifs (+ – | – +), avec la polarité inversée :

- a. *Il n'a pas réussi à entrer à Oxford.* □ *Il n'avait pas été admis à Oxford.*
- b. *Elle n'a pas évité de se faire prendre.* □ *Elle s'est fait prendre.*
- c. *Il n'a pas négligé de la rappeler.* □ *Il l'a rappelée.*
- d. *Kim a oublié de prendre son petit-déjeuner.* □ *Kim n'a pas pris de petit déjeuner.*

Il est intéressant de noter que ce qui est le plus pertinent de notre point de vue, et du point de vue du rôle de la prosodie dans l'interprétation des inférences, a été mentionné dans une note en bas de page :

En plus de leurs propriétés implicatives, toutes les constructions du tableau 2 [verbes implicatifs] suggèrent quelque chose d'autre. Par exemple, *se souvenir de VP* et *oublier de VP* impliquent tous deux que le protagoniste avait l'intention de VP ou était censé le faire. Cela peut conduire à des arguments tels que *Je n'ai pas oublié d'aller à la fête. Je n'ai jamais eu l'intention d'y aller*, qui ne portent pas sur ce qui s'est passé mais sur la question de savoir si l'oubli était impliqué.

(Karttunen, 2015 : 300, note 4)

D'une part l'emploi alternant de « suggérer » et de « impliquer » est déroutant, normalement, ce qui est suggéré peut être dénié et ce qui est impliqué non. Nous entendons bien que « impliquer » (*imply* en anglais) a, comme Karttunen l'a indiqué, différents sens, mais dans le contexte cité *suggérer* est considéré comme équivalent de *impliquer*, ce qui revient à dire en fait qu'on aurait affaire ici, au moins dans le cas des verbes implicatifs à implication bidirectionnelle et dans l'un des sens de « imply », à une implicature conversationnelle, ce qui paraît risqué comme constatation.

D'autre part, et d'une manière beaucoup plus importante du point de vue de nos analyses, ce qui est suggéré/impliqué dans ces phrases ne peut être dénié que si l'on met un accent d'insistance contrastif sur « oublier » et par conséquent un contour intonatif particulier de tout l'énoncé.

C'est une importance double : d'une part, parce qu'on voit la possibilité de la prosodie combinée ici avec négation mettre en cause ce qui est le fond de la conversation dans le cas d'un énoncé à intonation non marquée, ce qui est donc présupposé, et, d'autre part, les conséquences de cette constatation si on l'applique d'une manière conséquente à la classification des verbes implicatifs à implication bidirectionnelle et unidirectionnelle eux-mêmes.

Passant à l'analyse de 4 groupes de verbes à implication unidirectionnelle :

**Tableau 9**

*Quatre groupes de verbes à implication unidirectionnelle*

<u>+ + implicatives</u>	<u>+ - implicatives</u>	<u>- - implicatives</u>	<u>- + implicatives</u>
cause NP to	refuse to	can	hesitate to
force NP to	be unable to	be able to	
make NP to	prevent NP from		

(Karttunen, 2015 : 301)

Karttunen propose les exemples suivants et remarque tout de suite que l'inversion de la polarité supprime toute implication logique, contrairement aux exemples de verbes à implication bidirectionnelle de (8) et (9) :

- (10) a. *Ils ont forcé la foule à se disperser.*  *La foule s'est dispersée.*  
 b. *Jean a refusé de bouger.*  *Jean n'a pas bougé.*  
 c. *Il n'était pas capable de se lever.*  *Il ne s'est pas levé.*  
 d. *Elle n'a pas hésité à l'aider.*  *Elle l'a aidé.*

Comme le note Karttunen (2015 : 301), « [s]i une personne dit *J'ai été capable de me connecter*, on est enclin à conclure qu'elle l'a fait, et c'est peut-être ce que le locuteur a l'intention de dire. Cependant, il n'y a pas de contradiction à ce qu'elle continue *mais je ne l'ai pas fait* ».

Cette remarque, toute juste qu'elle est, paraît être un peu à côté du problème principal, parce que l'objet de la remarque est la phrase affirmative avec *être capable de faire quelque chose*, – et non pas négative, comme dans l'exemple original. L'interprétation de cette phrase affirmative ne fait aucun problème. Le problème est, cependant, ailleurs, dans le cas des phrases négatives.

Prenons le comportement du verbe de l'exemple (10a), un verbe qui est censé représenter la classe des verbes à implication unidirectionnelle, (+ +) d'après la combinatoire des valeurs de vérité proposée par Karttunen. D'après cette combinatoire la valeur de vérité de la complétive si la phrase est niée est indéterminée, toutes les deux, la vérité et la fausseté, devraient être possibles et une fois la phrase affirmée, comme dans (10a), la complétive est vraie :

- (10a2) *Ils n'ont pas forcé la foule à se disperser.*  $\sqsubset$  *La foule ne s'est pas dispersée*
- (10a3) *Ils n'ont pas forcé la foule à se disperser.*  $\sqsubset$  *?\* La foule s'est dispersée.*

La première phrase paraît naturelle, cependant, la seconde, si on l'énonce sans une prosodie particulière, marquée, ne paraît pas tout-à-fait spontanée. En revanche, si on la prononce avec un accent d'insistance sur « forcé », symbolisé ici en caractères gras et soulignés, et avec l'ajout de « d'elle-même » p. ex., elle serait acceptable :

- (10a4) *Ils n'ont pas forcé la foule à se disperser.*  $\sqsubset$  *La foule s'est dispersée d'elle-même.*

La combinatoire des valeurs de vérité de cette classe de prédicats à implication unidirectionnelle avec un contour prosodique non-marqué, normal, serait : (+ + // - -), identique donc à celle des verbes à implication bidirectionnelle, du type *réussir* (notons que, ici et par la suite, la barre singulière « / » renvoie à la possibilité des deux valeurs de vérité, et la barre double délimite les phrases affirmatives des phrases négatives).

Prenons maintenant le comportement du verbe de l'exemple (10b), un verbe censé représenter la classe des verbes à implication unidirectionnelle, (+ -), d'après la combinatoire des valeurs de vérité proposée par Karttunen. D'après cette combinatoire la valeur de vérité de la complétive si la phrase est niée est indéterminé, toutes les deux, la vérité et la fausseté, devraient être possibles, et une fois la phrase affirmée, la complétive est fausse :

- (10b2) *Jean n'a pas refusé de bouger.*  $\sqsubset$  *Jean a bougé.*
- (10b3) *Jean n'a pas refusé de bouger.*  $\sqsubset$  *?\* Jean n'a pas bougé.*

De nouveau, la première phrase paraît naturelle, mais la seconde paraît un peu forcée à moins d'être prononcée avec une intonation particulière. Si l'on met donc un accent sur le mot « refusé », et si on reflète graphiquement cela en le mettant en gras et en le soulignant, la phrase sonnerait plus naturel :

- (10b3) *Jean n'a pas refusé de bouger.*  $\sqsubset$  *Il était immobile depuis déjà un certain temps.*

La combinatoire des valeurs de vérité de cette classe de prédicats à implication unidirectionnelle avec un contour prosodique non-marqué, normal, serait :

(+ - / - +), donc elle est identique à celle des verbes à implication bidirectionnelle, du type *oublier de faire quelque chose*.

Analysons maintenant le comportement du verbe de l'exemple (10c), un verbe censé représenter la classe des verbes à implication unidirectionnelle, (- -), d'après la combinatoire des valeurs de vérité proposée par Karttunen. D'après cette combinatoire la négation du verbe entraîne la fausseté de la complétive, et la valeur de vérité de la complétive si la phrase est affirmée est indéterminé, toutes les deux, la vérité et la fausseté, devraient être possibles :

- (10c1) *Il n'était pas capable de se lever.*  $\square$  *Il ne s'est pas levé.*  
 (10c2) *Il était capable de se lever.*  $\square$  *Mais il ne s'est pas levé.*  
 (10c3) *Il était capable de se lever.*  $\square$  *Et il s'est levé.*

Toutes les trois phrases sont naturelles, n'exigeant pas une prosodie particulière, marquée, pour que les inférences présentées puissent être effectuées.

La combinatoire des valeurs de vérité est donc : (+ +/- // - -), donc celle des verbes à implication unidirectionnelle.

Passons maintenant à l'analyse du comportement du verbe de l'exemple (10d), un verbe censé représenter la classe des verbes à implication unidirectionnelle, (+ +/- // - +), d'après la combinatoire des valeurs de vérité proposée par Karttunen. D'après cette combinatoire la négation du verbe entraîne la fausseté de la complétive, et si la phrase est affirmée la valeur de vérité de la complétive est indéterminé, toutes les deux, la vérité et la fausseté, devraient être possibles :

- (10d) *Elle n'a pas hésité à l'aider.*  $\square$  *Elle l'a aidé.*  
 (10d) *Elle a hésité à l'aider.*  $\square$  *Mais elle l'a aidé.*  
 (10d) *Elle a hésité à l'aider.*  $\square$  *Et elle ne l'a pas aidé.*

Toutes les trois phrases sont naturelles et n'exigent pas une prosodie particulière, marquée, pour que les inférences indiquées puissent être effectuées.

La combinatoire des valeurs de vérité serait donc : (+ +/- // - +), donc celle des verbes à implication unidirectionnelle.

Ces analyses montrent que, une fois la prosodie prise en compte, la classification des inférences des verbes théoriquement possibles doit être modifiée.

On n'a donc pas 4 groupes de verbes à implication unidirectionnelle, mais 2, à savoir les groupes 6 et 7 :

- // + +/- du type *être capable, pouvoir*
- +/- // - + du type *hésiter à*.

Les deux groupes considérés comme constituant des groupes distincts des verbes à implication unidirectionnelle, à savoir les groupes 4 et 5 :

- + + // - +/- du type *forcer à*
- + - // - +/- du type *refuser de*

font partie des groupes canoniques de verbes à implication bidirectionnelle, respectivement : *forcer à* du groupe de verbes du type *réussir à* et *refuser de* du groupe des verbes du type *oublier de faire*.

Naturellement, cette différenciation de la classification aussi importante qu'elle est ne reflète qu'un comportement différent de certains types de prédicats, et c'est ce qui est l'élément le plus important de ces analyses en vue d'automatisation de la reconnaissance des inférences des prédicats.

## 5. En guise de conclusion

Il est important de noter que d'autres verbes à implication unidirectionnelle que Karttunen et ses collaborateurs présentent dans une base de verbes implicatifs simples qui a été rendue publique et accessible (cf. [http://web.stanford.edu/group/csli\\_lnr/Lexical\\_Resources/simple-implicatives/simple-implicatives.prn](http://web.stanford.edu/group/csli_lnr/Lexical_Resources/simple-implicatives/simple-implicatives.prn)) manifestent un même comportement inférentiel que les verbes analysés ci-dessus. Par exemple tout un groupe de verbes liés au concept général de *pousser quelqu'un ou quelque chose à faire quelque chose qu'il ne voudrait pas faire naturellement* dont l'exemple canonique est *forcer à* se comporte comme *forcer à*. Y appartiennent p. ex. les verbes *amener à, conduire à, contraindre, mener à, pousser, veiller à*, etc.

Ce qui est intéressant de voir, c'est que dans certains cas l'attribution à une catégorie d'implication n'était pas évidente et soulevait des doutes chez les chercheurs et signalés par un astérisque. Cf. p. ex. les cas de, pour rester au sein du groupe qui était originellement le groupe de *forcer*, donc le groupe de verbes censés être à implication unidirectionnelle défini au début comme (+ + // - +/-) : *choisir de* (comme dans : *Ed a choisi de fermer l'usine*), *confirmer que* (comme dans : *Ed confirme que Mary est arrivée*), *reconnaître que* (comme dans : *Ed a reconnu*

que *Mary avait raison*) ou *être influencé pour* (comme dans : *Ed a été influencé pour voter pour Bush*). Une analyse détaillée de ces cas reste à faire, mais il paraît de prime abord que la clé du problème réside dans la prosodie.

Le calcul des inférences à partir de verbes implicatifs a été déjà, comme le remarque Karttunen (2015 : 301), mis en œuvre dans Nairn, Condoravdi & Karttunen (2006). Il serait bien qu'on le passe en revue en fonction des analyses ci-dessus montrant le rôle de la prosodie dans le calcul des inférences, et ce sera aussi une autre tâche à accomplir.

Naturellement, l'un des buts de ce type d'analyse est d'aider l'application de ces connaissances et de cette conscience dans les systèmes automatiques de la génération de la parole, d'une part, quant à la génération des inférences, et d'autre part, de la génération de la parole qui serait capable de rendre compte des nuances prosodiques qui influencent les inférences dont nous avons parlé, ainsi que l'ordre des mots, surtout dans les langues à l'ordre des mots dit « libre », qui va de pair avec, et qui ensemble sont révélateurs de telle ou autre structure informationnelle de l'énoncé, indiquant ce qui est « at-issue » et ce qui en est une pré-condition, et ceci non seulement dans, pour ainsi dire, et ce qui est évidemment exagéré, « simple » synthèse du texte à la parole (« texte-to-speech synthesis » [TTS]), mais encore dans le cas d'une génération de la parole dialogique, au cours d'une conversation homme-machine. À l'époque de l'IA l'essor révolutionnaire des larges modèles de langue, il y a plus ou moins un an, avec le ChatGPT 4-e génération, le Bard de Google, le Bing de Microsoft, le Claude 2 de Anthropic, pour ne citer que les assistants les plus répandus actuellement, ces questions sont parmi les plus pressantes.

Bien que la synthèse vocale se soit beaucoup améliorée ces dernières années et puisse désormais produire une parole très réaliste, il est encore difficile pour elle de restituer toute la richesse expressive et les nuances prosodiques de la voix humaine. Les variations de ton, de rythme et d'accentuation qui permettent à notre voix de transmettre des émotions, des intentions précises, l'intonation porteuse du focus/rhème et influençant les inférences à tirer des énoncés produits demeurent encore un défi pour les systèmes de synthèse vocale les plus sophistiqués. Malgré les grands progrès accomplis, reproduire le naturel, l'expressivité complexe de la parole humaine et le jeu de la prosodie avec la structure informationnelle reste encore un objectif à atteindre pour les technologies de synthèse vocale.

Il est évident que c'est une tâche extraordinairement difficile. La création d'un modèle pour générer les accents toniques en fonction de la structure sémantico-pragmatique du texte est très difficile en raison, entre autres, de la variation potentielle des schémas accentuels. Les accents toniques peuvent varier en fonc-

tion du contexte, dans un dialogue où il y a un va-et-vient perpétuel de ce qui est donné, de ce dont on parle, de ce qui est nouveau et de la prosodie qui les représente. De plus, même dans un contexte identique, il peut y avoir une variation significative entre les locuteurs. Les modèles existants ne peuvent donc représenter, pour l'instant, que les accents toniques probables.

Il y a naturellement des travaux très prometteurs en cours sur une incorporation opérante de la prosodie dans les systèmes automatiques de la synthèse vocale (cf. p. ex. Lacheret-Dujour & Morel, 2011 ; Martin, 2019 ; Bertrand *et al.*, 2022 ; Sloan *et al.*, 2022 ; Pamisetty & Murty, 2023 ; Xin *et al.*, 2023).

La recherche sur la synthèse vocale a connu aussi une progression radicale grâce à l'apprentissage profond et les réseaux neuronaux comme modèles génératifs de la synthèse vocale. Les systèmes actuels, basés sur des réseaux neuronaux, produisent une parole synthétique qui est de plus en plus naturelle et intelligible, qui est aussi de plus en plus expressive et plus avancée que jamais (cf. p. ex. Hirose & Tao, 2015 ; Ning *et al.*, 2019 ; Sini *et al.*, 2022 ; Ren *et al.*, 2020 ; Nuthakki *et al.*, 2023 ; Kumar, Koul & Singh, 2023 ; Tan, 2023).

Pour que la synthèse vocale dialogique puisse rendre compte des relations entre la prosodie, les inférences et la structure informationnelle des énoncés que nous avons présentées ci-dessous, elle doit être de qualité réellement proche de la qualité discursive humaine, et même si certains chercheurs, p. ex. Li *et al.* (2018) déclarent que le système qu'ils proposent a déjà atteint ce niveau, tout efficace qu'il est, ce n'est pas encore le cas. Néanmoins, une collaboration encore plus renforcée entre linguistes de divers domaines et informaticiens, appuyée par les avancées récentes et futures en intelligence artificielle devrait donner des résultats satisfaisants.

## Références citées

- Abeillé, A. & Bonami, O. (2020). *Constraint-based Syntax and Semantics: Papers in Honor of Danièle Godard*. The University of Chicago Press.
- Abrusán, M. (2016). Presupposition cancellation: explaining the “soft–hard” trigger distinction. *Natural Language Semantics* 24(2), 165–202.
- Abrusán, M. (2022). Presuppositions. Dans D. Altshuler (éd.), *Linguistics meets Philosophy* (470–501). Cambridge University Press.
- Abusch, D. (2009). Presupposition triggering from alternatives. *Journal of Semantics* 27, 37–80.



- Ajdukiewicz, K. (1956). Okres warunkowy a implikacja materialna. *Studia Logica: An International Journal for Symbolic Logic* 4, 117–153.
- Lewis, C. I. (1917). The Issues Concerning Material Implication. *The Journal of Philosophy, Psychology and Scientific Methods* 14(13), 350–356.
- Altshuler, D. (éd.) (2022). *Linguistics meets Philosophy*. Cambridge University Press.
- Anscombe, J.-Cl. & Ducrot, O. (1983). *L'argumentation dans la langue*. Pierre Mardaga.
- Arnold, J. E. et al. (2013). Information structure: linguistic, cognitive, and processing approaches. *Wiley interdisciplinary reviews. Cognitive science* 4(4), 403–413.
- Atlas, J. D. (2005). *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface*. Oxford University Press.
- Avanzi, M., Simon, A. C., & Post, B. (2016). La prosodie du français : accentuation et phrasé. *Langue française* 3(191), 5–10.
- Baglini, R. B. & Francez, I. (2016). The Implications of Managing. *Journal of Semantics* 5(33), 541–560.
- Banyś, W. (1989). *Théorie sémantique et si... alors*. Wydawnictwo Uniwersytetu Śląskiego.
- Banyś, W. (2000). *Système de « si » en français moderne. Esquisse d'une approche cognitive*. Wydawnictwo Uniwersytetu Śląskiego.
- Banyś, W. (2020). Pragmatèmes au pays de la prosodie. *Neophilologica* 32, 89–117.
- Banyś, W. & Karolak, S. (éds) (1988). *Structure thème-rhème dans les langues romanes et slaves*. Ossolineum.
- Battistella, E. (1990). *Markedness. The evaluative superstructure of language*. State University of New York Press.
- Bauerle, U. Reyle, U. & Zimmermann, T. E. (éds) (2010). *Presuppositions and discourse: essays offered to Hans Kamp*. Elsevier.
- Beaver, D. (2010). Have you noticed that your belly button lint colour is related to the colour of your clothing?. Dans U. Bauerle, U. Reyle & T. E. Zimmermann (éds), *Presuppositions and discourse: essays offered to Hans Kamp* (65–99). Elsevier.
- Beaver, D. & Clark, B. (2008). *Sense and sensitivity: how focus determines meaning*. Wiley-Blackwell.
- Beaver, D. et al. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics* 3, 265–284.
- Beaver, D. I., Geurts, B. & Denlinger, K. (2021). Presupposition. Dans E. N. Zalta (éd.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Benz, A. & Jasinskaja, K. (2017). Questions Under Discussion: From Sentence to Discourse. *Discourse Processes* 54(3), 177–186.
- Bertrand, R. et al. (2022). La prosodie au Laboratoire Parole et Langage : histoire, recherches actuelles et perspectives. *TIPA. Travaux interdisciplinaires sur la parole et le langage* 38, 1–40.

- Beyssade, C. *et al.* (2015). Prosodic Realizations of Information Focus in French. Dans L. Frazier & E. Gibson (éds), *Explicit and Implicit Prosody in Sentence Processing. Studies in Honor of Janet Dean Fodor* (39–61). Springer International Publishing.
- Bianchi, C. (éd.) (2004). *The semantics/pragmatics distinction*. CSLI Publications.
- Biglari, A. & Bonhomme, M. (éds). (2018). *La Présupposition entre théorisation et mise en discours*. Garnier.
- Bocci, G., Bianchi, V. & Cruschina, S. (2020). Focus in wh-questions. *Natural Language & Linguistic theory* 39(2), 405–455.
- Bogusławski, A. (1977). *Problems of the Thematic- Rhematic Structure of Utterances*. PWN.
- Bogusławski, A. (1986). *Analiza zdań warunkowych a problem funkcji semiotycznych*. *Studia semiotyczne* 14–15, 215–224.
- Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press.
- Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse*. Edward Arnold.
- Büring, D. (2016). *Intonation and meaning*. Oxford University Press.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Dans C. N. Li (éd.), *Subject and Topic* (25–55). Academic Press.
- Chen, I.-H., Huang, C.-R. & Politzer-Ahles, S. (2018). Determining the Types of Contrasts: The Influences of Prosody on Pragmatic Inferences. *Frontiers in Psychology* 9, 1–9.
- Chungmin Lee, Ch., Gordon, M. & Büring, D. (éds). (2007). *Topic and Focus. Cross-linguistic Perspectives on Meaning and Intonation*. Springer.
- Cole, P. (1978). *Syntax and Semantics 9: Pragmatics*. Academic Press.
- Coleman, L. (1975). The case of the vanishing presupposition. *Proceedings of the first Annual Meeting of the Berkeley Linguistics Society* 1, 78–89.
- Cruschina, S. (2022). Focus and Focus Structures in the Romance Languages. *Oxford Research Encyclopedia of Linguistics* (1–48). Oxford University Press.
- Dalrymple, M., Lowe, J. J. & Mycock, L. (2019). Prosodic structure. Dans M. Dalrymple, J.J. Lowe & L. Mycock, *The Oxford Reference Guide to Lexical Functional Grammar* (395–435). Oxford University Press.
- Dargnat, M., Bartkova, K. & Jouvét, D. (2015). Discourse particles in French: Prosodic parameters extraction and analysis. Dans A. H. Dediu, C. Martín-Vide, R. Mitkov & B. Truthe, B. (éds), *Statistical Language and Speech Processing* (40–49). Springer Berlin Heidelberg.
- Dargnat, M. & Jayez, J. (2020). Presupposition Projection and Main Content. Dans A. Abeille & O. Bonami (éds), *Constraint-based Syntax and Semantics: Papers in Honor of Danièle Godard* (101–126). Center for the Study of Language and Information.
- Davidson, D. & Harman, G. (éds) (1975). *The Logic of Grammar*. Dickenson.

- Dediu, A.-H., Martín-Vide, C. & Vicsi, K. (éds) (2015). *Statistical Language and Speech Processing III*. Springer.
- Degen, J. & Tonhauser, J. (2022). *Are there factive predicates? An empirical investigation*. *Language* 98, 552–591.
- Delais-Roussarie, E. *et al.* (2015). Intonational phonology of French: Developing a ToBI system for French. Dans S. Frota & P. Prieto (éds), *Intonation in Romance* (63–100). Oxford University Press.
- Dendale, P., Vanderheyden, A. & Schuring, M. (2022). *Visiblement*, adverbe de phrase évidentiel ou *visiblement*, adverbe de constituant exprimant la manière. La prosodie comme nouveau critère. *Langages* 227, 57–78.
- Di Cristo, A. (1998). Intonation in French. Dans D. Hirst & A. Di Cristo (éds), *Intonation Systems. A Survey of Twenty Languages* (195–218). Oxford University Press.
- Di Cristo, A. (2000). Vers une modélisation de l'accentuation du français (seconde partie). *Journal of French language studies* 10(1), 27–44.
- Di Cristo, A. (2013). *La Prosodie de la parole*. De Boeck-Solal.
- Djäv, K. & Bacovcin, H. A. (2017). *Prosodic effects on factive presupposition projection*. *Journal of Pragmatics* 169, 61–85.
- Ducrot, O. (1972). *Dire et ne pas Dire*. Hermann.
- Ducrot, O. (1984). *Le Dire et le Dit*. Éditions de Minuit.
- Erdmann, P. (1974). Factive, implicative verbs and the order of operators. *Studia Linguistica* 28(1), 51–63.
- Eschenberg, A. (2007). Polish Narrow Focus Constructions. Dans C. Lee, M. Gordon & D. Büring (éds), *Topic and Focus: Cross-linguistic perspectives on meaning and intonation* (23–40). Springer Netherlands.
- Fasciolo, M. (2012). Présuppositions immédiates vs Présuppositions médiates. *Langages* 186(2/2012), 37–51.
- Féry, C. (2017). *Intonation and Prosodic Structure*. Cambridge University Press.
- Fillmore, Ch. (1975). An alternative to checklist theories of meaning. Dans C. J. Fillmore *et al.* (éds), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* (123–131). University of California.
- Fintel, K. V. (1999). *NPI Licensing, Strawson Entailment, and Context Dependency*. *Journal of Semantics* 16, 97–148.
- Frazier, L. & Gibson, E. (éds) (2016). *Explicit and Implicit Prosody in Sentence Processing. Studies in Honor of Janet Dean Fodor*. Springer International Publishing.
- Gardes Tamine, J. (2018). Chapitre 2. La prosodie. Dans J. Gardes Tamine (éd.), *La grammair : Tome 1 – Phonologie, morphologie, lexicologie* (37–47). Armand Colin.
- Fox, D. & Katzir, R. (2009). On the characterization of alternatives for implicature and for focus. *MIT Working Papers in Linguistics* 60, 101–111.
- Frota, S. & Prieto, P. (éds) (2015). *Intonation in Romance*. Oxford University Press.

- Gorka, E., Lee, C., Gordon, M. & Büring, D. (éds) (2007). *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*. Springer.
- Gotzner, N. (2017). The Mechanisms of Activation and Competitive Inhibition. Dans C. Palgrave Macmillan (éd.), *Alternative Sets in Language Processing* (71–101). Springer.
- Gotzner, N. (2019). The role of focus intonation in implicature computation: a comparison with *only* and *also*. *Natural Language Semantics* 27, 189–226.
- Grindrod, J. & Borg, E. (2019). *Questions Under Discussion and the Semantics/Pragmatics Divide*. *The Philosophical Quarterly* 69, 1–9.
- Grice, P. (1975). Logic and conversation. Dans D. Davidson & G. Harman (éds), *Speech acts* (41–58), Brill.
- Grice, P. (1978). Further notes on logic and conversation. Dans P. Cole (éd.), *Pragmatics*. Vol. 9 of *Syntax and Semantics* (113–127). Academic Press.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.
- Gussenhoven, C. & Chen, A. (éds), (2021). *The Oxford Handbook of Language Prosody*. Oxford Handbooks.
- Hajičová, E., Partee, B. & Sgall, P. (1998). *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*. Kluwer Academic Publishers.
- Harris, R. J. (1974). Memory for presuppositions and implications: a case study of 12 verbs of motion and inception-termination. *Journal of Experimental Psychology* 103, 594–597.
- Hasler, E. *et al.* (2017). A comparison of neural models for word ordering. Dans J. M. Alonso, A. Bugarín & E. Reiter (éds), *Proceedings of the 10<sup>th</sup> International Conference on Natural Language Generation* (208–212). Association for Computational Linguistics.
- Hazlett, A. (2010). *The Myth of Factive Verbs*. *Philosophy and Phenomenological Research*, 80, 497–522.
- Hedberg N. & Sosa J. M. (2008). The Prosody of Topic and Focus in Spontaneous English Dialogue. Dans E. Gorka, C. Lee, M. Gordon & D. Büring (éds) (éds), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*. Springer.
- Hesse, Ch. *et al.* (2022). Testing Focus and Non-at-issue Frameworks with a Question-under-Discussion-Annotated Corpus. *Proceedings of the 13<sup>th</sup> Conference on Language Resources and Evaluation (LREC 2022)*, 5212–5219.
- von Heusinger, K. & Turner, K. (éds) (2006). *Where Semantics Meets Pragmatics*. Elsevier Ltd.
- Hirose, K. & Tao, J. (2015). *Speech Prosody in Speech Synthesis: Modeling and Generation of Prosody for High Quality and Flexible Speech Synthesis*. Springer.

- Hirschberg, J. (2006). Pragmatics and Intonation. Dans L. R. Horn & G. Ward (éds). *The Handbook of Pragmatics* (515–537). Blackwell Publishing Malden.
- Hirst, D. & Di Cristo, A. (éds) (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press.
- Horn, L. R. (2006). The Border Wars: A Neo-Gricean Perspective. Dans K. von Heusinger & K. Turner (éds), *Where Semantics Meets Pragmatics* (21–49). Elsevier Ltd.
- Horn, L. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics* 32, 289–326.
- Horn, L. R. & Ward, G. (éds) (2006). *The Handbook of Pragmatics*. Blackwell Publishing Malden.
- Jeong, S. (2021). Prosodically-conditioned factive inferences in Korean: An experimental study. *Semantics and Linguistic Theory* 30, 1–21.
- Jeong, S. (2020). The Effect of Prosody on Veridicality Inferences in Korean. *Lecture Notes in Computer Science* 12331, 133–147.
- Kaiser, E. & Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition* 94, 113–147.
- Karolak S. (1984). Składnia wyrażen predykatywnych. Dans M. Grochowski, S. Karolak & Z. Topolińska (éds), *Gramatyka współczesnego języka polskiego. Składnia* (11–210). PWN.
- Karolak S. (2007). *Składnia francuska o podstawach semantycznych. T. 1*. Collegium Columbinum.
- Karolak S. & Bogacki K. (1991). Fondements d'une grammaire à base Sémantique. *Lingua e Stile* 26(3), 11–48.
- Karttunen, L. (1971a). Implicative verbs. *Language* 47, 340–358.
- Karttunen, L. (1971b). Some observations on factivity. *Papers in Linguistic* 4, 55–69.
- Karttunen, L. (1973). La logique, des constructions anglaises à complément prédicatif. *Langages* 8(30), 56–80.
- Karttunen, L. & Peters, P. (1979). Conventional implicature. Dans C. K. Oh & D. A. Dinneen (éd.), *Syntax and Semantics 11: Presupposition* (1–56). Academic Press.
- Karttunen, L. et al. (2012). *Simple-implicatives*. Stanford University.
- Karttunen, L. (2012). Simple and Phrasal Implicatives. *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 124–131.
- Karttunen, L. (2015). *From Natural Logic to Natural Reasoning*. *Conference on Intelligent Text Processing and Computational Linguistics*, 295–309.
- Karttunen, L. (2016). Presupposition: What went wrong. Dans M. Moroney, C.-R. Little, J. Collard & D. Burgdorf (éds), *Proceedings of Semantics and Linguistic Theory* (705–731). CLC Publications.

- Karttunen, L. & Cases, I. (2019). Teaching a Neural Network to Reason with Implicatives. *Presentation, CLASP, University of Gothenburg*.
- Kastner, I. (2015). *Factivity mirrors interpretation: The selectional requirements of presuppositional verbs*. *Lingua* 164, 156–188.
- Katz, J. & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from prosodic prominence in English. *Language* 87(4), 771–816.
- Keller, F. & Alexopoulou, T. (2001). Phonology competes with syntax: experimental evidence for the interaction of word order and accent placement in the realization of Information Structure. *Cognition* 79, 301–372.
- Kim, Ch., (2019). Focus. Dans C. Cummins & N. Katsos (éds), *The Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford Academic.
- Kiparsky, P. & Kiparsky, C. (1970). Fact. Dans M. Bierwisch & K. E. Heidolph (éds), *Progress in linguistics* (143–173). Mouton.
- Kleiber, G. (2012). Sur la présupposition. *Langages* 186(2/2012), 21–36.
- Koev, T. (2018). Notions of at-issueness. *Lang. Linguistics Compass* 12, 1–17.
- Köhler, W. (1920). *Gestalt psychology*. Liveright.
- Kumar, Y., Koul, A. & Singh, C. (2023). A deep learning approaches in text-to-speech system: a systematic review and recent research perspective. *Multimedia Tools and Applications* 82, 15171–15197.
- Kurumada, C. & Clark, E. V. (2016). Pragmatic inferences in context: learning to interpret contrastive prosody. *Journal of Child Language* 44(4), 850–880.
- Kügler, F. & Calhoun, S. (2023). Prosodic Encoding of Information Structure: A typological perspective. Dans C. Gussenhoven & A. Chen (éds), *The Oxford Handbook of Language Prosody* (454–467). Oxford Academic.
- Lacheret-Dujour, A. & Morel, M. (2011). Modéliser la prosodie pour la synthèse à partir du texte: perspectives sémantico-pragmatiques. Dans F. Neveu, P. Blumenthal & N. Lequerler (éds), *Au commencement était le verbe: syntaxe, sémantique et cognition: mélanges en l'honneur du Professeur Jacques François* (299–325). Peter Lang.
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Indiana University Press.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge University Press.
- Lacheret-Dujour, A. & Beaugendre, F. (1999). *La prosodie du français*. Éditions du CNRS.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar*. Stanford University Press.
- Lappin, Sh. (éd.) (2015). *The Handbook of Contemporary Semantic Theory*. Blackwell.
- Lee, C., Gordon, M. & Büring, D. (éds) (2008). *Topic and Focus. Studies in Linguistics and Philosophy*. Springer.
- Lee, Ch., Kiefer, F. & Krifka, M. (éds) (2017). *Contrastiveness in Information Structure, Alternatives and Scalar Implicatures*. Springer.

- Léon, P. R., Martin, P. & Baligand, R. A. (1969). *Prolégomènes à l'étude des structures intonatives*. Didier.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Levinson, S. C. (1995). Three levels of meaning, Grammar and Meaning. Dans F. R. Palmer (éd.), *Essays in Honour of Sir John Lyons* (90–115). Cambridge University Press.
- Lewis, C. I. (1917). The Issues Concerning Material Implication. *The Journal of Philosophy, Psychology and Scientific Methods* 14(13), 350–356.
- Li, C. N. (éd.) (1976). *Subject and Topic*. Academic Press.
- Li, N. et al. (2018). Close to Human Quality TTS with Transformer. *ArXiv, abs/1809.08895*.
- Lin, P. (2018a). *The Prosody of Formulaic Sequences. A Corpus and Discourse Approach*. Bloomsbury Academic.
- Lin, P. (2018b). Formulaic language and speech prosody. Dans A. Siyanova-Chanturia & A. Pellicer-Sánchez (éds), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (78–94). Routledge.
- MacCartney, B. & Manning, C. D. (2007). Natural Logic for Textual Inference. *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, 193–200.
- Mahler, T. (2019). Does at-issueness predict projection? It's complicated!. *Proceedings of the 49<sup>th</sup> Meeting of the North East Linguistics Society* 2, 245–255.
- Martin, Ph. (2009). *Intonation du français*. Armand Colin.
- Martin Ph. (2015). *The Structure of Spoken Language. Intonation in Romance*. Cambridge University Press.
- Martin, Ph. (2018). *Intonation, structure prosodique et ondes cérébrales*. ISTE.
- Martin, Ph. (2019). Génération automatique de la structure prosodique en français. *Journal of Speech Sciences* 7(2), 79–91.
- Mertens, P. (2004). Le prosogramme : une transcription semi-automatique de la prosodie. *Cahiers de l'Institut de Linguistique de Louvain* 30(1–3), 7–25.
- Mertens, P. (2008). Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours. *Travaux de Linguistique* 56(1), 87–124.
- Moeschler, J. (2018). Présupposition et implicature. Où passe la frontière?. Dans A. Biglari & M. Bonhomme (éds), *La présupposition entre théorisation et mise en discours* (7–82). Classiques Garnier.
- Munitz, M. K. & Unger, P. K. (éds) (1974). *Semantics and philosophy*. New York University Press.
- Nadathur, P. (2016). Causal necessity and sufficiency in implicativity. *Semantics and Linguistic Theory* 26, 1002–1021.
- Nelson, J. O. (1966). Is Material Implication Inferentially Harmless?. *Mind* 75(300), 542–551.

- Nairn, R., Condoravdi, C. & Karttunen, L. (2006). Computing relative polarity for textual inference. *Proceedings of Inference in Computational Semantics*, 67–76.
- Neveu, F., Blumenthal, P. & Lequerler, N. (éds) (2012). *Au commencement était le verbe : syntaxe, sémantique et cognition : mélanges en l'honneur du Professeur Jacques François*. Peter Lang.
- Ning, Y. *et al.* (2019). A Review of Deep Learning Based Speech Synthesis. *Applied Sciences* 9(19), 40–50.
- Nuthakki, P. *et al.* (2023). Deep Learning based Multilingual Speech Synthesis using Multi Feature Fusion Methods. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 1–17.
- Oh, Ch-K. & Dinneen, D. A. (éds) (1979). *Syntax and Semantics 11*. Academic Press.
- Palmer, F. R. (éd.) (1995). *Essays in Honour of Sir John, Lyons*. Cambridge University Press.
- Pamisetty, G. & Sri Rama Murty, K. (2023). Prosody-TTS: An End-to-End Speech Synthesis System with Prosody Control. *Circuits Syst Signal Process* 42, 361–384.
- Partee, B. H. (2009). Formal Semantics, APPENDIX to Lecture 7: Implicatures, Presuppositions, etc. *MGU April 17, 2009*, 1–2.
- Pavlick, E. & Callison-Burch, C. (2016). Tense Manages to Predict Implicative Behavior in Verbs. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2225–2229.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English Intonation*. Thèse de doctorat, M.I.T.
- Pierrehumbert, J. & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings International Conference of Spoken Language Processing* 2, 867–870.
- Portes, C. & Reyle, U. (2022). Combining syntax and prosody to signal information structure: the case of French. *Proceedings of the Conference on Speech Prosody 2022*, 87–91.
- Post, B. (2000). *Tonal and phrasal structures in French intonation*, Academic graphics.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.
- Recanati, F. (2004). ‘What is said’ and the semantics/pragmatics distinction. *WOC 2002: Semantics and Pragmatics, Oct 2002, Genoa*, 45–64.
- Ren, Y. *et al.* (2020). *FastSpeech 2: Fast and High-Quality End-to-End Text to Speech*. *ArXiv, abs/2006.04558*.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. Dans J.-H. Toon & A. Kathol (éds), *Papers in Semantics* 49 (91–136). The Ohio State University Working Papers in Linguistics.
- Roberts, C. (2012). *Information Structure: Towards an integrated formal theory of pragmatics*. *Semantics and Pragmatics* 5, 1–69.
- Rossi, M. (199). *L'Intonation. Le système du français : description et modélisation*. Ophrys.
- Persson, R. (2017). La prosodie comme ressource pour l'organisation de l'interaction : état des lieux et illustrations. *Revue française de linguistique appliquée* 22(2), 33–52.




- Ronai, E. & Xiang, M. (2021). *Pragmatic inferences are QUD-sensitive: An experimental study*. *Journal of Linguistics* 57(4), 841–870.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.
- Rooth, M. (1996). Focus. Dans Sh. Lappin (éd.), *The handbook of contemporary semantic theory* (271–97). Blackwell.
- Sadock, J. (1978). On testing for conversational implicature. Dans P. Cole (éd.), *Syntax and Semantics 9: Pragmatics* (281–297). Academic Press.
- Sander, T. (2022). Taxonomizing Non-at-Issue Contents. *Grazer Philosophische Studien* 99(1), 1–34.
- Saul, J. (2002). Speaker meaning, what is said, and what is implicated. *Noûs* 36, 228–248.
- Schwarz, F. (2019). Presuppositions, Projection, and Accommodation. Dans C. Cummins & N. Katsos (éds), *The Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford Academic.
- Simon, A. C. (2004). *La structuration prosodique du discours en français: une approche multidimensionnelle et expérimentielle*. P. Lang.
- Simons, M. et al. (2010). What projects and why. *Semantics and Linguistic Theory* 20, 309–327.
- Simons, M. et al. (2017). *The Best Question: Explaining the Projection Behavior of Factives*. *Discourse Processes* 54, 187–206.
- Sini, A. et al. (2022). Techniques de synthèse vocale neuronale à l'épreuve des données d'apprentissage non dédiées: les livres audio amateurs en français. *Traitement Automatique des Langues* 63(2), 41–65.
- Siyanova-Chanturia, A. & Pellicer-Sánchez, A. (éds) (2018). *Understanding Formulaic Language: A Second Language Acquisition Perspective*. Routledge.
- Sloan, R., Adigwe, A., Mohandoss, S. & Hirschberg, J. (2022). Incorporating Prosodic Events in Text-to-Speech Synthesis. *Proceedings of the Conference on Speech Prosody 2022*, 287–291.
- Spalek, K., & Oganian, Y. (2019). The neurocognitive signature of focus alternatives. *Brain and language* 194, 98–108.
- Stalnaker, R. C. (1974). Pragmatic presuppositions. Dans M. K. Munitz & P. K. Unger (éds), *Semantics and philosophy* (197–213). New York University Press.
- Strawson, P. F. (1966). *The Bounds of Sense*. Methuen.
- Strawson, P. F. (1971). *Logico-Linguistic Papers*. Methuen.
- Tan, X. (2023). *Neural Text-to-Speech Synthesis*. Springer.
- Teh, T. H. et al. (2023). Ensemble Prosody Prediction for Expressive Speech Synthesis. *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.
- Toon, J.-H. & Kathol, A. (éds) (1996). *Papers in Semantics* 49. The Ohio State University Working Papers in Linguistics.

- Tonhauser, J. (2019). Prosody and Meaning. Dans Ch. Cummins & N. Katsos (éds), *The Oxford Handbook of Experimental Semantics and Pragmatics* (494–511). Oxford Academic.
- Tonhauser, J. (2016). *Prosodic cues to presupposition projection*. *Semantics and Linguistic Theory* 26, 934–960.
- Tonhauser, J. *et al.* (2013). Toward a taxonomy of projective content. *Language* 89, 66–109.
- Wagner, M. & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes* 25(7–9), 905–945.
- Davis, W. (2019). Implicature. Dans E. N. Zalta (éd.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Wertheimer, M. *et al.* (éds) (1945). *Productive thinking*. Harper and Brothers.
- White, A. S. & Rawlins, K. (2018). *The role of veridicality and factivity in clause selection*. *Proceedings of The 48th Annual Meeting of The North East Linguistic Society*, 221–234.
- Witek, M. & Witczak-Plisiecka, I. (éds) (2019). *Varieties and Dynamics of Speech Actions*. Brill Rodopi.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. Brace & Company, Inc./Kegan Paul, Trench, Trubner & Co., Ltd.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen/Philosophical Investigations*. Basil Blackwell.
- Włodarczyk, M. (2019). Are Implicative Verbs Presupposition Triggers? Evidence from Polish. Dans M. Witek, I. Witczak-Plisiecka (éds), *Varieties and Dynamics of Speech Actions* (206–230). Brill Rodopi.
- Włodarczyk, M. (2021). Processing presuppositions. Are implicative verbs soft triggers?. *Research in Language* 19(1), 47–75.
- Xin, D. *et al.* (2023). Improving Speech Prosody of Audiobook Text-To-Speech Synthesis with Acoustic and Textual Contexts, *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023*, 1–5.
- Zalta E. N. (éd.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Ziembicki, D., Seweryn, K. & Wróblewska, A. (2022). Polish natural language inference and factivity: An expert-based dataset and benchmarks. *Natural Language Engineering*, 1–32.
- Ziembicki, D. (2022). *Lingwistyczna analiza zjawiska faktywności (na materiale współczesnej polszczyzny)*. Université de Varsovie.
- Zuber, R. (1972). *Structure présuppositionnelle du langage*. Dunod.
- Zwanenburg, W. (1965). *Recherches sur la prosodie de la phrase française*. Universitaire Pers Leiden.



**Sibilla Cantarini**

University of Verona  
Italy

 <https://orcid.org/0000-0002-1688-634X>

# Anglicisms in German – An Overview with Reference to the *Neologismenwörterbuch* in the *Online-Wortschatz-* *Informationssystem* *Deutsch* (OWID)

## Abstract

Neologisms indicate linguistic changes in the lexicon. Vocabulary is particularly prone to changes since language communities constantly need new words and phrases to denote new facts or circumstances or they require new expressions for concepts already established. New expressions are often borrowed from other languages. English has had a strong influence on German vocabulary in the last decades. Therefore, German has several English borrowings. These anglicisms differ in their degree of integration into the linguistic system and vary, for instance, with respect to usage, style, and diffusion in the language community. With the help of the *Neologismenwörterbuch*, it is possible to have an overview of neologisms of English origin in German that integrate different areas of language change.

## Keywords

Anglicisms, neologisms, morphological processes, formal integration, semantic change

## Anglicisms

Anglicisms have influenced the German language since the early Middle Ages and especially since the 17th century in several periods and with varying

intensity. The cultural and linguistic influence achieved exceptional growth in the post-World War II period. Particularly from the nineties onwards, English has been diffused into several areas of life. Thus, English vocabulary is part of the everyday language of every German-speaking person today. While some of those lexemes<sup>1</sup> come into use only temporarily and in certain restricted areas or for certain purposes only, for example, in advertising, others enter the German language and, after some time, are not recognizable as anglicisms anymore (cf. Busse, 2001, p. 131 f.) as they have lost their exogenous characteristics and have become an integrated, habitual part of the language system (cf. Burmasova, 2010, p. 218).

This article aims to examine the anglicisms since the nineties, in order to provide an overview of these loan expressions and their characteristics. The data is delivered by the online *Neologismenwörterbuch*. It includes 2,100 new words, phrases, and new meanings that have become part of common German between 1991 and today (*Neologismenwörterbuch* 2006–). The *Neologismenwörterbuch* has been accessible online via the OWID (*Online-Wortschatz-Informationssystem Deutsch*) portal since 2006. It is an expanding dictionary that continuously receives new word articles on current neologisms, neologisms that have been determined retrospectively, new meanings, and new phraseologisms (cf. Müller-Spitzer, 2008, pp. 8–12). Since the OWID portal focuses on general standard language, the new vocabulary in the *Neologismenwörterbuch* does not contain terms strictly belonging to technical jargon or language for special purposes (cf. Steffens, 2008, p. 104; Steffens, 2017, p. 282). It documents all new vocabulary since the nineties, that is, the new words, the new phraseologisms, and the new meanings that became widespread in the general language during a given period. The expressions included are determined on scientific and empirical grounds and based on corpora covering written contemporary language (cf. al-Wadi, 2017, p. 174). To distinguish the neologisms from occasional occurrences of new expressions or meanings and to fulfil the common language crite-

---

<sup>1</sup> Lexemes are linguistic units with independent lexical meaning. This includes all inflected variants of these units, e.g., [*der*] *Mann*, [*des*] *Mannes*, [*dem*] *Manne*, [*die*] *Männer* (Kessel & Reimann, 2010, p. 61). Inflected variants are attributed to lemmas that are regarded as basic forms of words (Lemnitzer & Zinsmeister, 2015, p. 65). Words constitute smaller units than word groups or phrases. However, fixed formations of words with a lexicalized meaning, such as *Antwort geben* have closed semantics; thus, such word groups are called lexemes, too. Whereas word and lexeme meaning correspond to each other, the phrasal meaning of word group lexemes is not necessarily composed of the individual word meanings in the phrase (cf. Herberg, 1980, p. 143 f.). Particularly in phraseologisms such as *jemanden in die Pfanne hauen* or *sich etwas durch die Lappen gehen lassen*, the meaning of the whole phrase cannot be derived from the meanings of the individual lexical units (cf. Haspelmath, 2011, p. 36).

tion, the entries in the dictionary must show a certain distribution in the corpora they were drawn from (cf. Steffens, 2017, p. 284 f.).

The degree of integration into German usage can be determined, among other things, by the integration into the norms of the German language. Accordingly, new lexemes occurring in the German language are included in the dictionary. They can be ascertained by the fact that they are found with a certain frequency and in various texts or sources. Particularly in the case of anglicisms, the degree of adaptation to the norms can be observed in various areas, such as pronunciation, spelling, and inflexion. For instance, the German noun *Waveboard* is pronounced [ˈve:fbɔ̃t] instead of English [ˈweɪvbɔ:d]. Furthermore, the German word is adapted to German spelling with capitalisation and the compounding of two nouns, and fits into the inflectional system by using the genitive *-s* as in *des Waveboards*. In this lexical area of anglicisms, it is also apparent that quotation marks, italics, or similar boundary signals marking a word as new and (yet) unusual are being dropped over time (cf. Steffens & al-Wadi, 2015, p. XIV f.).

In addition, the neologisms include pseudo-anglicisms. These are German formations appearing as if they were of English origin, for example, *Gigaliner* or *Handy* (cf. Steffens & al-Wadi, 2015, p. XV). On the one hand, the dictionary labels pseudo-anglicisms and, on the other hand, it points out the source of the expressions, for instance, that the pseudo-anglicism *Servicepoint* is translated into English with “information desk” (cf. al-Wadi, 2013, p. 22).

The extended search in the *Neologismenwörterbuch* includes several characteristics like time of appearance, the origin from other languages, the origin by type (loanword, semantic loan, loan translation or interpretable as such, partial loan translation, loan transfer, pseudo-anglicism and translation of a loan word). The word formation section distinguishes different types, that is, different compounds and derivations, abbreviations or particle verbs (*Präverbfügungen*), between different constituents, that is, different parts of speech as constituents of compounds, linking elements (*Fugenelemente*), affixes and abbreviations, as well as between different kinds of word formation productivity, that is, in compounds, derivations, abbreviations or particle verbs (cf. *Neologismenwörterbuch 2006–: Erweiterte Suche*).

To analyse anglicisms in the *Neologismenwörterbuch* a categorization is necessary. The categorization applied here is guided by criteria formulated by Busse (2001) and Burmasova (2010), briefly explained in the following section.

## Criteria

### Formal Criteria

#### *Partes orationis*

It is often stated that borrowed lexical units mostly belong to the parts of speech (*partes orationis*): nouns, verbs, adjectives, and adverbs. In her investigation, Burmasova (2010, p. 217 f.) could not attest to adverbs and, following functional aspects, assigned participles either to the category of verbs or to the category of adjectives. One result of the comparison of anglicisms in two German newspaper corpora from 1994 and 2004 was a significant increase in both the number of borrowed nouns and verbs, while the variance changed only slightly for both categories. The variance indicates that the use of nominal or verbal anglicisms did not change significantly or, to put it differently, the increase in numbers was not equivalent to an increase in use. The analysis of the adjectives in the corpora showed that adjectives were in use only sporadically. These results in growth rate concerning nouns, verbs, and adjectives contradict the common opinion that nouns are borrowed far more often than other parts of speech, an assumption relying on absolute figures only without consideration of growth rate and variance (cf. Burmasova, 2010, pp. 223–227).

#### *Type of Borrowing*

While Burmasova concentrates on the categories “direct borrowings,” “loan word formations” (“pseudo-borrowings”), “abbreviations,” and “mixed compounds” with both German and English components (cf. Burmasova, 2010, p. 217), Busse (2001, pp. 135–141) deals with “direct borrowings,” “substitutions and loan translations,” “partial substitutions,” “pseudo-anglicisms,” “reborrowings” and “internationalisms” (or “eurolatinisms”). The following sections give a short sketch of those concepts.

- *Direct Borrowings*

Direct borrowings probably constitute the type of borrowing occurring most often. This includes simple and compound words like *Babysitter*, *Bowling*, and *Engineering* or elements like *mini-* or *super-*. Some of those direct loans were adopted via further languages, especially French (e.g., *Budget* or *Jury*). Some of the direct loans are still recognizable as anglicisms because of their orthography and/or pronunciation, (e.g., *Bowling* or *Job*), while others are fully integrated into

the norms of German language use (e.g., *Boot* [bo:t]) and, what is more, form different derivations like *Test* > *testen*, *Tester* and compounds like *Testverfahren*. It is not necessarily the case that only older borrowings belong to the fully integrated end of the continuum and only younger ones are to be found at the partly integrated end. The morphological, phonological, and/or orthographic integration of borrowings and their degree of common acceptance can undergo different stages and, over time, this process can be accompanied by several variants, especially with regard to pronunciation but also in their morphological integration, cf. *Teste*, the more seldom plural variant of *Test* (cf. Busse, 2001, p. 135 f.).

- *Substitutions and Loan Translations*

Substitutions reproduce English expressions, for instance, simple and compound words like *Flutlicht* < *flood light*, phrases like *stehende ovationen* < *standing ovations*, *in einem Boot sitzen* < *to be in the same boat* and word formation elements like *-sicher* < *-proof*. Besides such morpho-semantic borrowings, substitutions and loan translations can also include semantic borrowings resulting in new meanings of existing words (e.g., *realisieren* < *realize*) (cf. Busse, 2001, p. 136 f.).

- *Partial Substitutions*

Partial substitutions are nominal compounds like *Campingplatz* < *camping site* or *Nonstopflug* < *non-stop flight* and particle verbs (e.g., *einchecken*). It might be possible to distinguish cases with an English equivalent, for example, *Computerspiel* < *computer game* that probably has been borrowed as a whole, from cases like *Managerkrankheit*, thus loan word formations revealing the productivity of the original English part that do not adapt an English example. The decision in those cases depends on the documentation of lexical change in dictionaries. A further category considers derivations with an original German element, for instance, *Computerei*, *stressig* (cf. Busse, 2001, p. 137 f.).

- *Pseudo-Anglicisms*

Pseudo-Anglicisms do not have formal or semantic English equivalents but are German coinages formed with English means of word formation, for example, *Dressman* (Engl. *male model*), *Handy* (Engl. *mobile* or *cellular phone*). Some of those lexemes are formed in analogy to formerly borrowed English loans, for instance, *Showmaster* (from *Quizmaster*) or *Pullunder* (from *Pullover*) (cf. Busse, 2001, p. 138 f.).

Among the possible morphological operations are abbreviations of simple words like *Profi* < *Professional* (Engl. *professional* > *pro*), short forms of compounds like English *cocktail party* > German *Cocktailparty* > German *Cocktail*, and short forms of phraseologisms or multi-word lexemes like English *last but not least* > German *last but not least* > German *last not least* (cf. Busse, 2001, p. 139).

- *Reborrowings*

Reborrowings usually involve semantic change. They are attestations of the cultural and lexical exchange between language communities. However, they do not seem to occur very often. German examples are *Hamburger* < *hamburger* “steak” < *Hamburger* “inhabitant of Hamburg” and *Superman* “superhero” < *superman* < *Übermensch* after the specific use by philosopher Friedrich Nietzsche (cf. Busse, 2001, p. 140, OED Online: “Hamburger, n.”, “superman, n.”).

- *Internationalisms or Eurolatinisms*

There are some lexical or word formation units, mostly of Ancient Greek or Latin origin, that are widespread in many languages, like *Administration*, *Global*, *Telegram*, *super-*, and *ultra-*. Formally, they do not seem to have an English source. However, it is possible to classify such expressions as anglicisms if English influence initiated the revitalisation of a former usage or the usage frequency has increased. It can be noticed that internationalisms having entered the German language through English usually retain their spelling but become integrated in phonological and phonetic respects (cf. Busse, 2001, p. 140 f.). Burmasova (2010, p. 216) also argues that the more distant etymology should not be decisive in determining an anglicism but the immediate source language from which a lexical unit was borrowed.

## Semantics

It cannot be expected that the entire semantic range of meanings of English is transferred in the course of borrowing. This is what a comparison of the word meanings of anglicisms with their English-language models reveals. Rather, usually, one meaning for a specific context is borrowed. In the case of multiple borrowings, different borrowings of the same language sign come along with borrowings of different meanings at different points in time, for example, *Twist* “darning



thread” and “dancing style.” In other cases, anglicisms, having gained a foothold in German, are influenced by their semantic field and develop meanings they did not have originally in English. Those anglicisms develop semantics of their own independently from the original language and are subject to the same conditions of language change as German lexical units (cf. Busse, 2001, p. 139).

### Sociolinguistic and Pragmatic Aspects

As the examples so far illustrate, formal integration into the linguistic system is a process during which borrowings are distributed across the language community and incorporated into usage. Thus, it is not necessary that a borrowing is fully integrated into the linguistic system to be classified as a common expression. Borrowings may receive German linguistic characteristics before they become usual expressions, whereas widespread expressions can still exhibit formal exogenous characteristics of their original language, especially in the areas of phonology or morphology. For this reason, from a synchronic perspective, a decision on the integration degree should take into account sociolinguistic and pragmatic aspects (cf. Burmasova, 2010, p. 219).

The degrees of distribution and acceptance of lexical borrowings can be approached by considering their use and their possible restrictions to certain areas or contexts. A typology of anglicisms should take into account their usage by speakers and writers to come to an estimate of their (sociolinguistic and pragmatic) status within the language system. Their usage can be described by their frequency, their acceptability, and their social and stylistic distribution. A combination of acceptance, frequency, and formal integration can result in different categories.

Category 0: The least degree of distribution and acceptance can be supposed only if speakers with a high command of both German and English can understand the English word and hence use it as a quotation in code-switching settings. In other cases of this category, the English words are known but only applied to refer to foreign circumstances, objects, or roles that cannot easily be transferred to situations in German-speaking countries and therefore retain the form they bear in the English-speaking world. For instance, this holds for *Detective Superintendent* in crime novels, *College* or *Earl*. Examples like (ironizing) *Sheriff* or *-gate* (after Nixon's *Watergate*) illustrate that such exotisms can be transferred to German and the cultural conditions of the language community (cf. Busse, 2001, p. 141 f.). Thus, while expressions in category 0 are not to be considered as a fixed part of the German linguistic system, they may enter the German

language in particular contexts. In general, a more regular use of borrowings with a lower variance across different contexts indicates that the sociolinguistic integration process moves forward. In contrast, a high variance in a new expression's occurrence suggests that it is used only once in a while (cf. Burmasova, 2010, p. 221 f.).

Expressions in categories 1 and 2 are restricted in usage with respect to style, frequency, group languages or registers, for example, advertising, youth language, oral vs. written language, or certain professional groups. Words in category 2 can be distinguished by their acceptance and a broader diffusion in different contexts and stylistic levels. Nevertheless, their English origin is still recognizable in their form, that is, their spelling, pronunciation, or morphology (e.g., *Jeans*, *Thriller*) (cf. Busse, 2001, p. 141 f.). With regard to spelling, pronunciation, and/or morphology, words in category 3 are not recognisable as being of English origin anymore (e.g., *Boot*, *Sport*, *Humor*, *Pudding*, *Rum*). Words in category 4 are formally identical to established German words, only that their semantics have been transferred from English, for instance, *Maus* ("Arbeitsmittel am Computer") (cf. Busse, 2001, p. 143).

Busse (2001, p. 143) points out some specific features of anglicisms in German. Firstly, their distribution across the lexicon is not consistently even. Secondly, the majority can be regarded as terminology of languages for special purposes and, except for data processing and communication technology, they are not widespread, not fully integrated, not stylistically marked and are more likely to be found in written language. Thirdly, another group of anglicisms in German are vernacular. They are also restricted to certain areas, namely youth language, journalism, and advertising. They are more likely to be found in spoken language and, although their semantics are vague for many speakers, they are prestigious.

According to Burmasova, it is hard to assign anglicisms to specific subject areas since many expressions are part of the general vocabulary, polysemous (cf. Burmasova, 2010, p. 217), or are not specific to certain subject areas (cf. Burmasova, 2010, p. 223). However, the distribution by topic is not pointless. It may demonstrate whether anglicisms are concentrated in certain subject areas or if they are evenly distributed across all areas. This can allow for assumptions about the course of their integration process (cf. Burmasova, 2010, p. 217). Burmasova, for instance, could observe an increase in nearly all topic areas (Politics, Economics, Sports, Culture, Science and Technology, and Society) except the section Science and Technology (cf. Burmasova, 2010, p. 223 f.).

## Anglicisms in the *Neologismenwörterbuch*

The *Neologismenwörterbuch* lists 740 anglicisms that have entered the German language since 1990. The anglicisms cover several topic areas, such as Communication on the Internet or with an Electronic Device (*adden, Follower, Internet-mobbing, Youtuber*), Economics and Finance (*Assessmentcenter, Crowdfunding, outsourcen, Turn-around*), Education (*Telelearning, Teleteaching*), Fashion, Food, and Beverages (*Alcopop, Frozen Yogurt, Slow Brewing, Wrap*), Health (*bleachen, Bracket*), Love and Sexuality (*Blind Date, Cybersex, Mingle, Womanizer*), Movies and TV (*Anchor, Infotainment, Scripted Reality*), Politics (*Brexit, BRICS, Gender-Mainstreaming, Landgrabbing*), Sports (*Aquajogging, nordicwalken, raften, walken*), Technology (*Bioprinter*), Transport (*Skymarshal, SUV*) etc. Some entries (e.g., *XL, XXL*) are not restricted to certain topic areas but are applicable in different contexts. In the case of *XL* and *XXL*, this is the result of a generalisation process from the meaning “Konfektionsgröße” to the meaning “Größe.” For this reason, the dictionary lists compounds like *XL-Bratwurst, XL-Sonnenbrille, or XXL-Koalition*.

Among the new lexemes, there are no phraseologisms but only single and compound words. 348 of the anglicisms emerged in the nineties, 215 during the first decade of the 21st century, and a further 177 during the second decade. The anglicisms include the six short-time expressions *Ins, Outs, Showview, Skymarshal, Vatileaks, and WAP*. While borrowings and pseudo-anglicisms contribute to the expansion of the lexicon, short-time words fall into disuse after a short while and are not involved in the expansion of the lexicon (cf. Steffens, 2017, p. 275).

### *Partes orationis* in the *Neologismenwörterbuch*

With 659 entries out of 740, the majority of the anglicisms in the dictionary have nominal status. In addition, there are 59 verbs, 12 adjectives, 5 adverbs, and 2 interjections. 3 entries do not have grammatical status at all: *at* (“Gliederungszeichen”), *XL, XXL*. Thus, on the one hand, the borrowings’ syntactic status goes beyond nouns, verbs, and adjectives. On the other hand, the often-cited finding can be confirmed that, in absolute figures, nouns form the majority of borrowings from English.

## Types of Borrowing in the *Neologismenwörterbuch*

Among the new expressions, 669 expressions can be classified as loan words (e.g., *Attachment*, *Clickbait*, *Hoodie*, *Scripted Reality*, or *Sitcom*). In the dictionary, the phrase “loan word” from English is understood as a word borrowed from English that has been adapted to the German linguistic system and may still bear features of the source language. In addition, the anglicisms include five new meanings: *Exit*, *Framing*, *ghosten*, *Pushback*, and *Update*. For instance, the verb *ghosten* has the new meaning “eine Beziehung oder Freundschaft plötzlich beenden, indem man sich aus jeglicher Kommunikation zurückzieht.” According to the dictionary entry, this semantic loan emerged under the influence of English *to ghost* used in its most recent meaning. New meanings, that is, the new semantics of a lexeme already established, may be accompanied by notes on the older or former meaning. In the case of a loan word and in the case of a new meaning to an established lexeme, the immediate language of origin is usually given for the entry. The English words on which the loan words are based are usually traced back to their basis and/or constituents and provided with their semantics (cf. *Neologismenwörterbuch* 2006–: *Benutzerhinweise/Herkunft*).

Furthermore, there are some entries that were formed in German and are presented as such in the dictionary, for instance, *Blogger*, a derivation from the verbal anglicism *bloggen*. As there is also an English noun *blogger* with the same meaning, there is a note in the word formation section of the entry *Blogger* in the dictionary that *Blogger* can be interpreted as a loan word. An analogous case would be *Tabatatraining* after English *Tabata training* (cf. *Neologismenwörterbuch* 2006–: *Benutzerhinweise/Wortbildung*).

There are no dictionary entries explicitly classified (and searchable in the extended search) as loan translations, partial loan translations, loan transfers, or translations of a loan word. The reason might be that the exact status of lexemes likely to be categorised as loan translation, partial loan translation, loan transfer, or translation of a loan word is not clear yet and needs further observation. However, there are some examples of possible loan translations among the words classified as anglicisms, for instance, *Samenbombe* formed after *seed bomb*, *zwitschern* after *twittern* or *to twitter* or *Biodrucker*, which, according to the dictionary, is translated from English *bioprinter*. Other loan translations are not to be found among the anglicisms in the dictionary, for example, *Helikoptereltern* which however is explicitly described as being translated from English *helicopter parents*.

Although strictly speaking, pseudo-anglicisms do not belong to the anglicisms among the German neologisms, they are related to them because the lexical or

semantic source material drawn back to is of English origin (cf. Steffens, 2017, p. 277). In German usage, the borrowed material has a new meaning or receives an additional meaning (cf. *Neologismenwörterbuch 2006–: Benutzerhinweise/Bedeutung und Verwendung*). The pseudo-anglicisms in the dictionary (e.g., *Bowl*, *Call-in*, *Factoryoutlet*, *flashen*, *Onliner*, *Outlet*, *Pad*, *Slackline*, *Speeddating*) are classified as words of English origin, as pointed out above. For instance, according to the dictionary, *Best Ager* denotes a middle-aged person with purchasing power and physical and mental strength in German. The noun emerged in the first decade of the 21st century. In English, however, the appropriate expression would be “consumer older than 40” (cf. *Neologismenwörterbuch 2006–: Best Ager*).

### Morphological Processes and Integration

As far as word formation and word formation productivity are concerned, 520 anglicisms form compounds, either as a first (e.g., *Alcopop* > *Alcopopflasche*, *Alcopopgetränk*, *Alcopopkonsum*, *Alcopopsteuer*, *Buzzer* > *Buzzerknopf*, *Buzzerkönig*, *Gameshow* > *Gameshowkandidat*, *Gameshowmoderator*) or as a second component (*Anchorman* > *ARD-Anchorman*, *RTL-aktuell-Anchorman*, *Tagesthemen-Anchorman*, *Banking* > *Direktbanking*, *Internetbanking*, *Onlinebanking*, *Telebanking*, *Telefonbanking*, *Morphsuit* > *Lycramorphsuit*, *Frosting* > *Frischkäsefrosting*, *Schokoladenfrosting*) or both (*Account* > *Accountdaten*, *Accountinhaber*, *Nutzeraccount*, *Uni-Account*, *E-Mail-Account*, *Downlight* > *Downlightelement*, *Downlightkonzept*, *Einbaudownlight*, *Halogendownlight*, *LED-Downlight*, *Stalking* > *Stalkingbeauftragte/r*, *Stalkingforschung*, *Stalkinggesetz*, *Stalkingopfer*, *Stalkingtäter*, *Internetstalking*, *Telefonstalking*). Among the entries, 101 form as base words explicit derivations (e.g., *chillen* > *chillig*, *Couchsurfer* > *Couchsurferin*, *inlineskaten* > *Inlineskater*, *nordicwalken* > *Nordicwalker*, *Nerd* > *Nerdette*, *nerdig*, *Nerdin*, *Nerdine*, *nerdmäßig*, *Nerdtum*, *Spa* > *Superspa*). Only 6 entries can (also) serve as the base of a conversion, that is, the word formation process changing the syntactic category without a word formation marker (cf. Elsen, 2011, p. 29) or a derivational affix (cf. Baeskow & Müller, 2014): *Doodle* > *doodeln*; *MMS* > *mimsen*; *Gender* > *gendern*, *SMS* > *simsen*, *uploaden* > *Upload*, *metrosexuell* > *Metrosexueller*. There are 8 entries in the dictionary that can form abbreviations: *Baseballcap* > *Basecap*, *Daily Soap* > *Daily*, *Daily Talk* > *Daily*, *Factoryoutletcenter* > *Factoryoutlet*, *FOC*, *Outlet*, *Flatrate* > *Flat*, *French Nail* > *French*, *Inlineskate* > *Inline*, *Transgender* > *TG*. Finally, 24 entries form particle verbs, for example, *casten* > *nachcasten*, *wegcasten*, *preppen* > *drauflospreppen*, *scrollen* > *durchscrollen*,

*herunterscrollen, weiterscrollen, zappen > durchzappen, herumzappen, wegzappen, weiterzappen.*

The degree of morphological integration<sup>2</sup> may vary between low, moderate, and high. High integration can be observed with *SMS*, which has the two meanings “Kurzmitteilung” and “Dienst.” With the meaning of “Kurzmitteilung” *SMS* has German plurals (*SMS*, *SMSe*, *SMSen*), produces derivations (*simsen*), and is productive in forming nominal compounds as a head word, for example, *Gratis-SMS*, *Handy-SMS*, *Werbe-SMS*, and as a first word, for instance, *SMS-Party*, *SMS-Sucht*, *SMS-Versand*. With the meaning “Dienst” *SMS* has no gender and the masculine gender in the IDS text corpora of the nineties is only documented for the long form. It produces derivations (*simsen*) and nominal compounds as a first word, for example, *SMS-Dienst*, *SMS-fähig*, *SMS-Meldung*, *SMS-Nachricht*. An example of a moderate degree of morphological integration is *Trainspotter*, an expression with genitive *-s* and German plural (*Trainspotter*), a rather low productivity in compounding as a first word, for example, *Trainspotterfachbuch*, *Trainspotterszene*, and no derivations. *Aquacycling* is an example of low degree of integration; it can take the German genitive *-s* (but also the genitive form *Aquacycling* is attested) and only seldom is involved in compounds (e.g., *Aquacyclingkurs*).

## Semantic Change

The *Neologismenwörterbuch* documents semantic change and distinguishes between neologisms with different meanings and neologisms with older meanings that are not explicitly presented in the entries of the dictionary (cf. *Neologismenwörterbuch 2006–: Benutzerhinweise/Lesart*).

New meanings have already been mentioned earlier. Among the anglicisms there is, for example, the lexeme *Pushback* (after English *pushback*) “häufig gewaltsames Zurückschieben bzw. Zurücksenden von Flüchtlingen gegen ihren Willen bei Grenzübertreten,” while earlier, the word denoted more generally the “act of pushing something back” (OED online: “push-back, adj. and n.”). Another

<sup>2</sup> Although in this article we do not deal with phonological aspects for reasons of space, it should be noted that it is unlikely that neologisms as new words with a certain usage frequency in general language retain their original pronunciation. For instance, a pronunciation comparison reveals minor vocal differences between British English *attachment* ([ə'tætʃmənt], OED online) and German *Attachment* ([ɛ'tɛʃmənt], *Neologismenwörterbuch*; [ə'tɛʃmənt], DWDS). However, other expressions and their pronunciation like *twerken* are far more integrated in the German phonological system.

example is the word *Framing*, according to the *Neologismenwörterbuch* originally a term for the embedding of frames in websites. The term became more widespread denoting “stark auswählende, in eine bestimmte Richtung deutende Aufbereitung von Informationen mit dem Zweck der Meinungsbeeinflussung.”

*Zerealien* is another case documenting the development of a new meaning. Moreover, the noun could even constitute a pseudo-anglicism. Formerly a German word borrowed from Latin that referred to the grain, it refers to the ingredients of a breakfast dish now. According to the dictionary, the new reference to breakfast has been promoted by advertising and could have been influenced by the English language (cf. the corresponding meaning of English *cereals*).

The *Neologismenwörterbuch* takes into account several procedures of semantic change: metaphorisation, metonymisation, appellativisation, generalisation, specialisation and euphemisation (cf. *Neologismenwörterbuch* 2006–: *Erweiterte Suche/Weitere Merkmale*). Currently, only metonymisation, metaphorisation, appellativisation, and generalisation are relevant for the anglicisms documented in the dictionary. Metonymisation is the process that changes the semantics of an expression by meaning transfer. The transfer takes place because of a factual relation between the source concept and the new concept, for example, temporal, local, instrumental, etc. (cf. Seiffert, 2017a). Entries in the dictionary that underwent metonymisation are *Bitcoin*, *Blu Ray*, *Bodypainting*, *Bodypiercing*, *Book-on-demand*, *Bowl*, *BRIC*, *BRICS*, *Cranberry*, *E-Book*, *E-Paper*, *Escaperoom*, *Hype*, *MMS*, *Onliner*, *Paintball*, *Piercing*, *Powerplate*, *RFID*, *Second Screen*, *Sexting*, *SMS*, *Spoiler*, *Start-up*, *Take-away*, *Work-out*, *Zentangle*. For instance, *Take-away* has two meanings “(frisch zubereitetes) Schnellgericht zum Mitnehmen” and “Straßenverkauf für (frisch zubereitete) Schnellgerichte, Snacks bzw. Imbisslokal, Restaurant mit Straßenverkauf,” and the relation between the two readings is metonymic. Metaphorisation also employs semantic transfer; however, it takes place because of a relation of similarity between the source concept and the new concept (cf. Seiffert, 2017b). There are two entries in the *Neologismenwörterbuch* as far as the process of metaphorisation is concerned: *Pad* and *flashen*. *Flashen* has the two meanings “blitzen, wie ein Blitz kommen bzw. wirken” and “emotional berühren, begeistern, beeindrucken” that clearly have a metaphoric relation. Generalisation is the process leading to altered semantics by omitting one or some meanings, while, at the same time, generalising one specific meaning (cf. Wischer, 2021). In the dictionary, this is the case, for example, for the anglicism *unplugged* whose meaning “unverstärkt” was generalised to “authentisch.” Other examples are *Factoryoutlet*, *outen*, *Outing*, *Outlet*, and *Update*. Appellativisation is the process of a proper name becoming a common noun. This process can happen when a brand name or a proper noun becomes so commonly used

that it starts to be used as a generic term for a particular type of product or service (cf. Nübling *et al.*, 2012, p. 49 ff.), this is the case with words like *Chuck*, *Croc*, *E-Reader*, *Flexibar*, *Flipflop*, *Heely*, *Palmtop*, *Post-it*, *Powerplate*, *Rollerblade*, *Showview*, *Slush*, *Smartboard*, *Spa*, *Speeddating*, *Spinning*, *Ugg-Boot*, *Waveboard*, *Whiteboard*, *Zentangle*. Semantic change results in a loss of exogenous characteristics. Therefore, employing such processes is an indication that the language community has integrated the borrowed expressions into their active lexicon and does not treat them as words with a privileged status.

## Style and Usage

### Style

The *Neologismenwörterbuch* lists the stylistic markers commonly used in general language dictionaries (*Neologismenwörterbuch 2006–: Benutzerhinweise/Stil*). Stylistic markedness encompasses the categories “colloquial,” “slang,” “crude,” “colloquially joking,” “colloquially mocking,” “joking,” “mocking,” and “ironic.”

There are no slang, crude, joking or mocking usages among the anglicisms in the dictionary. 10 entries are marked as being of colloquial style (e.g., *Blingbling*, *Burner*, *flashen*, *handeln*, *Losser*, *Mc-Job*, *Onliner*, *pimpen*, *Update*, *ups*), 1 word as colloquially joking (*Couchpotato*), 1 entry as colloquially mocking (*pampfern*) and 1 word as ironic (*Greenwashing*). Thus, 727 of the anglicism entries are not marked for style.

### Usage

Information on usage includes pragmatic or sociolinguistic aspects like the restriction of certain group languages or registers or the restriction of oral or written contexts (cf. *Neologismenwörterbuch 2006–: Benutzerhinweise/Gruppensprache*, *Benutzerhinweise/Kommunikationssituation*). There is no categorisation for students’ language in the dictionary, however, there are 5 entries attributed to youth language (e.g., *Chill-out*, *Chill-out-Room*, *chillen*, *dissen*, *yep*), 18 to slang (*Szenesprache*) (e.g., *DJing*, *Musthave*, *Rave*, *raven*, *Thirtysomething*, *Twentysomething*), 1 entry to administrative language (*PPP*, the abbreviation of *Public-Private Partnership*), 39 entries to technical jargon (e.g., *Budgethotel*, *Cashcow*, *Clickworker*, *Cloud*, *Couchsurfer*, *Coworking*, *Crowdfunding*, *Crowdworker*, *Darknet*, *Domain*, *Edutainment*, *Fixie*, *fracken*, *Get-together*, *Give-away*, *Infotainment*, *Mikrojob*, *Second Screen*) and 40 as originating from languages for special purposes (e.g., *Bracket*, *Cloudcomputing*, *Digital Native*, *Downsizing*, *ESM*, *Femizid*, *Flagshipstore*, *Gender-Mainstreaming*, *High Potential*, *hosten*, *Lean Management*,



*Malware, Management-Buy-out, Medical Wellness, Nutriscore, Overtourism*). Lastly, 6 entries are categorised as occurring mostly orally (e.g., *at, buzzern, nice, Update, ups, yep*) and 1 entry as being used in written language (*politically correct*). The word *Geoblocking* is classified both as technical jargon and as originating from a language for special purposes. Particularly the developments in the area of terms exhibit that anglicisms can diffuse from the original usage as terms of a language for special purposes to a more common usage in German not strongly restricted to specific areas.

### Frequency and Diffusion

Frequencies and diffusion in different areas of communication indicate the significance and acceptance of borrowings in different areas and point to the degree of integration into the German language. One example is the borrowing *Aftershowparty* (after English *after show* “after the show” and *party* “party”) is a formation analogous to several similar formations containing *Party* as a second component and the first component determining the party’s occasion, like *Afterconcertparty, Aftercontestparty, Afterfilmparty*. These nouns emerged from the nineties onwards; the new lexeme *Aftershowparty* joined this group with significant usage beyond sporadic occurrences only in the first decade of the 21st century. Its first occurrences were in 1995 with a relative frequency below 0.05 per 1 million words and the peak in relative frequency, for the time being, in 2014 at 0.9 per 1 million words. Between 1995 and 2004, the relative frequency slowly rose to the value of 0.5 per 1 million words and then between 2005 and 2014 reached its peak; after 2014 the relative word frequency decreased to about 0.65 between 2015 and 2019, before the frequency per 1 million words saw a drop below 0.3 in 2020. Both the rise in frequency and the number of parallel, analogous formations are indications that this group of words and the formation pattern found acceptance in the language community and diffused in general use. Other examples can be found among those words or expressions originally restricted to a language for special purposes, as already noted, or a group language that spread to more general use and hence saw a growth in frequency. Frequency is an important criterion for the degree of acceptance and diffusion in general language. Its development should be observed over a longer period and noticeable changes in frequency should be related to extra-linguistic developments. Temporal distribution of usage frequency is given for each entry in the dictionary.

The degree of acceptance and the diffusion in general language use can be low, moderate, or high as well. It can be determined by examining the usage

restrictions to a group language like technical jargon (e.g., *Fixie*, “minimalistisch ausgestattetes Fahrrad mit starrer Nabe, ohne Gangschaltung und ohne Bremsen, das nicht für den Straßenverkehr zugelassen ist”), street slang (e.g., *Rave*, “große, zeitlich ausgedehnte Tanzparty in einer Location oder im Freien, besonders Technoveranstaltung”) or youth language (e.g., *funzen*, “funktionieren”).<sup>3</sup> For instance, the term *Bag-in-Box*, a direct loan from English *bag-in-box* was originally used only in the language of the catering trade. However, from the middle of the first decade of the 21st century, it occurred in general language use and became more widespread. According to the dictionary, this went hand in hand with the fact that the product bag-in-box gradually began to be offered in retail and, therefore, became more common.

## Conclusion

Since 1990, the *Neologismenwörterbuch* in the OWID has been an expanding dictionary documenting new words, phrases, and meanings for established concepts in general German. English has been a relevant source for neologisms in recent decades. As a result, there are several anglicisms in German everyday language. Based on the dictionary and using criteria from different areas of the language system and of language use, it has been possible to analyse Anglicisms in German since the nineties and obtain a structured overview of them.

Apart from formal criteria concerning *partes orationis*, morphological processes and integration, types of borrowing, and semantic change, the analysis includes observations about the frequency and diffusion of neologisms in the general language community. The degree of acceptance of neologisms can be assessed by considering possible sociolinguistic and pragmatic aspects such as stylistic markedness and usage restrictions. Particularly the examples in the previous section show that formal integration, frequency, and diffusion in the language community are interdependent. With constant data observation, changes can be determined, for instance, the development from a rather restricted use in languages for special purposes to a more general use or from stylistically marked as colloquial to stylistically unmarked. Both developments can lead to the acceptance of new words and expressions, and they can be accompanied by semantic change.

---

<sup>3</sup> Restrictions to certain topic areas might indicate that expressions are not widespread. Nevertheless, this criterion should not be regarded alone.

## References

- Baeskow, H., & Müller, P. O. (2014). Konversion. In S. J. Schierholz *et al.* (Eds.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. De Gruyter. [www.degruyter.com/database/WSK/entry/wsk\\_iddae11706-5594-45fb-a081-a3f765eeb480/html](http://www.degruyter.com/database/WSK/entry/wsk_iddae11706-5594-45fb-a081-a3f765eeb480/html) (accessed 1st August 2023).
- Burmasova, S. (2010). *Empirische Untersuchung der Anglizismen im Deutschen am Material der Zeitung Die WELT (Jahrgänge 1994 und 2004)*. University of Bamberg Press.
- Busse, U. (2001). Typen von Anglizismen: von *der heilago geist* bis *Extremsparing* – aufgezeigt anhand ausgewählter lexikographischer Kategorisierungen. In G. Stickel (Ed.), *Neues und Fremdes im deutschen Wortschatz. Aktueller lexikalischer Wandel* (pp. 131–155). De Gruyter.
- Elsen, H. (2011). *Grundzüge der Morphologie des Deutschen*. De Gruyter.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31–80.
- Herberg, D. (1980). Wortbegriff und Orthografie. In D. Nerius & J. Scharnhorst (Eds.), *Theoretische Probleme der deutschen Orthographie* (pp. 140–161). Akademie-Verlag.
- Kessel, K., & Reimann, S. (2010). *Basiswissen Deutsche Gegenwartssprache*. 3., überarbeitete Auflage. Francke.
- Lemnitzer, L., & Zinsmeister, H. (2015). *Korpuslinguistik: Eine Einführung*. 3. Auflage. Narr.
- Müller-Spitzer, C. (2008). Der texttechnologische Aufbau von OWID. In A. Klosa (Ed.), *Lexikografische Portale im Internet* (pp. 45–56). Institut für Deutsche Sprache.
- Nübling, D., Fahlbusch, F., & Heuser, R. (2012). *Namen. Eine Einführung in die Onomastik*. Narr Francke Attempto.
- Seiffert, A. (2017a). Metonymisierung. In S. J. Schierholz *et al.* (Eds.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. De Gruyter. [www.degruyter.com/database/WSK/entry/wsk\\_id\\_wsk\\_artikel\\_artikel\\_17597/html](http://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_17597/html) (accessed 1st August 2023).
- Seiffert, A. (2017b). Metaphorisierung. In S. J. Schierholz *et al.* (Eds.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. De Gruyter. [www.degruyter.com/database/WSK/entry/wsk\\_id\\_wsk\\_artikel\\_artikel\\_17114/html](http://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_17114/html) (accessed 1st August 2023).
- Steffens, D., & al-Wadi, D. (2015). *Neuer Wortschatz. Neologismen im Deutschen, 2001–2010*. 2 Bände, 3. Auflage. Institut für Deutsche Sprache.
- Steffens, D. (2008). Zu den Verlinkungen in den Wörterbüchern des Internetportals OWID – Ein Erfahrungsbericht aus Sicht des Neologismenwörterbuches (NWB).

- In A. Klosa (Ed.), *Lexikografische Portale im Internet* (pp. 97–106). Institut für Deutsche Sprache.
- Steffens, D. (2017). Von Pseudoanglizismen und Kurzzeitwörtern. Zwei Aspekte der Beschreibung von neuem Wortschatz im Neologismenwörterbuch im IDS-Portal OWID. *Sprachwissenschaft*, 42(3), 275–304.
- al-Wadi, D. (2013). Zwergentechnologie, Alphamädchen, Zurückleaken. Verdeckte neue Wörter des Neologismenwörterbuchs besser zugänglich machen. *Sprachreport*, 29(3), 16–24.
- al-Wadi, D. (2017). Begegnungen mit neuen Wörtern: Zu lexikografischen Praktiken im Neologismenwörterbuch des IDS. In J. Dabrowska-Burkhardt, L. M. Eichinger, & U. Itakura (Eds.), *Deutsch: lokal – regional – global* (pp. 173–186). Narr Franke Attempto.
- Wischer, I. (2021). Bedeutungswandel. In S. J. Schierholz (Ed.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)*. De Gruyter. [www.degruyter.com/database/WSK/entry/wsk\\_id9f9dda81-6b2b-4aa3-972b-4979e429a81d/html](http://www.degruyter.com/database/WSK/entry/wsk_id9f9dda81-6b2b-4aa3-972b-4979e429a81d/html) (accessed 1st August 2023).


## Dictionaries

- DWDS – Digitales Wörterbuch der deutschen Sprache. <https://www.dwds.de/> (accessed 1st August 2023).
- Neologismenwörterbuch (2006–). In *OWID – Online Wortschatz-Informationssystem Deutsch*. (Ed.), Leibniz-Institut für Deutsche Sprache (IDS). [www.owid.de/docs/neo/start.jsp](http://www.owid.de/docs/neo/start.jsp) (accessed 1st August 2023).
- Oxford English Dictionary online* (OED online), [www.oed.com](http://www.oed.com) (accessed 1st August 2023).



*Françoise Collinet*

Université Jagellonne de Cracovie  
Pologne

 <https://orcid.org/0000-0001-7520-2000>

# Pour une grammaire des notions et de leurs transformations rhétoriques

**For a grammar of concepts and their rhetorical transformations**

## **Abstract**

In a 1955 article, Perelman and Olbrechts-Tyteca suggest that their theory of argumentation could provide a foundation for a study that they place under the sign of semantics: the study of notions. In the *Treatise on Argumentation*, the multiplicity of the examples cited can monopolise the reader's attention and prevent him from finding the lineaments of this semantic reflection. If a semantic project appears in the *Treatise*, it is in the manner of a watermark. The analysis that Ducrot proposes of the word *But* and his commentary on Pascal's *Refutation of the Dominicans* could offer a precious means of making the functioning of this study of notions promised by the New Rhetoric more visible. Ducrot's strictly linguistic approach will allow for a better understanding of the workings of the Perelmanian method, a method that, by contrast, we will call "rhetorico-grammatical".

## **Keywords**

Grammar, Rhetoric, Argumentative semantics, Perelman, Ducrot

## **1. Argumentation linguistique et argumentation rhétorique**

La géométrie élémentaire affirme que deux droites pourront être dites parallèles si et seulement si elles ne se rencontrent pas. Telle semble être la relation qui se tisse entre la trajectoire de la théorie de l'argumentation dans la langue et celle de la nouvelle rhétorique perelmanienne, alors même que les deux démarches trouvent leur racine dans une commune prise de distance face au

positivisme logique (Plantin, 1990 : 43). Une petite quinzaine d'années plus tard, Ducrot (2004) met en garde contre la confusion entre argumentation linguistique et argumentation rhétorique. Le désir de « comprendre le mécanisme de la pensée vivante » qui anime le *Traité de l'argumentation* (TA, § I, 7) se trouve certainement à contre-pied de l'approche résolument linguistique des tenants de *L'argumentation dans la langue*.

Pourtant, dans *Logique, structure et énonciation*, Ducrot (1989 : 7) affirme également que « Ce qui est passionnant, c'est de voir comment les théories [...] modifient la façon de vivre et de percevoir le langage ». C'est sous cet angle que nous aimerions placer la présente contribution. Le point de départ de la réflexion est d'examiner, dans la perspective de la nouvelle rhétorique, deux analyses proposées par les théoriciens de l'argumentation dans la langue. Le premier exemple concerne l'énoncé *Ce restaurant est bon mais cher* ; le second porte sur un texte à prétention rationnelle, un texte plus proche des argumentations habituellement étudiées dans le cadre de la Nouvelle Rhétorique. Il s'agit de la réfutation des dominicains par Pascal et de l'analyse qu'en propose Ducrot (2007 [1980], 77–80) dans *Les échelles argumentatives*.

Dans les deux cas, l'objectif sera de faire apparaître l'habile compromis sur lequel repose la structure du *Traité de l'argumentation* qui peut se lire non seulement comme une adaptation d'une rhétorique millénaire aux préoccupations de l'Honnête Homme du XX<sup>e</sup> siècle, mais aussi comme l'esquisse d'une sémantique argumentative.

Cette seconde possibilité est exposée dans un article de 1955 qui suggère que l'étude de l'argumentation est à considérer comme un fondement de l'étude des notions (2012 : 131) ; autrement dit, la théorie perelmanienne de l'argumentation rhétorique pourrait être mise au service d'une étude sémantique de la transformation que l'argumentateur fait subir aux notions dont il se sert.

## 2. Nouvelle rhétorique et sémantique argumentative

Le projet d'une étude sémantique des notions et de leurs transformations sous l'effet de l'argumentation informe la structure du *Traité* qui se révèle dès lors plus ambiguë qu'il n'y paraît : deux structures différentes semblent se superposer.

La lecture rhétorique incite à se concentrer sur les parties 1 et 3 du *Traité* : la première expose les enjeux et les choix théoriques (remise en question du réductionnisme logique qui va de Descartes au logicisme contemporain, réhabilitation

de l'auditoire) tandis que la troisième fait un inventaire d'une cinquantaine de techniques argumentatives qui recourent, en grande partie, les listes établies par la rhétorique traditionnelle ; à cette liste sont ajoutées des dissociations argumentatives qui ont suscité beaucoup de commentaires.

La deuxième partie, quant à elle, contient des remarques d'allure banalement grammaticale (Amossy, 2002 : 153) et semble dès lors moins importante. Or, pour activer la lecture sémantique proposée, il faudrait, au contraire, se concentrer sur cette « grammaire des notions » et sur la manière dont elle annonce et prépare l'étude des techniques argumentatives. Dans cette perspective, la cinquantaine de schémas argumentatifs exposés par le *Traité* se trouve subordonnée à deux types d'opérations bien plus générales : des associations ou des dissociations de notions. Ces dernières sont susceptibles de conduire, par composition ou décomposition, à de nouvelles configurations notionnelles. L'articulation des parties 2 et 3 peut alors s'envisager comme deux plans (l'un grammatical et l'autre rhétorique) où, par des moyens différents, l'argumentateur adapte les notions à ses besoins.

**Tableau 1**

*Deux lectures de la structure du Traité*

Lecture rhétorique	Lecture sémantique
Partie 1 = Cadres de l'argumentation Partie 2 = Le point de départ de l'argumentation Partie 3 = Liste de schémas argumentatifs Partie 4 = Ordre et méthode	Partie 1 = Cadres de l'argumentation Partie 2 = Le point de départ de l'argumentation [grammaire des notions] Partie 3 = Associations et dissociations de notions [rhétorique des notions] Partie 4 = Ordre et méthode

Couplée à l'obligation de choix face à laquelle se trouve souvent l'être humain, la possibilité d'activer une telle grille de lecture est d'ailleurs explicitement signalée par les auteurs :

L'ordre adopté dans notre étude nous a amenés à considérer en dernier lieu l'usage et la transformation des notions, c'est-à-dire l'aspect sous lequel le problème du choix nous oblige à repenser, dans une perspective rhétorique, la plupart des problèmes sémantiques.

(TA, § 35 : 189)

Pour mettre en évidence cette grille de lecture sémantique, nous proposons de revenir sur certains éléments constitutifs de cette « grammaire des notions »

que semble constituer la partie II du *Traité* (§ 3). Ensuite, le § 4 montrera comment s'articule la transition entre cette étude des notions au niveau grammatical et l'étude de la transformation des notions (plan rhétorique ; partie III du *Traité*).

### 3. Une grammaire des notions comme moyen d'étude des accords préalables à l'argumentation proprement dite

De la partie consacrée à l'étude des accords préalables fondés sur le langage, on retiendra à ce stade de la réflexion cinq éléments :

- a) C'est dans cette deuxième partie du *Traité* qu'apparaît l'idée de notions confuses, plastiques et malléables. Le concept est hérité de Dupréel. À notre avis, il apparaît aussi comme une contestation mais aussi, dès lors, comme un souvenir des idées claires et distinctes de la *Logique* de Port Royal. La démarche de Dupréel consiste, dans notre esprit, à réhabiliter ces idées « obscures » ou « confuses » que *L'Art de penser* (§ I.9 ; 1970 : 101–109) et toute la tradition qui en est issue semblent avoir dressées en repoussoirs. Même si les Messieurs de Port Royal ne sont guère cités dans le *Traité*, cette interprétation semble confortée par le fait que Dupréel (1911) ait, à l'occasion, parlé en des termes neutres ou connotés positivement d' « idées confuses » au point de leur consacrer une théorie. La réorganisation perelmanienne des oppositions *clair / obscur* et *distinct / confus* en un continuum *obscurcissement / clarification* des notions<sup>1</sup> et, surtout, l'idée que ces deux opérations sont le résultat d'un travail effectué par l'argumentateur (TA, § 34 : 178–185 ; cf. aussi Dominicy, 2002 : 8) vont également dans le sens d'une relecture partiellement infidèle des jansénistes.
- b) Les notions se définissent en extension et en compréhension (TA, § 33, 174 ; § 35 : 187). À nouveau, même si la tradition grammaticale a procédé à quelques changements d'étiquettes, cette terminologie rappelle étrangement les conceptions d'Arnauld et Nicole (§ I.9 ; 1970 : 87–88). Cette opposition de l'extension et de la compréhension est d'autant plus importante qu'elle est au cœur des variables qui permettent à l'argumentateur d'agir sur une notion (Perelman, 2012 : 121) ; cette opposition de l'extension et de la compréhension recoupe, à notre avis, les opérations de clarification / obscurcissement ou d'assouplissement / durcissement dont les auteurs du *Traité* parlent par

---

<sup>1</sup> La typologie des notions proposée par Perelman et Olbrechts-Tyteca est fondée sur un tel continuum (Perelman, 2012 : 113 ; cf. aussi Dominicy, 2005 : 6–7).



ailleurs. Une chose qui peut étonner est que Perelman semble penser qu'on peut, au choix<sup>2</sup>, modifier l'extension ou la compréhension d'une notion alors que la grammaire traditionnelle laisse espérer que les deux valeurs s'entrelaceraient de manière stable en un rapport inversement proportionnel, conformément à ce que l'on nomme parfois la loi de Port Royal : plus on précise la compréhension d'un mot, plus son extension s'en trouve restreinte. Inversement, plus on étend l'extension d'un mot, moins sa compréhension est précise. La précision de la compréhension et la restriction de l'extension qui la compense se produisent lorsqu'on « joint » deux idées ensemble ; par exemple, « triangle » et « qui possède un angle droit » (Arnauld & Nicole, 1970 : § I.9 : 88).

- c) Toujours en suivant Dupréel, Perelman et Olbrechts-Tyteca introduisent une hiérarchie entre notions-maîtresses et notions-béquilles (TA, § 56 : 324). Cette distinction semble capitale et nous ne saurions trop insister sur son importance pour notre propos. Si la modification des notions résulte uniquement de leurs associations et de leurs dissociations, il faut nécessairement faire la distinction entre les notions qui sont à considérer prioritairement comme des acteurs de celles qui apparaissent davantage comme modifiées par l'action des autres. Sous la plume de Dupréel (1939 : 67–71), notions-maîtresses et notions-béquilles (aussi nommées notions-étançons<sup>3</sup>) sont décrites comme des complémentaires dont le rapport est fondé en raison inverse sur l'axe de la clarté et de la confusion. Il faut cependant aussi admettre que l'identification d'une notion-maîtresse laisse une certaine part de liberté à l'utilisateur de la théorie de l'argumentation avec tous les risques d'imprécision que cela suppose : la notion-maîtresse dans un texte comme *L'apologie de Socrate* est-elle le Juste, la Mort, la figure du Philosophe idéal, l'impiété ou autre chose encore ?

- d) Avant-plan et arrière-plan

Un autre élément essentiel à une interprétation sémantique du *Traité* est la nature de cette relation très solide et potentiellement productive entre la partie II de l'ouvrage (Le point de départ de l'argumentation) et la partie III du *Traité* (Les techniques argumentatives). Mais la question se pose de savoir en quoi consiste la relation, complexe, entre les deux sections. Pour nous, le *Traité* est organisé selon un découpage qui peut se décrire de diverses manières :

---

<sup>2</sup> Cet apparent paradoxe se résout si l'on prend en compte l'opposition entre avant-plan et arrière-plan que nous abordons au point 3 d) : l'argumentateur agit, de manière plus explicite, à l'avant-plan, tandis que l'auditoire, dont l'attention est fixée sur les éléments d'avant-plan, risque de perdre de vue les conséquences qui se déroulent à l'arrière-plan.

<sup>3</sup> Pièce de bois ou de métal permettant d'étayer un mur ou une charpente.

**Tableau 2***Articulation des parties II et III du Traité*

Partie II	Arrière-plan	Plan grammatical	Choix et retravail des notions
Partie III	Avant-plan	Plan rhétorique	Associations ou dissociations de notions

L'idée d'une opposition entre un avant-plan et un arrière-plan apparaît en plusieurs endroits sous la plume des auteurs (par ex. TA, § 36 : 191 ; § 38 : 203 ; § 44 : 256). Dans la suite du propos, les termes « avant-plan » et « arrière-plan » seront utilisés pour décrire cet ancrage de l'étude rhétorique (et sémantique) des associations ou dissociations de notions sur une grammaire des notions.

- e) Un mot encore sur l'importance de l'opération de classement et la manière dont elle structure les relations entre l'avant-plan et l'arrière-plan dont nous venons de parler (Perelman, 2012 : 116–117). La division entre accords préalables et argumentation proprement dite est théorique : le choix d'une qualification, le choix des données sont d'ores et déjà argumentatifs. Cependant, cette remarque est encore bien insuffisante. Pour revaloriser comme il se doit le rôle primordial de l'arrière-plan, mieux vaudrait dire que les accords préalables constituent non seulement le point de départ mais aussi le point d'arrivée de l'argumentation puisque « d'un bout à l'autre l'analyse de l'argumentation concerne ce qui est censé admis par les auditeurs » (TA, § 15 : 87). Contrairement à ce que suggère l'expression « accords préalables », la relation ne va pas uniquement de l'arrière-plan à l'avant-plan, mais aussi de l'avant-plan à l'arrière-plan. Une nouvelle de Maupassant, parce qu'elle se déroule dans un prétoire, permettra de dire en deux mots ce qu'un philosophe du droit pourrait bien avoir en tête :

La fille Prudent (Rosalie), bonne chez les époux Varambot, de Mantes, devenue grosse à l'insu de ses maîtres, avait accouché, pendant la nuit, dans sa mansarde, puis tué et enterré son enfant dans le jardin<sup>4</sup>.

Face à des preuves accablantes, la Justice française s'apprête à classer l'affaire Prudent comme un infanticide. Pourtant les jurés, le Président et le procureur lui-même hésitent : la jeune femme s'est renseignée auprès de la sage-femme sur les gestes à accomplir au cas où elle devrait enfanter seule et, surtout, elle a patiemment confectionné un trousseau destiné au nouveau-né. De l'interrogatoire, il ressort que la bonne a donné la vie à des jumeaux et que cette décou-

<sup>4</sup> Guy de Maupassant, Rosalie Prudent, *Nouvelles choisies*, Moscou, 1965(1886), 91.

verte l'a plongée dans une panique indescriptible : seule comme elle l'est, elle pourrait assurer, avec difficulté sans doute, la vie d'un enfant, mais jamais de deux. Dans son effroi, ne sachant lequel sauver et lequel sacrifier, elle a étouffé les deux. La nouvelle se conclut sur ces mots : « la fille Rosalie Prudent a été acquittée ».

Parce qu'elle commence par décrire les représentations préalables des juges (arrière-plan), la nouvelle montre bien comment le récit de l'accusée (avant-plan) conduit à une modification des représentations initiales (arrière-plan modifié). Ce qui constitue le point d'arrivée du processus argumentatif, c'est que les actes de l'accusée ne sont plus à classer comme un (double) infanticide mais qu'il faut les insérer dans une classe du type « (double) homicide commis sous l'empire d'un trouble émotionnel ». Un raisonnement analogue pourrait prendre comme notion-maîtresse non plus l'acte commis par l'accusée mais le caractère de cette dernière : Rosalie Prudent n'est pas une matricide mais une jeune femme en détresse qui a agi sous le coup de la panique. Cette liberté dans la détermination de la notion-maîtresse à choisir peut être perçue comme un manque de précision de la théorie perelmanienne ; il nous semble bien plutôt qu'il s'agisse d'une ambiguïté recherchée et calculée de manière à assurer la souplesse du système. L'enjeu est de lui permettre de s'adapter aux objets à étudier en le dotant de principes régulateurs. L'ambiguïté peut en effet être régulée si l'on admet que le lien entre la personne et les actes (TA, § 68 : 394) est une technique argumentative qui permet l'interaction entre des notions mutuellement déformables ; cette technique argumentative fait en effet partie de la famille des associations de notions fondées sur la structure du réel. Dans cette logique, l'argumentation sera, en fin de compte, un processus qui mène des accords préalables présents à l'arrière-plan (terme *a quo*) à une (nouvelle) représentation de l'arrière-plan indépendamment du fait que les accords se trouvent modifiés ou simplement renforcés (terme *ad quem*). Le fait que ce terme *ad quem* puisse ne correspondre qu'à la représentation qu'en propose l'argumentateur et non à celle des auditeurs qui restent libres de ne pas y adhérer n'est pas non plus gênant en soi : ce qu'il s'agit d'étudier à ce stade, c'est la modification que l'argumentateur cherche à imprimer aux notions ; l'efficacité de cette démarche est un autre problème à traiter comme une question à part.

#### 4. Une rhétorique des notions comme moyen d'étude des transformations sémantiques subies par les notions sous l'empire de l'argumentation

Une fois admise la possibilité d'étudier les déformations de notions à partir de certaines opérations grammaticales (modification de l'extension et de la compréhension, qualification, classement), l'examen des transformations que l'argumentateur peut faire subir aux notions peut se prolonger au niveau rhétorique.

La troisième partie du *Traité* peut se lire comme un inventaire reprenant une cinquantaine de techniques rhétoriques consacrées par la tradition (argument d'autorité, pétition de principe, argument de l'homme de paille, etc.); l'inventaire serait ensuite complété par un type de techniques jusqu'alors négligé : les dissociations de notions. L'attention portée aux remarques grammaticales sur des notions malléables selon que l'argumentateur agit sur leur extension ou leur compréhension peut cependant tout aussi bien conduire à réduire l'ensemble des techniques argumentatives à deux grands types d'opérations : les associations de notions et les dissociations de notions. Nous proposons d'associer chacun des types d'associations / dissociations possibles à un symbole :

##### A. Les associations de notions

**Tableau 3**

*Trois types d'associations de notions*

●	quasi-logiques	Définition, contradiction, tautologie, etc.
■	fondées sur la structure du réel	Cause, conséquence, autorité, etc.
◆	qui fondent la structure du réel	Exemple, illustration, analogie, métaphore, etc.

À ces trois types d'associations s'opposent alors

##### B. Les dissociations de notions

**Tableau 4**

*Une quatrième possibilité : les dissociations de notions*

Φ	dissociations	Opinion / vérité, Apparence / réalité, Lettre / esprit, etc.
---	---------------	--

L'articulation entre l'étude des notions au plan grammatical et l'étude des associations et dissociations de notions au plan rhétorique est exposée avec beaucoup de soin au § 15 et, surtout, au § 44. À dire vrai, toute la présente réflexion naît de cet énigmatique passage du § 44 :

Psychologiquement et logiquement *toute* liaison [de notions] implique une dissociation et inversement : *la même forme* qui unit des éléments divers en un tout bien structuré les dissocie du *fond neutre dont elle les détache*. Les deux techniques sont complémentaires et *toujours* à l'œuvre en même temps ; mais l'argumentation peut *mettre l'accent* sur la liaison ou la dissociation qu'elle est en train de favoriser *sans expliciter* l'aspect complémentaire qui résultera de la transformation recherchée.

(TA, § 44 : 256 ; nous soulignons)

- De ce passage, nous retiendrons trois éléments pour guider la suite du propos :
- la relation entre *association* / *dissociation* de notions semble marquée par une complémentarité jamais prise en défaut ;
  - l'idée d'une association (ou dissociation) sur laquelle l'argumentateur met l'accent alors que l'opération complémentaire se déroule, de manière plus implicite, à l'arrière-plan ;
  - l'idée d'une *forme* déterminée se détachant d'un fond indéterminé nous apparaît comme la description d'une perception visuelle.

À dire vrai, cette description semble à la fois suffisamment claire et suffisamment intrigante, pour susciter le désir d'observer de ses propres yeux ces chassés-croisés de notions. Voyons si l'Argumentation dans la langue peut venir à notre secours.

## 5. De la transformation des notions dans un énoncé du type *p mais q*

Que donnerait l'analyse de l'énoncé *Ce restaurant est bon mais cher* si on tentait de l'étudier sous l'angle de la transformation des notions ?

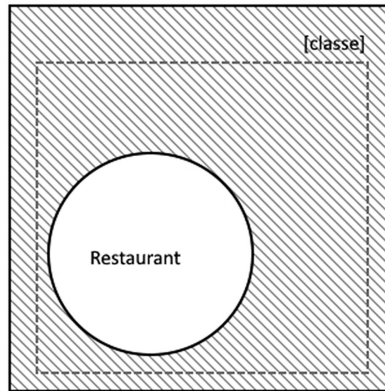
L'instruction de lecture fournie par la Nouvelle Rhétorique ne saurait être d'identifier (ou de construire), à partir des propositions *p* et *q* qui sont données, une proposition – *r* qui révélerait, dans un contexte donné, le sens de l'énoncé.

La première tâche que la théorie néo-rhétorique demande à son usager d'accomplir est de distinguer notion-maîtresse et notions-outils. Avec la part de

liberté que nous laisse la théorie sur ce point, considérons que *Restaurant* est la notion-maîtresse et que tous les autres mots de la phrase correspondent à des notions-outils. Installons donc la notion *Restaurant* sur le fond neutre de tous les sujets possibles mais que l'argumentateur éloigne du champ de l'attention de son auditoire.

**Figure 1**

*L'attention de l'auditoire est attirée vers une notion qui se détache du fond neutre de la multiplicité des sujets de discussion possibles*

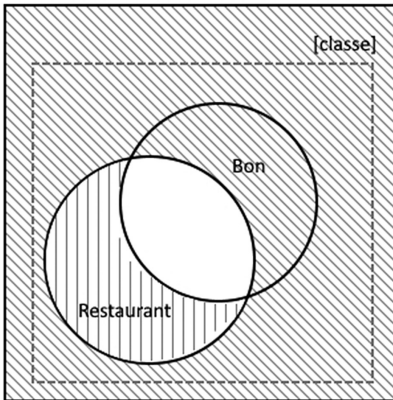


Ouvrons cependant une place pour la « classe » dans laquelle cette notion-maîtresse pourrait avoir à être inscrite. Cette décision qui peut sembler paradoxale découle de l'hypothèse émise au point e) de notre § 3 au sujet de l'importance des opérations de classification. Remarquons aussi que pour les notions-maîtresses auxquelles songent prioritairement les auteurs du *Traité*, les argumentateurs commenceront volontiers par proposer un premier classement de la notion discutée ; par exemple, le *Bonheur* est un sentiment, le *Bonheur* est un concept philosophique, le *Bonheur* est le but de toute éducation, etc.

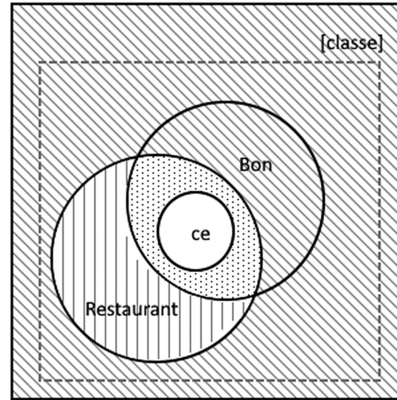
La proposition *Ce restaurant est bon* est l'occasion de montrer comme les préceptes de la grammaire la plus élémentaire et la plus traditionnelle correspondent bien aux notions qui s'associent et se dissocient en se détachant ou en se confondant avec un arrière-plan :

**Figure 2**

*L'association des notions Restaurant et Bon entraîne une restriction de l'extension assimilable à un mouvement dissociatif*

**Figure 3**

*Le champ de l'attention est à nouveau rétréci et orienté vers un restaurant particulier connu de l'argumentateur et de son auditoire*

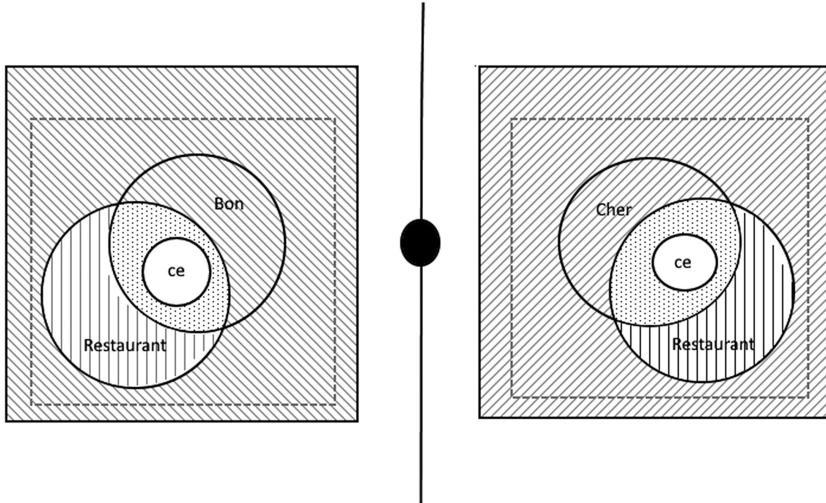


Il est important de remarquer qu'alors que cette figure correspond toujours au seul plan grammatical, le mouvement observé est, d'ores et déjà, une association de notions qui, implicitement, entraîne une rupture avec d'autres notions potentiellement disponibles. Par exemple, dire *Ce restaurant est bon* éloigne du champ de notre conscience d'autres configurations possibles : le restaurant en question pourrait « être à deux pas », être un établissement « qui vient d'ouvrir » ou « qui est recommandé dans le guide touristique ». Autrement dit encore, l'association des notions *Restaurant* et *Bon* qui conduit à préciser la compréhension de la notion-maîtresse est immédiatement compensée par une réduction de l'extension de cette notion.

La question qui se pose est de savoir si cette grille d'analyse pourra dépasser le plan grammatical. L'apparition du connecteur *mais* fournit une seconde instruction. Cependant cette fois encore, il ne s'agit pas d'identifier une proposition – *r*. Pour la NR, l'instruction consiste à placer les deux configurations notionnelles (*Restaurant x Bon* et *Restaurant x Cher*) en compétition. Les deux configurations que, accessoirement, nous disposons en opposition symétrique se retrouvent aussi, et surtout, unies par une association de type quasi-logique (●) : l'incompatibilité. Il est très important de remarquer que cette incompatibilité est une association de notions et même celle qui occupe la première place dans l'inventaire néo-rhétorique (TA, § 46 : 262–264).

**Figure 4**

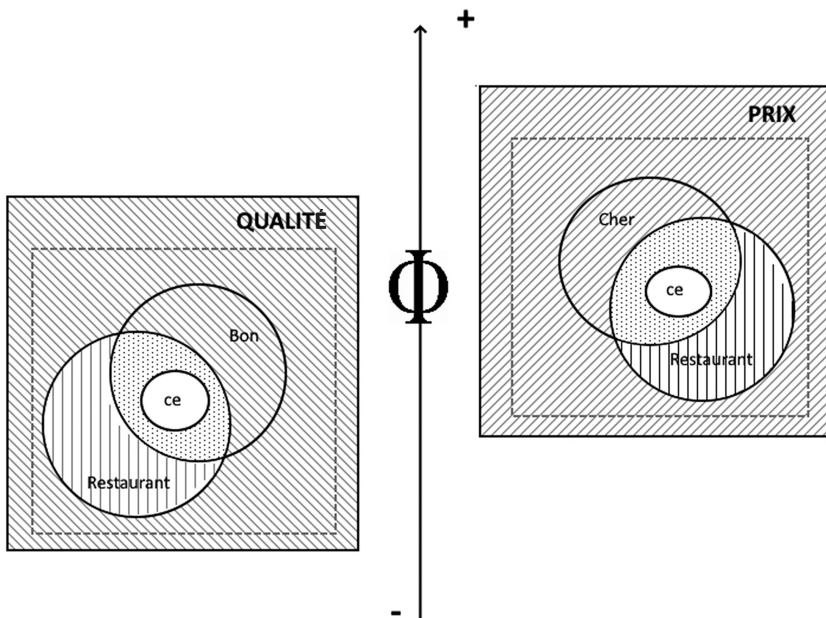
*L'incompatibilité en tant qu'association de notions de type quasi-logique*



L'incompatibilité a cependant cette particularité de favoriser, de manière très régulière, l'apparition d'une dissociation de notions (TA, § 89 : 552-553 ou 555).

**Figure 5**

*La dissociation de notions proprement dite*





Pour opérer cette dissociation, il a néanmoins fallu disposer de critères. Dans le cas étudié, l'argumentateur se présente comme ayant dû choisir entre la « qualité » du restaurant et son « prix » ; autrement dit, la notion *Restaurant* a été classée comme un objet de bonne « qualité » mais aussi comme un objet dont le « prix » est élevé. Un point important est que le critère de classement choisi correspond au critère de décision. Pour résoudre l'incompatibilité, l'argumentateur a aussi recours à un lieu du préférable qu'il présente comme allant de soi et qu'il adapte à ses besoins : dans ce cas particulier, il faut préférer l'équilibre de notre porte-monnaie à la promesse d'un bon repas. Les lieux font, en principe, partie des accords préalables à l'argumentation proprement-dite et on voit, une nouvelle fois, comment ce préalable fait son entrée en scène alors que l'argumentation a bel et bien commencé voire alors qu'elle est sur le point de se clore.

Que retenir de ce point de contact que nous avons favorisé entre ADL et NR ?

Sans chercher à généraliser, constatons que si chez Ducrot le connecteur *mais* unit des propositions, chez Perelman le *mais* unit des configurations de notions qui semblent s'insérer dans des classes mises en concurrence. Le connecteur *mais* n'oppose certes pas directement les prédicats « être bon » et « être cher » (Plantin, 2016 : 377), mais il oppose la *qualité* et le *prix* en tant que, après interaction entre les notions, ces deux concepts offrent un critère de choix parmi des valeurs concurrentes. Dans le cas présent, ces opérations de classement semblent coïncider avec le choix que l'argumentateur a, en réalité, déjà effectué et qu'il présente comme évident. Les diverses associations de notions effectuées au plan grammatical sont compensées par autant de restrictions de l'extension ; au plan rhétorique, l'association de notions est compensée par une dissociation de notions qui éloigne du champ de l'attention la bonne qualité du restaurant. Il est également possible que la méthode perelmanienne donne une information sur la manière dont l'utilisateur de l'ADL construit les propositions – *r* et *r* : si *q* (*ce restaurant est cher*) « est un argument pour » – *r* (*n'y allons pas*), c'est que l'utilisateur de l'ADL sait qu'un restaurant est un endroit dont on parle notamment au moment où l'on songe à s'y rendre et qu'il a aussi reconnu que l'argumentateur recourait à un lieu du préférable où la qualité doit céder le pas au prix.

Si, dans l'ADL, la manipulation des notions que suppose l'identification des propositions – *r* et *r* semble jouer le rôle de variable d'ajustement ; dans la NR, c'est au contraire, la nature des expressions linguistiques signalant la présence d'une dissociation à opérer qui est libéralement laissée à l'appréciation de l'utilisateur : certaines expressions permettraient de reconnaître, au premier coup d'œil, le terme I et le terme II d'une dissociation argumentative à opérer (TA, § 93, 580–581). On peut raisonnablement penser que, du point de vue néo-rhétorique, l'ensemble des connecteurs de concession-opposition sont à considérer comme le signal

d'une dissociation à opérer ; toute la question étant de savoir si cette dissociation concerne, ou non, des notions suffisamment proches de la notion-maîtresse. Dans le même ordre d'idées, la NR ne s'intéresse aux nuances que les linguistes cherchent à dégager entre *mais*, *cependant* ou *pourtant* que dans la mesure où elles seraient un moyen d'interpréter la déformation des notions de rang supérieur.

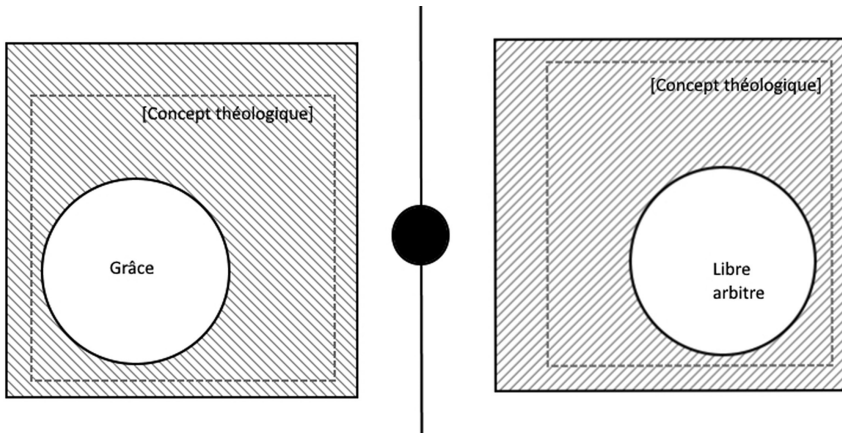
Les phénomènes que la nouvelle rhétorique cherche à observer sont aussi bien plus étendus. Comme la dissociation de notions est aussi fréquente que l'association de notions, ce motif ne se limite pas aux connecteurs de concession ou d'opposition : des oppositions stabilisées en couples « philosophiques » telles que *la lettre / l'esprit* ou *nature / culture* permettront d'atteindre le même résultat. C'est la raison pour laquelle la réfutation des dominicains par Pascal que Ducrot étudie dans *Les échelles argumentatives* (2007 : 77–80) permettra de prolonger la réflexion même si elle n'est pas liée à la présence d'un connecteur tel que *mais*.

## 6. De la Réfutation des dominicains par Pascal

Au XVII<sup>e</sup> siècle, les jésuites discutent de la théorie de la grâce héritée des Pères de l'Église. En termes perelmaniens, on pourra dire que les jésuites en distinguant *Grâce suffisante* et *Grâce efficace* introduisent une dissociation de notions. Dans l'esprit des jésuites, cette dissociation résulte de la perception d'une incompatibilité entre les notions de *Grâce* et de *Libre arbitre*.

**Figure 6**

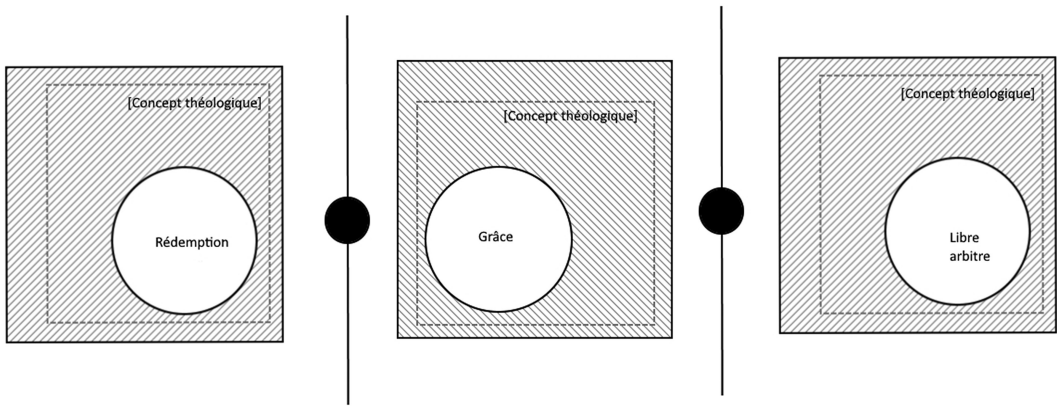
*Incompatibilité entre deux concepts théologiques*



La situation est d'ailleurs plus complexe qu'il n'y paraît à première vue. La transformation de la notion de *Grâce* par les jésuites résulte d'une double incompatibilité : la notion de *Grâce* apparaît également incompatible avec la théorie de la *Rédemption* selon laquelle, par son sacrifice, le Christ aurait racheté l'ensemble du genre humain. C'est ce qu'indique la figure suivante :

**Figure 7**

*Double incompatibilité entre la notion de Grâce et deux autres concepts théologiques*



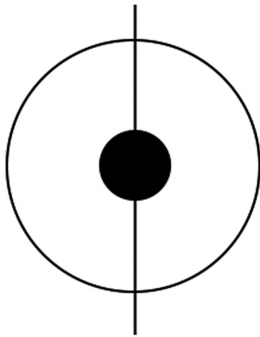
Cependant, contrairement à l'énoncé *Ce restaurant est bon mais cher*, l'incompatibilité est plus proche d'une contradiction formelle en ce sens que, du point de vue des théologiens, il est impossible de la résoudre par une décision subjective (TA, § 46 : 263). Il est en effet exclu d'établir une hiérarchie fondée sur l'expression d'une préférence entre ces trois notions ; ces trois notions sont perçues comme des concepts théologiques d'égale importance dans la tradition examinée. L'incompatibilité se reporte donc sur une des trois notions, la notion de *Grâce*, qui est appelée à être redéfinie, réaménagée. Pour commenter cette transformation, deux solutions s'offrent à l'observateur et leur complémentarité est très révélatrice :

- a) Parmi les associations de notions, une association de type quasi-logique pourrait faire l'affaire. Il s'agit de la division du tout dans ses parties ;
- b) Il pourrait aussi s'agir d'une dissociation proprement dite qui, par la suite, entraîne de nouvelles associations de notions.

La nuance entre ces deux possibilités serait également aisément schématisable :

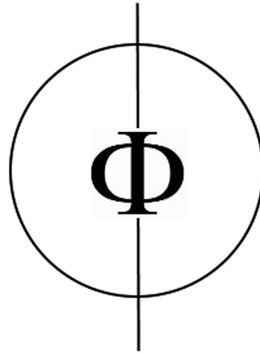
**Figure 8**

*La division du tout en parties en tant qu'association de notions*



**Figure 9**

*Dissociation d'une notion*



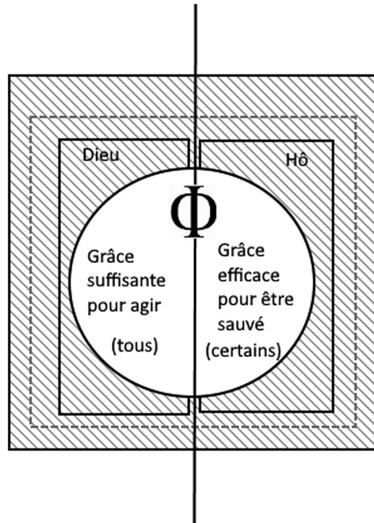
Dans le premier cas, l'association apparaît au premier plan et la dissociation semble seconde ; dans l'autre cas, c'est l'inverse<sup>5</sup>. À notre avis, il est plus raisonnable de considérer qu'il s'agit d'une dissociation dans la mesure où elle apparaît comme une innovation de la part des molinistes. Les Figures 4 et 5 avaient déjà montré la complémentarité et l'ambiguïté qui sont ménagées entre l'incompatibilité (association de type quasi-logique) et la dissociation. La nuance entre la définition (en tant qu'association de notions de type quasi-logique) et la définition dite « dissociative » (en tant que technique de dissociation présentée au § 95) montrerait que l'inventaire des techniques argumentatives est construit de manière à mettre en évidence la complémentarité entre associations et dissociations de notions.

Pour en revenir aux querelles théologiques du XVII<sup>e</sup> siècle, la double incompatibilité opérée entraîne une dissociation interne à la notion de Grâce :

<sup>5</sup> Il faut également prendre en compte la possibilité d'une association de notions qui aille de la partie au tout.

**Figure 10**

*Dissociation d'une notion résultant du constat d'une double incompatibilité*



Un point intéressant est que comme l'incompatibilité était dédoublée (cf. Figure 5), la dissociation repose sur la création des deux sous-classes qui se superposent. La dissociation correspond ainsi à la combinaison de deux critères : a) identification de la source de ce bienfait et identité du bénéficiaire. La première forme de grâce vient de Dieu seul mais, à un moment donné, elle a été accordée à tous les hommes (adaptation de la notion à la théorie du rachat). Au contraire, la seconde forme de grâce dépend des actes et donc des choix humains (adaptation de la notion à la théorie du libre arbitre). Dans ce second cas, seuls certains hommes seront sauvés, ce qui semble conforme avec les conséquences de la définition traditionnelle de la grâce. L'équilibre, délicat, de ce réaménagement notionnel est formulé avec prudence : Grâce suffisante « pour agir » / Grâce efficace « pour être sauvé ».

Pascal, comme les jansénistes, refuse cette distinction. La grâce est, de toute éternité, donnée par Dieu seul et seulement à certains hommes ; les hommes sont prédestinés. Ce sont donc les notions de *Libre arbitre* et la théorie du rachat qui doivent être réaménagées ou, simplement, gardées à distance de manière à ne pas attirer l'attention.

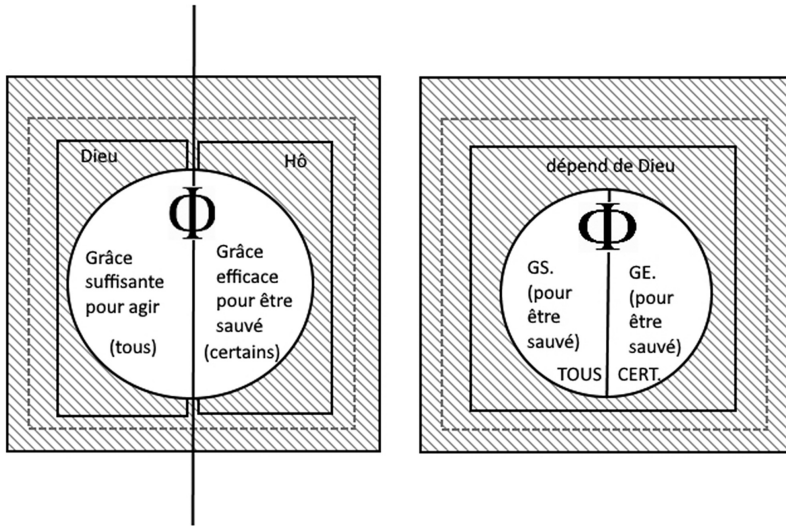
Enfin, la position des dominicains, celle qui au premier chef fait l'objet de la réfutation pascalienne, tente un compromis. Comme chez les jésuites, la grâce suffisante est accordée à tous les hommes mais, cette fois, cette grâce est simplement dite suffisante, de sorte qu'elle semble suffisante « pour le salut » et non plus

suffisante « pour agir ». Cependant, cette grâce dite « suffisante » apparaît comme insuffisante puisqu'elle devrait encore être complétée par une grâce efficace accordée cette fois par Dieu seul et à certains hommes seulement. La contradiction que Molina avait cherché à articuler réapparaît, sous une autre forme.

La structure correspondant à la Figure 10 reste globalement identique simplement les titres des notions-outils et des reclassements opérés se trouvent modifiés :

**Figure 11**

*Comparaison de la position des jésuites et de celle des dominicains*



La réfutation de Pascal consiste alors à montrer que dans le découpage notionnel proposé par les dominicains se cache une contradiction qu'il présente comme insurmontable et rédhibitoire. Dans notre manière d'aborder les choses, le problème se situe au niveau de la construction des sous-classes : à partir du moment où la Grâce efficace ne dépend plus de l'homme mais de Dieu, la grâce ne peut plus être dite suffisante « pour agir », suffisante pour que l'homme puisse agir pour son propre salut.

Il est intéressant de comparer les démarches de Ducrot et de Perelman sur ce point. Pour Ducrot, à peu près tout ce que nous avons dit jusqu'à présent est un élément de contextualisation ; le cœur de l'analyse repose sur la formule : *tout le suffisant donc tout le nécessaire*. Ducrot se sert alors d'outils formels pour montrer que Pascal fait allusion à la distinction entre condition nécessaire et condition suffisante et que l'utilisation que le mathématicien fait du terme « suffisant » est un coup de force argumentatif. En langage perelmanien, on pourrait dire que

c'est un argument quasi-logique en ce sens qu'il tente de capter le prestige généralement accordé aux démonstrations formelles qu'elles soient d'ordre logique ou mathématique. Cependant, d'un point de vue néo-rhétorique, cette lecture n'est pas la seule possible et sûrement pas la première à laquelle on songe. En effet, les notions de condition suffisante / condition nécessaire ne sont pas directement disponibles<sup>6</sup> dans l'inventaire (certes ouvert) des arguments quasi-logiques. Dans une terminologie qui est étrangère au TA, la démarche pourrait être décrite comme participant d'une analyse énonciative. Ainsi, ce qui est au cœur de l'analyse, ce sont 4 configurations notionnelles portées par 4 discours concurrents : les argumentations des jésuites, des jansénistes ou des dominicains et, finalement, la réfutation pascalienne de la version défendue par les dominicains. Dans cette perspective, la position des dominicains apparaît comme une citation déformée du raisonnement des jésuites : du point de vue pascalien, à la rupture avec les textes de Saint Augustin, elle ajoute le défaut de l'incohérence interne.

## 7. Remarques finales

À notre connaissance, dans la littérature francophone, les commentateurs se sont peu intéressés à la complémentarité entre associations et dissociations de notions. C'est la dissociation qui attire toute la lumière car elle apparaît comme doublement originale. D'une part, comme les auteurs y insistent eux-mêmes (Olbrechts-Tyteca, 1979 : 82), il s'agit d'une proposition novatrice. D'autre part, la dissociation serait une condition indispensable à toute pensée originale ; cette dernière remarque qui ne manque pas d'attirer l'attention n'implique cependant en rien que l'utilisation d'une dissociation garantisse l'originalité de la configuration notionnelle proposée puisque certains couples philosophiques sont consacrés par une longue tradition (TA, § 91 : 562). Les auteurs anglo-saxons (e.g. Franck, 2020) sont, quant à eux, plus sensibles à la complémentarité entre associations et dissociations de notions qui a guidé la présente réflexion. Cependant, une fois repérée la citation de Gourmont et de Burke, cette complémentarité est ensuite rapidement insérée dans un réseau de références propres à la tradition anglo-saxonne. On risque alors de perdre de vue, une spécificité que le TA (§ 89 : 552) assigne

---

<sup>6</sup> Si ces termes figurent dans l'index, ils ne sont pas étiquetés comme des schèmes mais comme des concepts qui permettent de décrire l'usage des « techniques argumentatives proprement dites ».

à sa conception de l'association et de la dissociation des notions : les notions dans l'argumentation ne sont pas à considérer comme des briques qu'auraient laissées intactes les opérations d'associations et de dissociations subies sous la pression de l'argumentation, mais comme des réaménagements de notions décomposables et déformables. Cette dimension est, par contre, très présente dans les commentaires francophones (e.g. Herman & Micheli, 2003, 20 ; Dominicy, 2002 : 9).

La complémentarité entre associations et dissociations de notions gagnerait à être aussi examinée dans le cadre de ce projet de réévaluation des problèmes de sémantique dans une perspective rhétorique. Comme le suggèrent les auteurs eux-mêmes (TA, § 35 : 189), ce projet donne sa structure au *Traité*. De notre point de vue, les chapitres II et III de la seconde partie offrent un fondement grammatical à l'étude des notions qui sera ensuite prolongée au niveau rhétorique.

Deux exemples empruntés à Ducrot ont été utilisés comme un moyen de prendre de la distance par rapport aux exemples du TA qui sont souvent relativement complexes, porteurs de spécificité et, surtout, qui se succèdent avec rapidité. Un des mérites de cette mise à distance est sans doute qu'elle permet d'illustrer le principe de complémentarité qui structure l'inventaire perelmanien des techniques argumentatives : pas d'association sans dissociation et vice-versa. Au cours de cette brève contribution dont la valeur reste avant tout exploratoire, on a observé que certaines courroies de transmission avaient été aménagées pour respecter ce principe de complémentarité : l'association quasi-logique la plus prototypique est l'incompatibilité ; or, l'une des spécificités de l'incompatibilité est de déclencher, de manière régulière, une dissociation de notions. La définition qui est, elle aussi, une association de type quasi-logique trouve son pendant parmi les dissociations sous le label : définition dissociative. Il y a tout lieu de penser que la nuance entre les deux se ramène à déterminer si la dissociation a lieu à l'avant-plan (on parlera alors de définition dissociative) ou à l'arrière-plan (on aura alors affaire à une définition ordinaire qui attire notre attention sur une association de notions tout en rejetant, dans le vague et de manière plus implicite, d'autres associations possibles).

Le présent exposé a avant tout traité des associations de type quasi-logique, mais certaines correspondances sont faciles à repérer entre associations fondées sur la structure du réel et dissociations argumentatives : le lien causal proprement dit (qui suscite relativement peu de développements) est inclus dans une classe « liaisons de succession » où la part belle est faite à des associations qui constituent des couples dissociatifs : fin / moyen ; fin / conséquence. Un second groupe d'associations fondées sur le réel joue un rôle très important dans l'économie du système : il permet de transférer les qualifications attribuées à un acte donné à la personne qui a accompli cet acte (ou inversement). Dans la brève analyse que



nous avons proposée du procès de Rosalie Prudent (§ 3<sup>e</sup>), on a vu comment cette technique permet d'assurer un passage entre des notions comme le caractère de l'accusée et les faits jugés. Il faudrait encore ajouter un mot sur le lien entre ces éléments et la sanction autorisée par la Loi : la manière dont la notion-maîtresse (qu'il s'agisse du portrait de l'accusée ou de la qualification des faits à juger) aura été stabilisée et précisée restreindra, dans un sens ou dans un autre, la gamme (en principe préétablie) des condamnations possibles. On a pu s'étonner (§ 6) du peu de cas qui est fait dans le TA à la condition et, *a fortiori*, à des enchaînements d'allure formelle du type *Si p alors q*. Dans le même esprit, on pourrait constater que le système n'est pas non plus prioritairement fondé sur une cellule du type *A donc C*. S'il en est ainsi, c'est que le système repose d'abord sur des opérations de déformations et de reclassement des notions ; la mise à distance du syllogisme qui, elle aussi, suscite parfois l'étonnement (Bouchard & Valois, 1983) est une autre conséquence de ce choix théorique.

Le prototype des associations qui fondent la structure du réel est sans aucun doute l'exemple en tant qu'il permet de créer une règle et de nouveaux classements à partir d'un ou de plusieurs cas particuliers. On devine que, au contraire des exemples étudiés jusqu'à présent, le mécanisme d'élargissement de l'extension sera à l'œuvre dans ce type de schèmes. Les cas les plus difficiles à représenter de manière stabilisée resteront sans doute les analogies et les métaphores.

Convenons-en, les reconstructions proposées n'ont pas la précision et la rigueur des analyses linguistiques de Ducrot. La liberté laissée à l'usager de la théorie est très grande d'autant que, si l'on en croit Dupréel (1911 : 517), une idée (une notion) peut « correspondre à un mot ou à des affirmations explicites de toute espèce » ; on peut alors se demander quel type de correspondance existe, au niveau supérieur, entre les segments explicites d'un énoncé et les associations de notions. L'usager de la théorie peut avoir, à chaque instant, l'impression de produire une explication *ad hoc*. Mais n'est-ce pas précisément un raisonnement *ad hoc*, un raisonnement conforme à ses besoins, que l'argumentateur cherche à former ? Le mérite de la théorie consiste alors à trouver des régularités face à ce phénomène éminemment mouvant : la difficulté vient bien moins de l'absence d'outil pour décrire une argumentation que des choix à faire parmi les notions à observer, parmi les multiples manières de se représenter un même argument ou parmi les niveaux d'analyse à distinguer (TA, § 44 : 251–252). Le TA cherche, ainsi, à classer la multitude des manipulations notionnelles possibles en une cinquantaine de schèmes argumentatifs. Ce premier inventaire « rhétorique » est ensuite restreint à un petit nombre de motifs « sémantiques » (trois types d'associations et des dissociations), eux-mêmes réductibles à deux grands principes complémentaires. La formulation que Perelman et Olbrechts-Tyteca donnent

à cette complémentarité qui unit les mécanismes associatif et dissociatif n'est pas sans rappeler l'élégance de ce précepte de dessinateur : *une ligne unit autant qu'elle divise*. Un principe analogue, fondé sur la complémentarité des notions mises en œuvre, vient prolonger la réflexion néo-rhétorique et la distinction, à notre avis essentielle, qu'elle instaure entre un avant-plan et un arrière-plan :

Un concept n'est possible que par une sorte de refoulement de tout ce qu'on ne fait pas entrer dans sa compréhension ; il appelle le correctif de son anti-concept. Ce mot ne veut pas dire son contraire, mais son complément.

(Dupréel, 1939 : 73)

## Références citées

- Amossy, R. (2002). Nouvelle Rhétorique et linguistique du discours. Dans R. Amossy & R. Koren (éds), *Quelles politiques pour les nouvelles rhétoriques?* (153–172). L'Harmattan.
- Arnaud, A. & Nicole, P. (1970[1662]). *La logique ou l'art de penser*. [Introduction de Louis Marin]. Flammarion.
- Bouchard, G. & Valois, R. (1983). (Nouvelle) rhétorique et syllogisme. *Laval théologique et philosophique* 39(2), 127–150.
- Dominicy, M. (1993). De Toulmin à Perelman : analyse d'un réseau argumentatif. *Verbum*. 1, 2, 3, 241–50.
- Dominicy, M. (2002). La dimension sémantique du discours argumentatif : le travail sur les notions. Dans R. Amossy & R. Koren (éds), *Quelles politiques pour les nouvelles rhétoriques?* (121–150). L'Harmattan.
- Ducrot, O. (2004). Argumentation rhétorique et argumentation linguistique. Dans M. Doury & S. Moirand (éds), *L'argumentation aujourd'hui. Positions théoriques en confrontation* (17–34). Presses Sorbonne Nouvelle.
- Ducrot, O. (2007/1980). *Les échelles argumentatives*. Minuit.
- Ducrot, O. (1989). *Logique, structure, énonciation. Lectures sur le langage*. Minuit.
- Dupréel, E. (1939). *Esquisse d'une philosophie des valeurs*. Alcan.
- Dupréel, E. (1911). Sur les rapports de la logique et de la sociologie ou théorie des idées confuses. *Revue de métaphysique et de morale* 19(4), 517–522.
- Franck, D. A. (2020). The Origins of and Possible Futures for Chaim Perelman and Lucie Olbrechts-Tyteca's Dissociation of Concepts. *Philosophy and Rhetoric* 53(4), 385–99.

- Herman, T. & Micheli, R. (2003). Renforcement et dissociations des valeurs dans l'argumentation politique. *Pratiques* 17(118), 9–28.
- Olbrechts-Tyteca, L. (1968). Les couples philosophiques : une nouvelle approche. *Revue Internationale de Philosophie* 33(127/128), 81–98.
- Perelman, C. (2012/1955). Les notions et l'argumentation. Dans M. Meyer (éd.), *Rhétoriques* (109–131). Presses de l'Université.
- Perelman, C. & Olbrechts-Tyteca, L. (2008/1958). *Traité de l'argumentation. La nouvelle rhétorique*. Préface de Michel Meyer. 6<sup>e</sup> éd. Presses de l'Université de Bruxelles.
- Plantin, C. (2016). *Dictionnaire de l'argumentation. Une introduction aux études d'argumentation*. ENS éditions.
- Plantin, C. (2010). Les instruments de structuration des séquences argumentatives. *Verbum* 32(1), 31–51.
- Plantin, C. (1990). *Essais sur l'argumentation. Introduction linguistique à l'étude de la parole argumentative*. Kimé.




**Marco Fasciolo**

Sorbonne Université  
France

 <https://orcid.org/0000-0001-9575-9594>

**Qianqian Zeng**

Sorbonne Université  
France

 <https://orcid.org/0009-0001-3798-1493>

# La synonymie : une relation lexicale comme les autres<sup>\*</sup>

**Synonymy: a lexical relation like any other**

## Abstract

This article deals with the epistemological consistency of synonymy. When synonymy is defined as equality of meaning, it turns out to be a factually impossible relation. Therefore, scholars distinguish between full synonymy (ideal concept) and partial synonymy (real phenomenon). We argue that this distinction must be discarded in order to maintain consistency in synonymy. We propose that synonymy does not consist in the absence of different semantic nuances between meanings, but rather in the condition for finding different semantic nuances between meanings. This condition is interpreted as sharing the same “predicative schema” in the sense of Gaston Gross.

## Keywords

Synonymy, meaning, equivalence, connotative meaning, lexicon, lexical relations

---

\* L'article *Pour une véritable fonction « synonymie » dans un traitement de texte* a été le premier texte de Gaston Gross que l'auteur de cette contribution a lu. Il en est né un lien d'amitié et une collaboration constante. L'autrice de cet article, en revanche, a connu les travaux de Gaston Gross à l'occasion de son mémoire de Master 2, portant sur la synonymie, sous la direction du premier auteur. Gaston Gross était une personne généreuse, qui n'hésitait pas à partager ses connaissances et à offrir ses intuitions aux élèves. Il nous a paru donc qu'écrire un article ensemble – sur la synonymie – était une façon appropriée de lui rendre hommage, en montrant que son attitude et ses idées continuent.

## 1. Introduction

### 1.1. Synonymie : une relation lexicale impossible ?

La synonymie, apparemment, n'est pas une relation lexicale comme les autres<sup>1</sup>. Lyons (1977 : 203–204) présente les relations lexicales sur le modèle des postulats de signification (*meaning postulates*) de Carnap (1952). L'implication simple (ou l'inclusion) est le modèle de l'hyponymie ; les termes contraires sont le modèle de l'antonymie au sens strict (l'opposition graduelle) ; les termes contradictoires sont le modèle de l'opposition complémentaire ; l'identité (la double implication ou l'équivalence logique) est le modèle de la synonymie. Dès que la synonymie est définie comme une identité de sens, cependant, elle devient une relation lexicale impossible, car, dans les faits, il est (trop) facile de dégager des différences entre des candidats synonymes (cf. Honeste, 2007 : 170). La synonymie semble donc s'autodétruire à cause de sa propre définition. Cette conclusion paradoxale est acceptée passivement ; elle est même transformée en un principe théorique : *the principle of non synonymy* (cf. Goldberg, 1997 : 67)<sup>2</sup>.

La démarche standard consiste à affaiblir la définition de départ en distinguant entre *synonymie totale* ou *absolue* et *synonymie partielle*. Si la première est une limite virtuelle, peut-être inatteignable, la seconde est un phénomène productif, dont les sémanticiens peuvent décrire les réalisations avec minutie. Le travail classique de Girard (1718) est, à ce propos, exemplaire. Une telle description est sans doute cruciale pour la lexicographie, mais elle ne doit pas cacher la donnée de fond : à savoir que les dimensions de différenciation entre les synonymes, répertoriées par les auteurs (cf. entre autres : Baldinger, 1984 : 187 ; Lehmann & Martin-Berthet, 1998 : 57 ; Apresjan, 2000 : 26–34 ; Ježek, 2016 : 174–175), sont autant de raisons de *partialité* de la synonymie. Mais, dans la mesure où des synonymes sont partiels, ils ne sont pas complètement synonymes. La synonymie semble donc une relation lexicale asymptotique qui ne se réalise jamais, sauf peut-être dans des cas exceptionnels. La poussière théorique balayée sous le tapis de la description est toujours là<sup>3</sup>. Si l'on soulève ce tapis, deux difficultés majeures se présentent.

La première difficulté est que la distinction partielle/totale est unique à la synonymie. Toutes les autres relations lexicales se réalisent tout court : il n'y a pas

<sup>1</sup> Pour une synthèse de l'histoire de la notion de synonymie, cf. Doualan (2013, 2014).

<sup>2</sup> La validité du principe de non synonymie est débattue dans le cadre de la grammaire de constructions : Cf. Uhrigh (2015) pour une critique et Leclercq et Morin (2023) pour une défense.

<sup>3</sup> L'expression est de F. Neveu, que nous remercions.

d'hyponymie, d'opposition, etc. partielle plutôt que totale. Le mot *animal* est plus général que *mammifère*, qui est plus général que *chien*, qui, à son tour, est plus général que *teckel*. Mais *animal* n'est pas plus hyperonyme que *chien* et *chien* n'est pas un hyperonyme partiel sous prétexte qu'il est moins général qu'*animal*. *Laid* et *moche* (qui peuvent être synonymes) sont contraires de *beau*, mais l'un n'est pas plus contraire que l'autre. Dans *froid*, *frais*, *tiède* et *chaud*, il y a certainement une gradation, mais la relation lexicale d'antonymie n'est pas pour autant partielle : elle est univoque et parfaitement réalisée. À cet égard, le titre de Zhurauliova (2021) est révélateur : *La synonymie exacte existe-t-elle ?* À notre connaissance, aucun auteur n'a posé cette question pour les autres relations lexicales.

La seconde difficulté concerne le modèle théorique de la synonymie. Ce modèle est rendu explicite par Cruse (1986) :

Synonyms [...] are lexical items whose senses are identical in respect of 'central' semantic traits, but differ, if at all, only in respect of what we may provisionally describe as 'minor' or 'peripheral' traits [...].

(Cruse, 1986[1997 : 267])

À bien y regarder, l'illustration parfaite du modèle centre/périphérie évoqué ci-dessus est le rapport ensemble à sous ensemble ou une taxinomie basée sur genre et différences spécifiques. Si elle est prise à la lettre, la précédente est une définition de l'hyponymie. Cette conclusion n'est pas tout à fait étonnante, car, comme Rey-Debove (1997 : 92) le souligne, la co-hyponymie est bien la synonymie au sens d'Aristote (*Catégories I, 1a*) (cf. Doualan, 2014 : 410). Cependant, reste le fait qu'une telle conception de la synonymie n'est pas la nôtre : les co-hyponymes ne sont pas des synonymes ! Si, d'un point de vue métaphysique, *homme* et *bœuf* sont synonymes parce qu'ils sont deux espèces d'animaux, du point de vue du lexique, le fait qu'*homme* et *bœuf* désignent deux espèces d'animaux est précisément la raison pour laquelle ces mots ne sont pas synonymes. L'adoption de la définition précédente de la part de Cruse (*cit.*), par ailleurs, est autant plus étonnante que cet auteur, juste une page avant, aborde implicitement la difficulté que nous venons de soulever : « [...] it does not follow that the more semantic traits a pair of word share, the more synonyms they are » (Cruse, 1986[1997 : 266]). Nous y reviendrons au §3.2.

Pour le moment, nous en retenons la conclusion suivante : le fait que la synonymie est la seule relation lexicale qui ne se réalise pas complètement et le fait qu'il n'y a pas un modèle clair pour la concevoir sont des indices que sa conception est affectée par un vice épistémologique de fond.

## 1.2. But, structure et limites de la contribution

Le but de cette contribution est d'offrir une façon de penser la synonymie qui dépasse les difficultés soulevées sous §1.1. Dans cette conception alternative, la synonymie est une relation lexicale comme les autres, c'est-à-dire une relation lexicale qui ne se vérifie pas seulement dans des cas exceptionnels, et qui, toutes les fois qu'elle se vérifie, elle se vérifie tout court.

La structure de la contribution est la suivante. Le §2 est consacré à introduire notre idée de fond : la synonymie comme présupposé. Le §3 illustrera cette idée à travers l'approche de Gross & Clas (1997), Gross (1998) et, plus en général, de la théorie des classes d'objets (cf. Gross, 2012). Au §4, finalement, nous reviendrons sur une série de conséquences concernant : le lien entre synonymie, interchangeabilité et polysémie (§4.1) ; le caractère fondamental (ou pas) de la synonymie (§4.2) ; la place de la synonymie parmi les autres relations lexicales (§4.3) ; l'impulsion à trouver toute sorte de différences entre les synonymes (§4.4). Avant de commencer, cependant, nous aimerions préciser trois points.

Premièrement, même si notre outil d'étude sera la structure de la phrase (cf. §3.1), nous nous bornerons à la synonymie comme relation entre lexèmes : nous ne considérerons donc pas la relation de paraphrase ou reformulation au niveau des énoncés.

Deuxièmement, nous nous situons bien à un niveau sémantique et non pragmatique. En parlant de *synonymie*, nous entendons une propriété stable des signifiés des lexèmes et non le référent des groupes nominaux. Cette précision s'impose à cause d'une ambiguïté terminologique spécifique à la tradition sémantique française, où le terme *référent* est polysémique. D'un côté, *référent* peut désigner l'objet indiqué, dans une situation contingente, en employant un groupe nominal : par exemple, la même personne à laquelle, ici et maintenant, on pointe en disant *La Comtesse*, *Ma belle-mère*, *Il Duce*, etc. Il est clair que plusieurs groupes nominaux peuvent avoir le même référent, mais cela ne rend pas leurs noms synonymes ; au contraire, l'une des raisons de l'intérêt d'indiquer le même référent avec plusieurs groupes nominaux est précisément que leurs noms ne sont pas synonymes<sup>4</sup>. De l'autre côté, *référent* peut désigner le concept (signi-

---

<sup>4</sup> Cela est vrai, à plus forte raison, pour des énoncés communiquant le même message : *La fenêtre est ouverte* et *Le voisin fait un barbecue* peuvent communiquer le même message (par exemple : le chat s'est sauvé chez le voisin), mais ils ne sont ni des synonymes, ni des paraphrases. Nous observons, incidemment, une différence entre synonymie et paraphrase. La synonymie est une relation entre les significations de plusieurs lexèmes. Une paraphrase, en revanche, n'est pas une relation entre les significations de plusieurs énoncés, mais la reformulation (littérale) du message qu'un énoncé veut communiquer.

fié) codé par un lexème : il s'agit du sens de *référent* dans l'expression *sémantique référentielle*. C'est ce dernier sens qui est pertinent pour notre discussion, ainsi que dans l'expression *équivalence référentielle* appliquée à la synonymie (cf. §2.1).

Troisièmement, nous ne prenons pas en considération les prépositions ou expressions prépositionnelles en position de connexion transphrastique. Considérons, à titre d'exemple : *Grand-père a réparé la chaise avec de la ficelle*. Dans cet exemple, *avec* fait l'objet d'un processus d'enrichissement inférentiel qui, dans les faits, rend cette préposition sémantiquement équivalente à *à l'aide de*. Ce processus d'enrichissement, par ailleurs, n'est pas pragmatique, mais systématique : il s'active par défaut. Quoi qu'il en soit, ce phénomène restera au-delà de la portée de cette contribution.

## 2. Pour une conception alternative de la synonymie

### 2.1. Synonymie comme présupposé

Considérons la question suivante :

- 1) a. *Quelle est la différence entre « flingue » et « pistolet » ?*

Même pour quelqu'un qui n'est pas un linguiste professionnel, la réponse à (1a) est immédiate : entre *flingue* et *pistolet*, il y a une différence de registre. Observons maintenant la question (1b) :

- 1) b. *Quelle est la différence entre « flingue » et « voiture » ?*

Même pour quelqu'un qui est un linguiste professionnel, de prime abord, cette question est déroutante ; et cela bien que la réponse soit identique à la précédente. Nous ne sommes pas en train de soutenir que, face à (1b), l'on ne remarquera pas la différence de registre en jeu ; nous sommes en train de suggérer, plus simplement, que cela est moins immédiat qu'en (1a). L'asymétrie entre les questions (1a) et (1b) montre un fait décisif. Dans un cas, la différence de registre familier/standard est saillante parce que les mots *flingue* et *pistolet* sont synonymes ; dans l'autre, la même différence n'est pas également saillante parce que les mots *flingue* et *voiture* ne sont pas synonymes. La synonymie n'est donc pas l'absence de différences (sémantiques), mais la condition préalable pour qu'il soit



pertinent de mettre en valeur la présence de différences (sémantiques). Cela est, *in nuce*, l'idée que nous défendons.

Cette idée nous semble implicite dans la conception de la synonymie proposée par Tamba (1988). Observons le passage suivant :

L'équivalence référentielle n'est donc pas un effet, mais une condition de la synonymie puisque celle-ci ne peut que surajouter un trait sémantique connotatif à une dénotation préétablie.

(Tamba, 1988 : 85, notre soulignement)

Dans ce passage, et dans d'autres du même texte, la synonymie apparaît comme un phénomène à deux faces : une identité de sens dénotatif et une différence de sens connotatif. Ces deux faces ne sont pas au même niveau : le partage du socle dénotatif est le présupposé de la synonymie, alors que l'ajout de nuances connotatives est le posé de la synonymie. Tamba (1988 : 85–88) insiste sur ce dernier aspect en décrivant la synonymie comme un processus actif, qui *ajoute* et *surajoute* des nuances. Dans ce cadre, *être synonymes* consiste dans le fait de manifester des différences sémantiques sur le présupposé d'une identité de fond. Certes, il faudra revenir sur l'objet de cette identité (cf. §3.2). Cependant, une chose est déjà claire : les différences évoquées ci-dessus ne tracent pas une limite à la synonymie, en la diminuant ou en la rendant partielle, mais elles en sont une conséquence.

Nous comprenons par là le vice épistémologique affectant la conception standard discutée sous §1.1. Si l'on conçoit les synonymes comme des mots dont les sens partagent un centre et se différencient pour une périphérie, l'on admet implicitement que ce centre et cette périphérie soient au même niveau : de sorte que la seconde limite le premier, en rendant ainsi ces mots *synonymes jusqu'à un certain point*. Selon nous, en revanche, l'objet de l'identité qui définit la synonymie et les différentes nuances sémantiques entre les synonymes ne sont pas au même niveau, mais les secondes présupposent le premier. Dans cette perspective, si les synonymes peuvent être comparés par rapport à un éventail de différences, la synonymie n'est pas pour autant une question de degrés, car elle reste présupposée. Il n'y a donc pas, à proprement parler, des *synonymes partiels* opposés à des *synonymes totaux*, mais simplement des *synonymes tout court*.

## 2.2. Contre la distinction entre synonymes partiels et totaux

Rejeter la notion de synonymie partielle implique *ipso facto* le rejet de la notion de synonymie totale. À ce propos, il nous paraît intéressant de revenir sur la po-

sition de Kleiber (2009), qui défend le bien-fondé de la relation de synonymie en argumentant notamment en faveur de l'existence de synonymes totaux.

Selon Tamba (1988 : 85), on l'a vu, des mots comme *flingue* et *pistolet* sont synonymes parce qu'ils partagent le sens dénотatif et diffèrent exclusivement au niveau du sens connotatif. Par *composante connotative*, il faut comprendre toute composante sémantique non vériconditionnelle : des informations relevant du registre, de la fonction interpersonnelle, etc. Inversement, pour *composantes dénотatives* il faut comprendre toute composante sémantique ayant un impact sur les conditions de vérité. Kleiber (2009), pour sa part, souligne que lesdites *composantes connotatives* ne sont pas, à bien y regarder, des traits sémantiques, mais des caractéristiques du mot en tant que symbole – ou objet social – appartenant à un code linguistique :

[...] ces marques connotatives diachroniques, géographiques, diastratiques, expressives, etc., ne sont pas l'apanage du sens du mot synonymique, mais concernent le mot dans sa globalité et donc ne touchent nullement à l'identité sémantique qu'il peut y avoir ou non entre deux synonymes. Il n'y a donc pas lieu de parler à leur propos de *signifié* qu'il soit connotatif ou non.

(Kleiber, 2009 : 16)

Par là, Kleiber (*cit.*) accomplit un pas ultérieur par rapport à Tamba (*cit.*). Selon Kleiber (*cit.*), des mots comme *flingue* et *pistolet* sont synonymes *a fortiori*, car ce qui les différencie ne relève même pas de la sémantique. Ils sont donc des synonymes totaux. Cependant, dans cette approche, si la synonymie totale n'est plus un mythe, elle demeure quand même rare, car elle se réduit essentiellement à des cas comme *flingue/pistolet*, *mère/maman*, *voiture/bagnole*, etc., c'est-à-dire aux mots où les seules différences sont connotatives au sens strict du terme<sup>5</sup>. Cela entraîne une situation paradoxale.

Si les couples *flingue/pistolet*, *mère/maman*, etc. sont des synonymes totaux, alors ils sont également les cas les plus prototypiques de synonymie (car la synonymie se réalise pleinement seulement dans les synonymes totaux). Mais considérons des exemples comme les suivants :

- 2) a. *Ta veste est un peu étroite / juste.*
- b. *La défaite du PSG m'a abattu / découragé / démoralisé.*
- c. *Il lui murmurait / chuchotait / susurrait quelque chose.*

<sup>5</sup> La même remarque est valable pour Zhurauliova (2021), qui défend une thèse analogue à Kleiber (*cit.*).

d. *Elle est une petite fille douée / brillante.*

e. *Il a poussé des cris aigus / perçants.*

D'une part, les différences entre les mots soulignés ne peuvent pas être réduites à des composantes connotatives au même sens de *flingue/pistolet* : la différence entre *aigu* et *perçant* en (2e), entre *doué* et *brillant* en (2d), entre *murmurer* et *chuchoter* en (2c), etc. semble bel et bien sémantique. D'autre part, nous avons l'intuition que ces mots fonctionnent ici comme des synonymes. Force en est d'en conclure qu'il s'agit de synonymes partiels – c'est-à-dire, de cas non-prototypiques de synonymie – et que cela se vérifie pour la plupart des synonymes. Le paradoxe est que, d'un point de vue sémantique, les exemples sous (2) sont précisément les plus intéressants, alors que, les couples *flingue/pistolet*, etc. ne le sont guère ! Nous appelons ceci *paradoxe de la pertinence de la synonymie*. Le paradoxe de la pertinence de la synonymie montre que le prix à payer pour pouvoir admettre l'existence de synonymes totaux est très élevé : la synonymie devient une relation sémantiquement inintéressante.

Selon nous, pour maintenir la relation de synonymie, il n'y a pas besoin de prouver la possibilité de l'existence de synonymes totaux, mais il faut rejeter la distinction entre synonymie totale et synonymie partielle elle-même. Ce pas peut être accompli en adoptant la conception de la synonymie comme pré-supposé suggérée au §2.1. Dans ce cadre, non seulement la synonymie se réalise pleinement aussi bien dans *flingue/pistolet* que, par exemple, dans *perçant/aigu* sous (2e), mais le paradoxe évoqué ci-dessus disparaît, car les cas prototypiques de synonymie sont bien ces derniers. De notre point de vue, si les synonymes ayant des différences exclusivement connotatives sont peut-être les premiers qui viennent à l'esprit, ils sont également les cas les plus marginaux de cette relation lexicale.

### 3. Synonymie dans la théorie des classes d'objets

#### 3.1. Schéma prédicatif

La théorie des classes d'objets est décrite avec clarté dans un grand nombre de contributions : pour une présentation synthétique, nous renvoyons à Gross (1994), Le Pesant et Mathieu-Colas (1998) et Gross (2008) ; pour une présentation plus étendue, nous renvoyons à Gross (2012 : ch. 1 à 4) et, en italien, à Fasciolo et

Gross (2021 : ch. 1 à 5). Dans le cadre de cette théorie, l'unité de base pour l'étude du lexique n'est pas le mot, mais le schéma prédicatif : un prédicat saturé par des classes d'arguments. Ces classes d'arguments peuvent être des ensembles conceptuellement cohérents correspondant aux distinctions ontologiques de base (<humains>, <animaux>, <végétaux>, etc.), ou des classes sémantiquement homogènes construites à l'intérieur des précédentes (<moyens de transports publics>, <vêtements>, <bâtiments>, <arbres>, etc.). Les premières sont appelées *hyper-classes*, les secondes sont appelées *classes d'objets*. Un schéma prédicatif, en tant que tel, est une entité eidétique socialement partagée. La structure linguistique permettant de visualiser un schéma prédicatif de la façon la plus claire est le noyau de la phrase, et, en particulier, le noyau de la phrase où le prédicat est aligné avec le verbe. Voici les représentations de trois schémas prédicatifs du verbe *abattre* (cf. Gross, 2012 : 34–35) :

	SUJET	V	COD
	Arg.	Préd.	Arg.
3)	a. <humain>	<i>abattre</i>	<bâtiment>
	b. <humain>	<i>abattre</i>	<arbre>
	c. <événement>	<i>abattre</i>	<humain>
	...		

Soulignons deux aspects de la notion de schéma prédicatif pertinents pour notre discussion.

Tout d'abord, chaque schéma prédicatif est le pivot d'un éventail de propriétés : parmi ces dernières, nous nous intéressons maintenant au sens associé et aux relations lexicales impliquées. Dans le cadre de (3a), par exemple, il est impossible de confondre le sens d'*abattre* avec celui pertinent en (3b) ou (3c). Et inversement. Dans le cadre de (3a), *abattre* est contraire de *construire*, mais cela n'est pas vrai pour (3b) ou (3c). Dans le cadre de (3b), *abattre* est hyponyme de *couper*, mais cela n'est pas vrai pour (3a) ou (3c). Et ainsi de suite. Chaque schéma prédicatif est comme un fil rouge suivant lequel on peut mettre en évidence une région lexicale. Cette région lexicale – l'ensemble des propriétés impliquées par un schéma prédicatif – est ce que Gross (2012 : ch.2) appelle *emploi* : chaque schéma prédicatif sous (1) est donc le bout d'un emploi du verbe prédicatif *abattre*. Une telle notion d'emploi, on le voit, n'est pas discursive, énonciative ou pragmatique, mais sémantique et stable.

Ensuite, la notion de schéma prédicatif repose sur une conception de la signification qu'on peut qualifier de *fonctionnelle* par opposition à une conception *absolue* de la signification (cf. Fasciolo & Gross, 2021 : 36). Dans une conception

absolue, un mot a un sens et des relations lexicales dans l'abstrait : dans ce cadre, par exemple, on dira que *vieux* est contraire de *jeune* tout court. Dans une conception fonctionnelle, en revanche, un mot a un sens – et il entre dans des relations lexicales – par rapport à un schéma prédicatif. Dans ce cadre, on ne dira pas que *vieux* est le contraire de *jeune*, mais que *vieux* fonctionne comme contraire de *jeune* par rapport au schéma prédicatif <humain> *vieux*. Par rapport au schéma prédicatif <outil électronique> *vieux*, en revanche, *vieux* fonctionne comme contraire de *récent*, etc. Certes, dans la pratique, il peut s'avérer que *vieux* et *jeune* fonctionnent comme contraires par rapport à plusieurs schémas prédicatifs : c'est le cas, par exemple, des <humains> et des <vins>. Cette circonstance, cependant, est accidentelle : il s'agit d'une donnée *a posteriori* dépourvue de toute valeur prédictive.

Les remarques précédentes sont censées faire ressortir un point central dans le travail de G. Gross : à savoir, que le sens d'un lexème, tout en étant stable, n'a pas un contenu fixé dans l'abstrait, mais seulement en fonction d'un schéma prédicatif spécifique. La polysémie, elle, n'identifie pas un contenu plus général, qui engloberait les autres ou à partir duquel les autres peuvent être déduits. Le phénomène de la polysémie est plutôt un effet des ressemblances de famille entre les sens mobilisés par les différents schémas prédicatifs. Si des analogies plus ou moins étendues peuvent être parfois dégagées entre les différents sens, ces analogies n'autorisent pas pour autant à postuler un noyau commun ou une forme schématique de base (cf. Venant & Victorri, 2012). La même chose est vraie, *mutatis mutandis*, pour les relations lexicales. Puisqu'elles relient des sens des lexèmes, les relations lexicales reçoivent un contenu seulement en fonction de chaque schéma prédicatif.

Nous pouvons maintenant revenir à la synonymie, qui est l'une des relations lexicales.

### 3.2. Application à la synonymie

Au §2.1, nous avons laissé une question en suspens : si la synonymie est une forme d'identité, quel est l'objet de cette identité ? Dans le cadre théorique que l'on vient de présenter, la réponse est claire : il s'agit de l'identité du schéma prédicatif. Observons les exemples suivants :

- 3) a <humain> *abattre* <bâtiment> vs. *construire*  
*démolir* *bâtir*  
*ériger*

Affirmer qu'*abattre* et *démolir* sont synonymes signifie affirmer qu'il y a un schéma prédicatif – et plus en général un emploi – que ces mots partagent. Par rapport à ce schéma prédicatif, *abattre* et *démolir* sont contraires de *construire*. Le verbe *construire*, à son tour, partage un schéma prédicatif avec *bâtir* ou *ériger* (où les classes d'arguments sont, encore une fois, des <humains> et des <bâtiments>). Par rapport à ce schéma, *construire*, *bâtir* et *ériger* sont synonymes, et contraires d'*abattre* ou *démolir*.

Si la synonymie est définie comme une identité de schéma prédicatif, il est immédiatement évident que les différentes nuances sémantiques entre *démolir* et *abattre* par rapport au schéma <humain> *abattre* <bâtiment>, ou entre *décourager*, *démoraliser* et *abattre* par rapport au schéma <événement> *abattre* <humain>, ou entre *doué* et *brillant* par rapport au schéma <humain> *brillant*, ne rendent pas ces mots moins synonymes, mais présupposent notamment leur synonymie (cf. §2.1). Par là, l'opposition entre synonymies partielle et totale est privée de fondement. Insistons sur un point. Refuser cette opposition ne signifie pas qu'on ne puisse pas comparer les synonymes, au contraire ! (cf. §4.4). Les synonymes peuvent avoir toute sorte de différences et être plus ou moins ressemblants entre eux. La seule chose que le rejet de l'opposition entre synonymies partielle et totale implique est la suivante : les ressemblances et les différences ne rendent pas les synonymes plus ou moins synonymes, mais elles supposent la synonymie tout court.

Les exemples ci-dessus mettent en jeu des prédicats, mais des considérations semblables s'appliquent aux arguments. Affirmer que *flingue* et *pistolet* sont synonymes signifie affirmer qu'ils partagent la même classe d'objets des <armes à feu>, où ils sont co-hyponymes de *mitrailleuse*. À ce point, cependant, une précision s'impose. *Flingue* et *pistolet* ne sont pas co-hyponymes de *mitrailleuse* en tant que deux membres distincts de la classe d'objets des <armes à feu>, mais en tant que le même membre. Autrement dit, affirmer que *flingue* et *pistolet* sont synonymes signifie affirmer qu'ils sont deux mots en compétition pour la même position de co-hyponyme à l'intérieur du champ sémantique des armes à feu. En l'espèce, la différence entre ces mots est que l'un est marqué comme familier, alors que l'autre ne l'est pas. Des remarques analogues s'appliquent également à l'exemple d'*abattre* et *démolir*. Affirmer qu'*abattre* et *démolir* sont synonymes par rapport au schéma prédicatif sous (1a) ne signifie pas affirmer qu'ils sont deux contraires de *construire*, mais qu'ils sont le même contraire : ils sont en compétition pour la même position dans un champ oppositif.

Nous pouvons comprendre, par là, la différence entre synonymie et hyponymie évoquée au §1.1. À ce propos, Cruse (1986) écrit :

The key to this conundrum [= différencier la synonymie de l'hyponymie] lies in the nature of the differentiating characteristics : synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of contrastiveness

(Cruse, 1986[1997 : 266], notre soulignement)

Tamba (1988) est plus claire, car elle distingue explicitement entre :

[...] la synonymie qui ajoute une nuance relative à une catégorie et l'hyponymie qui ajoute un trait distinctif sous-catégorisant dans une hiérarchie catégorielle [...].

(Tamba, 1988 : 86, nos soulignements)

Or, pour qu'un mot puisse *ajouter une nuance sémantique sans pour autant ajouter un trait distinctif* par rapport à un autre mot, il faut qu'il soit en compétition avec ce dernier pour occuper la *même* position de co-hyponyme. En ce sens, la synonymie nous confronte à des différences sémantiques non-distinctives.

Nous pouvons également comprendre la raison derrière le paradoxe de la pertinence de la synonymie (cf. §2.2). En comparant les arguments élémentaires avec les prédicats, on remarque en effet une différence d'intérêt. Dans le cas des arguments élémentaires, constituant des champs sémantiques d'hyponymes avec un nombre limité de caractéristiques différentielles, la synonymie tend à se réduire à des différences non pas sémantiques, mais connotatives : *mère/maman, voiture/bagnole, pistolet/flingue*. Dans le cas des prédicats qui remplissent la fonction de sélectionner des classes d'arguments il semble raisonnable de s'attendre à ce que la synonymie concerne toute une palette de différences véritablement sémantiques : *démolir vs. abattre, abattre vs. décourager vs. démoraliser*. Ces derniers s'avèrent donc les exemples les plus riches et intéressants. Quoi qu'il en soit, deux points restent fermes : ils sont tous des synonymes tout court, au même sens, et ils sont des alternatives en compétition pour le même item dans un certain champ sémantique ou oppositif.

## 4. Conséquences

### 4.1. Synonymie vs. interchangeabilité vs. polysémie

Dans ce paragraphe, nous voudrions examiner un argument généralement accepté (cf. entre autres Mortureux, 1997 : 81 ; Niklas-Salminen, 1997[2003 : 110–111] ; Ploux & Victorri, 1992 : 2 ; Murphy, 2010 : 110–111). Considérons les exemples suivants :

- |    |                                    |    |                               |
|----|------------------------------------|----|-------------------------------|
| 4) | a. <i>Une doctorante brillante</i> | 5) | a. <i>Un diamant brillant</i> |
|    | b. <i>Une doctorante douée</i>     |    | b. * <i>Un diamant doué</i>   |

Selon cet argument, *brillant* et *doué* seraient des synonymes partiels parce qu'ils sont interchangeables seulement dans le « contexte » (4), mais non dans le « contexte » (5). Si *brillant* et *doué* étaient interchangeables également dans le « contexte » (5) – ainsi que dans tous les autres « contextes » possibles –, alors ils seraient des synonymes totaux. À la lumière de la notion de schémas prédicatif mis en avant au §3.1, ces affirmations sont trompeuses.

Tout d'abord, comme Kleiber (2009 : 20) le remarque également, la synonymie – et toutes les relations lexicales – n'est pas la propriété d'un mot, mais d'un sens d'un mot. Le sens est fonction d'un schéma prédicatif spécifique (cf. §3.1). Par conséquent, même en voulant utiliser la notion trompeuse de synonymie partielle, reste le fait que *brillant* et *doué* sont synonymes partiels dans la mesure où ils partagent un même schéma prédicatif (ils supposent la classe d'arguments des humains), et non parce que l'un entre dans un schéma prédicatif où l'autre n'entre pas ! Quoi qu'il en soit, la façon correcte de décrire le phénomène sous (4) est que *brillant* et *doué* sont synonymes – tout court – par rapport à la classe d'arguments des humains, et que cela est le présupposé rendant pertinent de mettre en valeur leurs différentes nuances sémantiques. Les sens de *brillant* en (4) et (5) ne sont pas directement comparables, car ils relèvent de schémas prédicatifs différents. La source de l'équivoque réside, on le voit, dans le manque d'une définition rigoureuse du « contexte », ou l'« environnement », où l'interchangeabilité a lieu. Si le terme *contexte* est précisé à travers la notion de *schéma prédicatif* (comme proposé par Gross 2010), cette équivoque s'éclaircit.

Ensuite, considérons les couples suivants :

- |    |                           |    |                             |
|----|---------------------------|----|-----------------------------|
| 6) | a. <i>Une pente douce</i> | 7) | a. <i>Une lumière douce</i> |
|    | b. <i>Une pente tenue</i> |    | b. <i>Une lumière tenue</i> |



Si l'on suit l'argument évoqué ci-dessus, force en est d'en conclure que *doux* et *tenue* sont synonymes, car ils sont bien interchangeables à travers les couples (6) et (7). Certes, il peut y avoir d'autres mots qui, dans les mêmes contextes, ne sont pas mutuellement interchangeables. Par exemple :

- 6) c. *Une pente facile*                      7) c. \**Une lumière facile*

Mais le point n'est pas là. Le point crucial est que les sens de *doux* et *tenue* en (6a-b) et (7a-b) – ou de *brillant* en (4a) et (5a) – ne sont déjà pas synonymes ! Nous sommes confrontés, en revanche, à une polysémie. *Doux* et *tenue* fonctionnent comme synonymes dans le cadre du schéma prédicatif sous (6) et dans le cadre du schéma prédicatif sous (7), séparément. La synonymie est, pour ainsi dire, contenue à l'intérieur de chaque schéma prédicatif. Le fait que *doux* et *tenue* soient interchangeables à travers (6) et (7) – sans être synonymes – est suffisant pour dissocier la synonymie de l'interchangeabilité, et pour montrer que tous les mots peuvent, en principe, être interchangeables dans tous les schémas prédicatifs, sans que cela n'affecte le fait qu'ils aient des sens virtuellement incommensurables<sup>6</sup>. Autrement dit, tous les mots peuvent être indéfiniment polysémiques et manifester la même polysémie, c'est-à-dire être synonymes dans les mêmes schémas prédicatifs. Cependant, *être synonymes dans les mêmes schémas prédicatifs* (= avoir la même polysémie) n'est pas à son tour une forme de synonymie (car cela suppose des sens différents pour chaque schéma prédicatif).

## 4.2. Synonymie et fondement du lexique

Revenons à la conception fonctionnelle de la signification présentée sous §3.1. Comme nous l'avons souligné, par rapport à un schéma prédicatif de référence, un sens est immédiatement identifié, et la polysémie est immédiatement levée. S'il est vrai que le verbe *abattre* est très polysémique, il est également vrai que – en connaissant les schémas prédicatifs <humain> *abattre* <bâtiment> et <humain> *abattre* <humain> – les sens de ce verbe ne peuvent plus être confondus. Il s'ensuit que, pour la compréhension de ces sens, l'existence d'un synonyme pour chaque emploi n'est pas requise. Le lexique peut bien offrir des synonymes

<sup>6</sup> Même si l'on suppose que *tenue* et *doux* remplissent une même fonction lexicale d'intensification (ou de-intensification) au sens de la théorie Sens-Texte (Magn), reste le fait que le contenu sémantique de cette fonction est très différent : l'intensité d'une pente est une inclination, mais non l'intensité d'une lumière. La difficulté à définir ce type de fonctions lexicales est qu'elles risquent d'être, elles-mêmes, polysémiques.

(de même qu'il peut offrir des hyperonymes), mais le codage de ces synonymes n'est pas nécessaire pour le fonctionnement du lexique (de même que le codage de tous les hyperonymes n'est nécessaire non plus).

Cette considération nous permet de revenir sur une affirmation de Spark Jones (1986 : 136) : « synonymy is fundamental to language ». De notre point de vue, si la synonymie est certainement une relation lexicale très productive, elle n'est pas pour autant plus fondamentale que les autres, car, comme toutes les autres relations lexicales, elle ne préexiste pas aux schémas prédicatifs, mais elle est fonction des schémas prédicatifs.

À bien y regarder, considérer la synonymie comme le fondement du lexique revient à commettre l'erreur consistant à essayer de définir le sens d'un mot (par exemple *prendre*) en donnant un autre mot (par exemple *saisir*). Comme Gross (2012 : 12) l'écrit, cela équivaut à sélectionner arbitrairement le mot qui vient le plus spontanément à l'esprit. En revanche, si nous fournissons des arguments, par exemple, *prendre un avion* (<moyen de transport public>), *prendre un chemin* (<route>), *prendre une aspirine* (<médicament>), les sens du mot se manifestent clairement et ils peuvent être illustrés par des éventuels synonymes.

Tout ensemble de synonymes (*row*, cf. Spark Jones, 1986 : 79 ; *synset*, cf. Milner, 1998 : 24) est un ensemble de mots. Si de tels ensembles peuvent être un moyen pratique pour organiser une base de données, ils ne peuvent pas être considérés comme les unités de base du lexique, car ils ne préexistent pas aux schémas prédicatifs, mais ils en sont des conséquences<sup>7</sup>. Dissocier les ensembles de synonymes de leurs schémas prédicatifs est donc un peu comme se retrouver quelque part sans ne plus savoir exactement ni où on est ni comment on est arrivé là. C'est pourquoi toute liste de synonymes sans schéma prédicatif associé (cf. Spark Jones, 1986 : 231–242) reste toujours plus ou moins vague (Gross, 1998 : 104, 2012 : 86–87).

En somme, espérer qu'une liste de synonymes puisse aider à identifier un sens précis revient à parier que cette liste n'identifie pas également des synonymes pour d'autres sens. Dans certaines limites, un tel pari peut être raisonnable, mais, comme on l'a vu au §4.1, en principe, cela n'est pas du tout nécessaire.

---

<sup>7</sup> WordNet est une base de données, un outil : il n'a donc pas l'engagement théorique de la théorie de Spark Jones (1986) qui l'inspire. Les *synsets*, en particulier, n'ont pas la prétention de définir ou identifier le sens d'un lexème, mais seulement de l'illustrer : chaque sens reste fixé par une définition traditionnelle, et la présence de *synsets* identiques n'est pas un problème réel. De ce point de vue, WordNet et la théorie des classes d'objets ne sont pas en conflit, mais complémentaires : remplacer les définitions traditionnelles du premier par les schémas prédicatifs de la seconde renforcerait, nous semble-t-il, la pertinence des *synsets*.

### 4.3. Synonymie : une relation seconde

Dans cette contribution, nous avons suggéré que la synonymie est une relation lexicale comme les autres, et cela en deux sens : d'une part, comme les autres, elle n'est pas partielle ; d'autre part, elle n'est pas plus fondamentale que les autres. Il y a cependant un sens où la synonymie est effectivement une relation spéciale. Revenons au cas de figure discuté sous §3.2. Affirmer que *démolir* est synonyme d'*abattre* par rapport au schéma prédicatif <humain> *abattre* <bâtiment> revient à affirmer que, par rapport à ce schéma, ils sont en compétition pour la même position de contraire de *construire*. Affirmer que *flingue* est synonyme de *pistolet* par rapport à la classe d'objets des <armes à feu>, revient à affirmer que, par rapport à cette classe, ils sont en compétition pour la même position d'hyponyme du champ sémantique des armes. Etc.

Mais si cela est vrai, alors ce n'est pas incorrect d'affirmer que la synonymie est une relation « seconde » qui suppose les autres. Autrement dit, les synonymes sont des signifiés différents fonctionnant comme le *même hyponyme* dans un champ sémantique, comme le *même contraire* dans un champ d'opposition, etc. Les synonymes supposent donc la co-hypomimie, l'opposition, etc.

### 4.4. Nuances synonymiques

Nous concluons cet article en revenant sur l'intuition qu'il est très – trop – facile d'envisager des nuances différentes entre les synonymes (cf. §1.1). Cette intuition est généralement justifiée sur la base d'un principe d'économie du langage : cf. Neveu (2004 : 281) et Murphy (2010 : 112) « it takes more effort to learn two words when one is sufficient ». L'idée de la synonymie comme présupposé, exprimée à travers la notion de schéma prédicatif, offre une justification alternative, mais naturelle, de cette intuition.

La synonymie, comme nous l'avons anticipé au §2.1, est un phénomène à deux faces. L'identité du schéma prédicatif en est le présupposé. Les différentes nuances sémantiques en sont le posé ou la face fonctionnelle. Si cela est clair, alors il est également clair nous sommes libres de chercher des différences entre les synonymes précisément parce que l'identité du schéma prédicatif n'est pas remise en cause. Autrement dit, pour justifier l'attitude naturelle évoquée ci-dessus, il n'y a pas besoin d'invoquer un principe d'économie, car c'est la synonymie elle-même – conçue comme identité de schéma prédicatif – qui nous invite à différencier les synonymes avec toute sorte de nuance

sémantique<sup>8</sup>. Ces différences ne se situent pas au niveau des valeurs oppositives qui séparent les items d'un champ sémantique. Au contraire, la présupposition que les synonymes sont le même item d'un champ sémantique est ce qui nous pousse à multiplier – et même inventer – les différentes nuances synonymiques.

## Références citées

- Apresjan, J. (2000). *Systematic Lexicography*. Oxford University Press.
- Baldinger, K. (1984). *Vers une sémantique moderne*. Klincksieck.
- Carnap, R. (1952). Meaning postulates. *Philosophical Studies* 3, 65–73.
- Cruse, D. A. (1986)(1997). *Lexical semantics*. Cambridge University Press.
- Doualan, G. (2013). La synonymie, relation d'équivalence, un artefact de la pensée ?. *Équivalences* 40(1&2), 15–42.
- Doualan, G. (2014). Éléments pour une lecture de l'histoire de la synonymie. *4<sup>e</sup> Congrès Mondial de Linguistique Française, SHS Web of Conferences* 8, 409–424.
- Fasciolo, M. & Gross, G. (2021). *La sintassi del lessico. Manuale di linguistica aperta all'informatica e alla filosofia*. UTET.
- Girard, G. (abbé) (1718). *La Justesse de la langue françoise ou Les différentes significations des mots qui passent pour synonymes*. Veuve d'Houry.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago University Press.
- Gross, G. (1994). Classes d'objets et description des verbes. *Langages* 28(115), 15–30.
- Gross, G. (1998). Pour une véritable fonction « synonymie » dans un traitement de texte. *Langages* 32(131), 103–104.
- Gross, G. (2008). Les classes d'objets. *Lalies* 28, 111–165.
- Gross, G. (2010). Sur la notion de contexte. *Meta* 1(55), 187–198.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses Universitaires du Septentrion.
- Gross, G. & Clas, A. (1997). Synonymie, polysémie et classes d'objets. *Meta* 42(1), 147–154.
- Honeste, M.-L. (2007). Entre ressemblance et différence : synonymie et cognition. *Le Français Moderne* 75(1), 160–173.
- Kleiber, G. (2009). La synonymie « identité de sens » n'est pas un mythe. *Pratiques* 141/142, 9–25.

---

<sup>8</sup> Cela n'empêche que, parfois, ces nuances peuvent s'avérer discursivement plus pertinentes du schéma prédicatif partagé. C'est le cas de la différence de domaine qui peut entraîner une acceptabilité différente de *vélo de course* et (?)*bicyclette de course* (cf. Kleiber, 2009 : 19–20).

- Ježek, E., (2016). *The lexicon. An Introduction*. Oxford University Press.
- Leclercq, B. & Morin, C. (2023). No equivalence: A new principle of no synonymy. *Constructions* 15, 1–16.
- Lehmann, A. & Martin-Berthet, F. (1998). *Introduction à la lexicologie. Sémantique et morphologie*. Dunod.
- Le Pesant, D. & Mathieu-Colas, M. (1998). Introduction aux classes d'objets. *Langages* 32(131), 6–33.
- Lyons, J. (1977). *Semantics I*. Cambridge University Press.
- Milner, Ch. (1998). Nouns in WordNet. Dans Ch. Fellbaum (éds), *WordNet. An Electronic Lexical Database* (23–46). MIT Press.
- Mortureux, M.-F. (1997). *La lexicologie entre langue et discours*. SEDES.
- Murphy, M. L. (2010). *Lexical Meaning*. Cambridge University Press.
- Neveu, F. (2004). *Dictionnaire des sciences du langage*. Armand Colin.
- Niklas-Salminen, A. (1997) (2003). *La lexicologie*. Armand Colin.
- Ploux, S. & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement Automatique des Langues* 39(1), 161–182.
- Rey-Debove, J. (1997). La synonymie ou les échanges de signes comme fondement de la sémantique. *Langages* 128, 91–104.
- Spark Jones, K. (1986). *Synonymy ans semantic classification*. Edinburgh University Press.
- Tamba, I. (1988) (2005). *La sémantique*. PUF.
- Uhrig, P. (2015). Why the principle of no synonymy is overrated. *Zeitschrift Für Anglistik Und Amerikanistik* 63(3), 323–337.
- Venant, F. & Victorri, B. (2012). La synonymie comme accès à la structure sémantique du lexique adjectif et verbal du français. Dans F. Berlan & G. Berthomieu (éds), *La synonymie* (73–90). PUPS.
- Zhurauliova, H. (2021). La synonymie exacte existe-t-elle?. *Le français moderne* 89(2), 267–280.



*Aude Grezka*

CNRS, LIPN

Université Sorbonne Paris Nord

France

 <https://orcid.org/0000-0002-4582-3428>

# Morfetik – mises à jour et évolutions d'une ressource en ligne

**Morfetik – updates and upgrades to an online resource**

## Abstract

In this article, a morphological linguistic resource for contemporary French called Morfetik is presented. The evolution of the resource and its various linguistic and technological characteristics are discussed. Additionally, an overview of the numerous tools integrated into the resource is provided. Morfetik represents a continuously evolving platform designed to progress and enhance the processing of textual data.

## Keywords

Morphological linguistic resource, Morfetik, NLP, language variation, French

## 1. Introduction

La création d'une ressource lexicale ne se résume pas à un listage d'éléments lexicaux mais dans une définition rigoureuse et actualisable du lexique et de ses diverses flexions. Ces ressources morphologiques sont utilisées dans divers domaines de la linguistique et de la recherche en Traitement automatique du langage (TAL). Elles sont essentielles pour l'analyse et la génération de mots, la lemmatisation, l'étude des relations entre les mots, la modélisation linguistique et bien d'autres applications linguistiques.

Dans cette perspective, est né le projet Morfetik. Ce projet a débuté en 2008<sup>1</sup> avec l'informatisation de la ressource lexicale de M. Mathieu-Colas (M. Mathieu-Colas, 2009 ; P.-A. Buvet *et al.*, 2009) et de sa structuration, travail de plus d'une vingtaine d'années de collecte et de description. En 2015, A. Grezka a repris la direction du projet et a repensé la ressource avec les technologies actuelles<sup>2</sup>. La base de données a également été actualisée et enrichie considérablement, notamment avec l'ajout des unités polylexicales, des verbes pronominaux et des néologismes, totalement absents auparavant (A. Grezka, 2017 ; 2020). Morfetik est une ressource en ligne en accès libre<sup>3</sup> qui présente les caractéristiques suivantes : une large couverture, des informations précises et fiables, le respect des normes et une évolutivité garantie. La ressource lexicale Morfetik est un dictionnaire morphologique des unités lexicales simples et polylexicales du français contemporains (noms, adjectifs, déterminants, pronoms, verbes, adverbes, prépositions, conjonctions, interjections, locutions, etc.). Il est ainsi possible d'obtenir toutes les formes associées à n'importe quel mot français, que ce soit le pluriel des noms, le féminin et le pluriel des adjectifs, ou encore les formes conjuguées des verbes, et bien d'autres. De même, Morfetik permet d'identifier le mot de base, appelé « lemme », correspondant à n'importe quelle forme fléchie.

Les données sont organisées en tables et constituent le point de départ d'un système de traitement qui comprend un moteur de flexion, un dictionnaire des formes fléchies, des interfaces de consultation et d'interrogation, ainsi qu'une interface d'édition réservée aux éditeurs pour mettre à jour la base de données (assurer la maintenance et l'exploitation de la ressource). Grâce à ce système, il est possible de générer automatiquement toutes les formes simples et complexes de la langue française et d'apporter également des informations sémantiques lorsque cela est nécessaire (domaines, par exemple), de contexte (analyse et suivi du mot dans la presse), de fréquence, et bien d'autres éléments. Depuis quelques années Morfetik s'est imposé comme un outil linguistique essentiel, offrant des analyses morphologiques précises et des informations linguistiques approfondies.

L'article présente la ressource par le biais de ses évaluations linguistiques et informatiques. Nous proposons tout d'abord de revenir brièvement sur le fonctionnement général de Morfetik. Puis, nous présentons les différentes nouveautés. Nous verrons notamment les outils qui ont été ajoutés à la ressource. Mor-

---

<sup>1</sup> Le projet a débuté au laboratoire LDI (Lexiques, Dictionnaires, Informatique, UMR 7187 CNRS), anciennement LLI (Laboratoire de Linguistique Informatique) dirigé par Gaston Gross.

<sup>2</sup> Ce travail bénéficie partiellement d'une aide de l'IdEx Université Paris Cité (ANR-18-IDEX-0001) au titre du Labex Empirical Foundations of Linguistics – EFL.

<sup>3</sup> Lien vers la ressource (version 0.2(2022)) : <https://tal.lipn.univ-paris13.fr/morfetik>.

fetik a pour vocation de s’enrichir progressivement pour améliorer la chaîne de traitement des données textuelles.

## 2. Fonctionnement de Morfetik

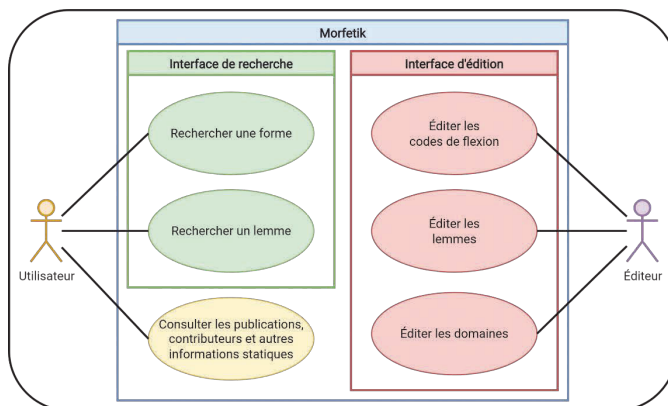
La plateforme Morfetik est découpée en trois grands blocs : l’interface de recherche de la base de données, la fonction principale de Morfetik ; des pages d’informations (page d’accueil, publications, etc.) et une interface d’édition accessible uniquement par les éditeurs pour permettre la mise à jour de la base de données. Dans cette partie, nous détaillons l’architecture générale du système et ses principaux modules. Du point de vue technique, l’un des plus gros travail que nous avons dû réaliser, a été la reconstruction du moteur de flexion et la création d’une interface éditeur pour faciliter le travail collaboratif.

### 2.1. Cas d’utilisation

Morfetik à deux cas principaux d’utilisation : la recherche par mot (lemme) et la recherche par forme. À ces cas s’ajoute la gestion des données, avec une interface d’édition pour que les éditeurs de la ressource puissent ajouter des mots ou des informations plus facilement :

**Figure 1**

*Diagramme des cas d’utilisation de Morfetik (v.0.2)*



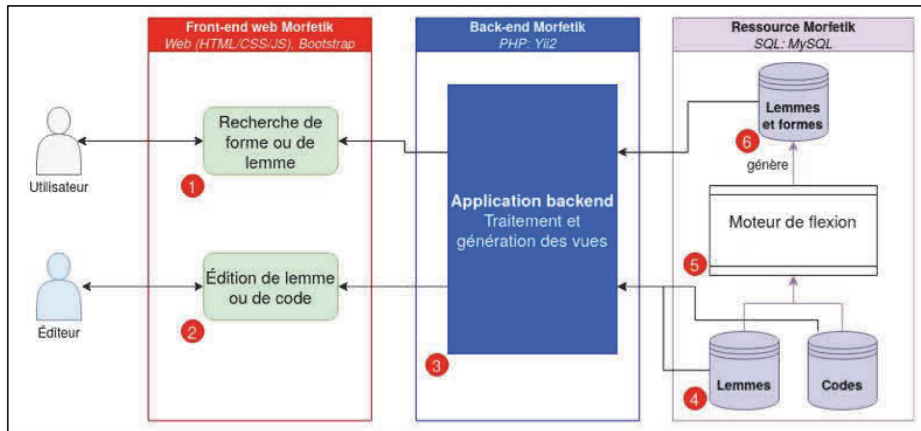


## 2.2. Architecture de Morfetik

L'architecture de Morfetik est la suivante :

**Figure 2**

*Architecture de Morfetik (v.0.2)*



1. Les utilisateurs peuvent rechercher une forme ou un lemme. [SEP]
2. Les éditeurs peuvent modifier les lemmes ou les codes. [SEP]
3. Dans les deux cas (utilisateur et éditeur), le back-end<sup>4</sup> va s'occuper de récupérer les informations. [SEP]
4. Dans le cas d'une demande d'édition, le moteur va permettre de modifier les lemmes et les codes de la ressource Morfetik.
5. Du côté de la ressource, le moteur de flexion, écrit en SQL, s'occupe de générer la table des formes et des lemmes. [SEP] Le moteur effectue la flexion des lemmes : il permet de passer d'un lemme et d'un code à des formes.
6. Cette table est utilisée pour les requêtes de formes et de lemme. L'application [SEP] back-end effectue de la logique simple pour récupérer les résultats.

Le choix de cette architecture, d'inclure le moteur de flexion non pas dans l'application back-end mais dans la base de données elle-même, présente un avantage. La valeur de Morfetik réside dans les données linguistiques et non dans ses interfaces d'utilisation. Le moteur de flexion est primordial au fonctionnement de Morfetik. Ainsi, si le front-end<sup>5</sup> ou encore le framework<sup>6</sup> utilisés viennent

<sup>4</sup> La partie invisible de l'application web (côté serveur).

<sup>5</sup> La partie visible de l'application web (côté client).

<sup>6</sup> Ensemble d'outils et de composants logiciels à la base d'une application.

à changer, l'accès aux données, le plus important, restera toujours présent, rendant ainsi Morfetik modulaire. L'objectif dans la conception de Morfetik est de créer une application qui puisse durer dans le temps. Les frameworks, le back-end ou le front-end viendront peut-être à changer mais la base ne changera pas.

### 2.3. Reconstruction et mise en fonctionnement du moteur de flexion

L'un des problèmes majeurs et techniques que nous avons dû régler à la reprise du projet fut la reconstruction du moteur de flexion. Depuis de nombreuses années, Morfetik était statique puisque le générateur automatique de formes fléchies ne fonctionnait plus. Le système permet maintenant de générer automatiquement l'ensemble des formes simples et composées du français et donc de mettre à jour les données.

Les formes sont générées automatiquement à partir d'une table des lemmes et d'une table de modèles morphologiques représentés par des codes de flexion (pour les mots invariables, comme les adverbes, un simple listage suffit). Les codes (environ 400 actuellement) représentent les formes que vont prendre chaque lemme dans des situations différentes : temps, masculin ou singulier, pluriel des noms<sup>7</sup>, féminin/pluriel des adjectifs, etc. Le générateur de formes fléchies est la composante principale du système qui prend en charge l'interprétation des tables pour générer le lexique des formes fléchies (c'est-à-dire le résultat de la recherche). À chaque forme sont associés un lemme ainsi que différentes informations de catégorie, de genre et de nombre (noms, adjectifs) ou de temps, de mode, de personne et de genre (verbes). Il comprend un moteur principal ainsi que des modules. Le moteur principal se charge de la collecte des informations

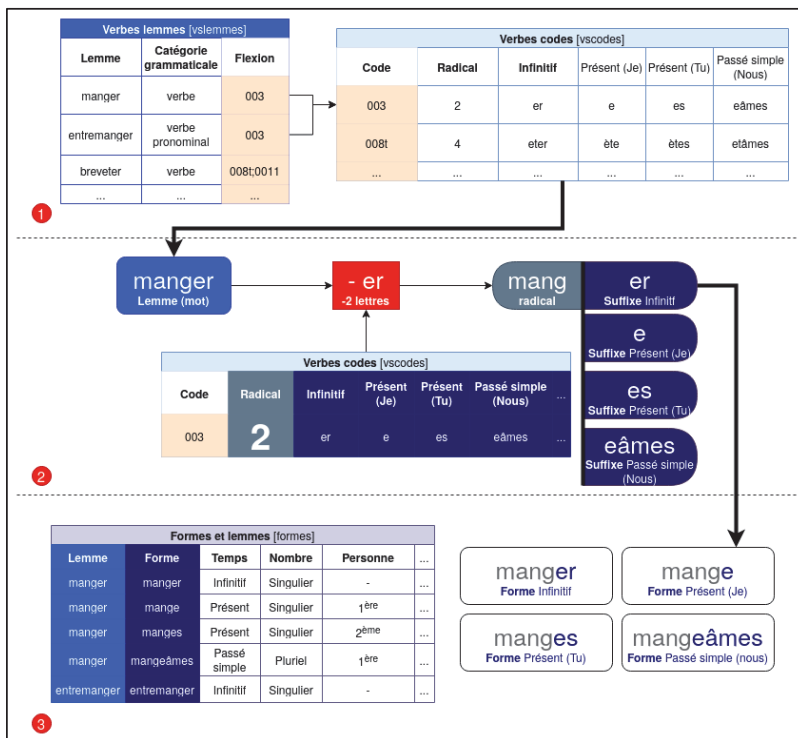
---

<sup>7</sup> Pour des raisons à la fois théoriques et pratiques, les noms n'ont pas été fléchis en genre, à la différence d'autres dictionnaires, il n'y a donc dans la nomenclature que des noms masculins et des noms féminins (M. Mathieu-Colas, 2009). Du point de vue théorique, le genre n'a pas le même statut pour les noms et les adjectifs. Il est inhérent au nom, d'où la variété des formes : à côté de noms épiciques (*un élève, une élève*), toutes sortes de substitutions ou de suffixes peuvent apparaître (*un loup, une louve*). Il y a même une rupture morphologique entre les deux genres (*un dieu, une déesse ; un roi, une reine ; un garçon, une fille*). Du point de vue pratique, le traitement du féminin en termes de flexion pose de nombreuses difficultés, dues à la polysémie. Si l'on veut regrouper en une seule entrée *maître* et *maîtresse*, il faut préciser que cette correspondance, usuelle dans le domaine scolaire (« le maître / la maîtresse a dit... »), ne vaut pas pour tous les emplois : ainsi, il est des cas où *maître* ne peut être qu'au masculin (« les grands maîtres de la peinture » vs « \*les grandes maîtresses de la peinture »), ou à l'inverse au féminin (« il a rejoint sa maîtresse » vs « \*elle a rejoint son maître »). Même des formes apparemment simples peuvent être problématiques : par exemple, si *boulangier* désigne toujours quelqu'un qui fabrique et qui, le cas échéant, vend du pain, *boulangère* peut aussi signifier « femme d'un boulangier » (Larousse, TLFi).

dans les tables, de la création des formes fléchies à partir du lemme et du code de flexion approprié, de l'attribution des catégories grammaticales correspondantes (genre, nombre, etc.) et des autres métadonnées, et enfin il centralise l'exécution des modules. Ces modules sont de plusieurs ordres. Chaque partie du discours nécessitant une flexion dispose d'un module spécifique permettant de gérer finement les particularités des tables qui lui correspondent, comme les différents types de codes pour la rareté, ainsi que les différents moyens de signaler les formes défectives ou les variantes. Cette architecture modulaire permet de pouvoir assez facilement envisager la réutilisation du moteur pour d'autres langues ou d'autres ressources.

L'intérêt principal de ce système, est qu'au lieu de rentrer toutes les flexions, toutes les formes d'un mot, d'un verbe ou d'une expression, les éditeurs n'ont besoin de rentrer que le lemme dans la base de données en lui associant un ou des codes de flexion qui vont ainsi permettre de créer automatiquement la forme fléchie. Le schéma suivant permet de comprendre le moteur de flexion :

**Figure 3**  
*Fonctionnement du moteur de flexion pour un verbe*



Le processus de flexion se fait en trois étapes :

1. Chaque lemme possède une catégorie grammaticale et une flexion, correspondant à un code. Un lemme peut posséder plusieurs flexions. Le moteur de flexion identifie tous les codes correspondants au lemme.
2. Chaque code possède un champ « Radical », indiquant le nombre de lettres à enlever pour avoir le radical du mot. Le moteur de flexion associe ensuite le radical à chaque suffixe.
3. Des formes fléchies sont créées. Le moteur de flexion joint les résultats dans une table, avec des informations syntaxiques, comme le temps et le nombre si le lemme est un verbe, ou le genre si le lemme est un nom.

Prenons l’exemple d’un mot simple qui n’a que 2 formes fléchies : *sandwich*. Au pluriel, ce nom masculin peut prendre la forme *sandwichs* ou la forme *sandwiches*. Pour que le moteur de flexion puisse créer toutes les formes de ce mot, il a été codé avec les codes 01 (qui ajoute « s » au pluriel) et avec le code 25 (qui ajoute « es » au pluriel). Le moteur de flexion va donc créer les 2 formes avec le code 01 (un *sandwich*, des *sandwichs*) puis les 2 formes avec le code 25 (un *sandwich*, des *sandwiches*). Pour éviter d’avoir des doublons dans la base de données, lorsqu’une forme existe déjà dans la base, on ne la crée pas. Ainsi dans cet exemple, le moteur de flexion ne crée que 3 formes (un *sandwich*, des *sandwichs/sandwiches*) puisque le singulier est identique malgré 2 codes différents. La majorité des codes ont également pour fonction de retirer une partie du lemme pour obtenir le radical avant d’ajouter la terminaison (par exemple le code 03 des noms retire la dernière lettre et ajoute « -ux » au pluriel, on peut le voir sur le mot *idéal* → *idéaux*).

Chaque code est associé à plusieurs lemmes, et un lemme est associé à un seul code.

## 2.4. Création d’une interface éditeur

Dans les premières versions de Morfetik, les éditeurs modifiaient la base de données directement. Un de nos objectifs dans ce projet a été que la ressource soit facile d’utilisation pour les utilisateurs comme pour les éditeurs. Nous avons donc créé une interface CRUD (Create Read Update Delete : interface permettant d’interagir avec des données, on peut consulter, créer, mettre à jour et supprimer des données) pour interagir directement avec Morfetik :

**Figure 4**

Édition directe de la table avec menu déroulant (image prise sur le site Morfetik)

Verbes: lemmes Affichage de 1-20 sur 10,230 éléments.

\* Vous pouvez redimensionner la table en prenant les bordures des colonnes.

#	+	Lemme	Catégorie grammaticale	Sous-catégorie grammaticale	Flex	Notes	Pronominal	Actes
		<input type="text"/>	vr <b>b</b>	<input type="text"/>				
1	+	abaisser	vr <b>b</b>	(non défini)			2	
2	+	abalourdir	vr <b>b</b>	(non défini)				
3	+	abandonner	vr <b>b</b>	(non défini)				
4	+	abasourdir	vr <b>b</b>	(non défini)				
5	+	abâtardir	vr <b>b</b>	(non défini)				
6	+	abattre	vr <b>b</b>	(non défini)	120	(non défini)	2	

Modifier Sous-catégorie grammaticale

- (Non rempli)
- vi (Verbe intransitif)
- vt (Verbe transitif)
- vt (vpr) (Verbe transitif (verbe pronominal))
- loc v (Locution verbale)
- (Non rempli)

L'interface d'édition permet maintenant d'éditer la base Morfetik et sa structuration, de rechercher, modifier ou bien ajouter de nouveaux mots. L'interface d'édition possède également des menus déroulants qui sont modifiables directement. La grille d'édition permet l'édition de fiche, la suppression d'un lemme ou la suppression de masse. La recherche est directement intégrée à la grille. On peut également exporter en format TXT ou CSV, pour permettre le traitement par Excel.

L'interface éditeur permet ainsi aux utilisateurs de modifier ou de personnaliser la ressource de manière conviviale et intuitive, sans avoir à manipuler directement le code ou les paramètres techniques. Cela ouvre la porte à un plus large éventail d'utilisateurs, même ceux qui ne possèdent pas de connaissances approfondies en programmation. Enfin, l'utilisation de l'interface éditeur permet de réduire les erreurs potentielles. Les utilisateurs peuvent effectuer des modifications dans un environnement contrôlé et guidé, ce qui diminue les risques d'introduire des erreurs de syntaxe ou d'autres problèmes liés à la manipulation directe du code.

### 3. Caractéristiques linguistiques de la ressource

Nous présentons ici les caractéristiques linguistiques de la ressource, ainsi que les nouveautés (veille linguistique, implémentation des unités polylexicales, des domaines et des verbes pronominaux).

### 3.1. Large couverture et fiabilité des données lexicographiques

Morfetik est actuellement la ressource morphologique du français la plus exhaustive. La ressource a en effet été comparée avec des ressources lexicales analogues en français<sup>8</sup> (A. Grezka *et al.*, 2015). La couverture lexicale a également été validée par comparaison avec trois corpus du français suffisamment volumineux et représentatifs de la langue générale (les 10 ans du Monde, le Wikipédia français et la version française de Wacky). La ressource contient plus de 1.000.000 de formes.

L’une des spécificités de Morfetik réside dans la nature des sources. Plutôt que de faire table rase de l’héritage lexicographique en s’en remettant exclusivement à des procédures d’extraction automatique, il nous a semblé préférable de nous appuyer, dans un premier temps, sur la richesse des données consignées dans les dictionnaires, qui ont le double mérite d’exister et d’être particulièrement fiables. Toutes les variantes graphiques attestées dans les ouvrages ont été retenues. En effet, l’inclusion des variantes graphiques (A. Grezka, 2020) dans la base de données vise à fournir une ressource complète, précise et respectueuse de la diversité linguistique, tout en répondant aux besoins des différents utilisateurs de la langue, qu’ils soient novices ou experts, en leur donnant une vue complète des possibilités orthographiques. Les principales sources consultées ont été les suivantes :

- le *Petit et le Grand Robert*, le *Petit Larousse illustré* et le *Lexis* ;
- le *Trésor de la langue française* (avec toutefois un filtrage des entrées, pour exclure certaines données trop datées ;
- le *Harrap’s* et le *Robert & Collins* ;
- des dictionnaires d’argot (le *Dictionnaire de l’argot de Larousse*) ;
- des tables de conjugaison (dont le *Bescherelle*) ;
- *Le Bon Usage* de GREVISSE et des dictionnaires de « difficultés » pour le traitement des cas problématiques.

Nous avons également pris en compte le DELAS (*Dictionnaire électronique du LADL*, cf. B. Courtois, 1990), avec, ici encore, un filtrage des données, afin d’éliminer certains artefacts.

Afin d’amorcer l’intégration des langues de spécialité au sein de la base, nous avons mis à contribution, dans une large mesure, le *Grand Dictionnaire encyclopédique Larousse* (GDEL), et exploité systématiquement, pour deux domaines particuliers – la médecine et la minéralogie –, des dictionnaires spécialisés.

---

<sup>8</sup> Notamment avec les ressources lexicales : GLAFF (Hathout *et al.*, 2014 ; Sajous *et al.*, 2013, 2014), Leff (Clément *et al.*, 2004 ; Sagot, 2010), Morphalou et Dicolecte.

Ce noyau historique de Morfetik étant constitué, nous avons pu, dans un deuxième temps, entreprendre l'actualisation et l'élargissement des données, sur plusieurs plans simultanés :

- l'intégration des néologismes, des nouvelles graphies (cf. les Rectifications de l'orthographe 1990) et des nouveaux féminins ;
- l'exploration d'autres spécialités ;
- l'exploitation des grands corpus.

### 3.2. Mises à jour régulières : veille linguistique

Cet inventaire n'est pas définitif et continue à évoluer. Il était en effet important pour nous d'effectuer une veille linguistique, qui ne se faisait plus sur Morfetik. Les dictionnaires, comme toutes les ressources linguistiques, nécessitent une confrontation continue avec des corpus, pour surveiller, collecter des informations sur les évolutions, les tendances et les nouveautés dans le domaine de la langue. Cette veille est essentielle pour rester à jour dans un monde linguistique en constante évolution.

Ainsi, nous actualisons quotidiennement la base de données par l'ajout de néologismes et l'intégration de nouvelles spécialités, rendu possible par les programmes liés à Morfetik, grâce à la connexion que nous avons établie avec les ressources Néoveille (<https://tal.lipn.univ-paris13.fr/neoveille>) et France Terme (<https://www.culture.fr/franceterme>)<sup>9</sup> :

**Figure 5**

*Interface de recherche de Morfetik (image prise sur le site Morfetik)*

The screenshot shows the Morfetik search interface. At the top, there is a navigation bar with links: Rechercher, Edition, Admin, Publications, A propos, Contact, Déconnexion (admin), and Admin. Below this, the search bar contains the text 'avions' and a 'Rechercher' button. There are also links for 'Guide d'utilisation' and 'Recherche avancée'. A checkbox labeled 'Sensible aux accents' is present. Below the search bar, the results are displayed in a table format, showing 2 elements out of 1-2.

	Lemme	Catégorie	Sous-catégorie	Temps	Nombre	Genre	Personne	Notes	Ressources externes
+	avion	Nom (n)	Nom masculin (nm)		P	M			
+	avoir	Verbe (vrb)		Ind-imp	1		P		

<sup>9</sup> Cette partie du projet a obtenu des subventions de la DGLFLF entre 2015 et 2019. La DGLFLF élabore la politique linguistique du Gouvernement en liaison avec les autres départements ministériels.

La plateforme Néoveille (E. Cartier, 2011, 2019 ; E. Cartier & J.-F. Sablayrolles, 2009) vise à détecter automatiquement, décrire linguistiquement et suivre l’évolution des innovations lexicales en corpus dynamique. Sa connexion avec Morfetik permet une mise en parallèle permanente avec des corpus afin de suivre l’évolution fréquentielle des lexies et de repérer ainsi celles qui semblent s’implanter (c’est-à-dire des néologismes qui atteignent une fréquence suffisante, sur une période donnée) pour les intégrer à Morfetik.

La ressource FranceTerme, quant à elle, est une base de données terminologiques de la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF) du ministère de la Culture, qui rassemble les récents néologismes spécialisés et avalisés par la Commission d’enrichissement de la langue française et publiés au Journal officiel, remplaçant les termes importés d’autres langues. La base terminologique comporte plus de 8.500 termes français dans différents domaines scientifiques et techniques. Elle a pour mission de promouvoir l’utilisation de mots en français, pour enrichir la langue française et éviter ainsi son recul dans le monde. Ainsi, pour désigner les réalités nouvelles qui se créent constamment, des termes français sont recommandés par des spécialistes. Par exemple, le domaine de l’informatique est notoirement connu pour user d’anglicismes, alors même que des équivalents français existent parfois. Ces termes font donc leur entrée dans Morfetik : *plateforme de données* (pour *data hub*), *apprentissage profond* (pour *deep learning*), *banque de données* (pour *data bank*), etc.

Ces ressources nous permettent donc de compléter les lacunes et d’enrichir la terminologie qui ne cesse d’évoluer, en particulier dans les langues de spécialités (nouvelles informations, nouveaux concepts ou nouvelles découvertes).

### 3.3. Implémentation des unités polylexicales

Aucun traitement informatisé n’est concevable si les unités complexes ne sont pas clairement identifiées, ce qui suppose un recensement très large et une description minutieuse. Nous avons, en 2022, implémenté dans Morfetik les unités polylexicales, quel que soit leur degré de figement (mots composés, mots à trait d’union, locutions, etc.)<sup>10</sup> :

---

<sup>10</sup> Nous ne faisons pas de différence, ici, entre les diverses appellations des séquences figées : noms ou adjectifs composés, locutions nominales ou adjectivales, noms ou adjectifs polylexicaux, etc. On conviendra de parler d’« unités polylexicales » dans un sens indifférencié pour désigner tous les syntagmes qui comportent, à quelque titre que ce soit, une part de figement.



**Figure 6**

Recherche d'une expression (image prise sur le site Morfetik)

The screenshot shows a search interface with a search bar containing 'copain comme cochon' and a 'Rechercher' button. Below the search bar, there is a checkbox for 'Sensible aux accents' and a 'Recherche avancée' link. The results section shows 'Affichage de 1-1 sur 1 élément.' and a table with columns: Lemme, Catégorie, Sous-catégorie, Temps, Nombre, Genre, Personne, Notes, and Ressources externes. The main entry is 'copain comme cochon' categorized as 'Adjectif Locution'. Below this, there is a detailed table for the lemma 'copain comme cochon' with columns: Lemme, Notes, Variante, and Informations sémantiques. The table lists four forms: Masculin singulier, Masculin pluriel, Féminin singulier, and Féminin pluriel, all with the same lemma and notes.

	copain comme cochon	Lemme	Notes	Variante	Informations sémantiques
Masculin singulier	copain comme cochon	copain comme cochon	Source : FRlex-adjc.csv		
Masculin pluriel	copains comme cochons				
Féminin singulier	copine comme cochon				
Féminin pluriel	copines comme cochons				

Sur le même principe que pour les mots simples, un dictionnaire des unités complexes a été construit (M. Mathieu-Colas, 2014 ; G. Gross, 1996). La tâche n'était pas simple, compte tenu de la diversité des formes (plus de 700 types morphologiques pour les seuls noms composés). Il fallait compter aussi avec la complexité de certains schémas flexionnels, notamment pour les mots à trait d'union, qui comportent de nombreuses exceptions. Prenons l'exemple de l'adjectif composé *franc-comtois* : on trouvera au masculin pluriel *FRANCS-comtois* (des *fromages FRANCS-comtois*) mais au féminin pluriel *FRANC-comtoises* (des *horloges FRANC-comtoises*). L'autre difficulté rencontrée lors de l'inventaire, est que l'orthographe des mots composés est loin d'être bien définie dans les dictionnaires traditionnels : un nombre non négligeable d'entrées sont écrites soudées dans certains dictionnaires, avec un trait d'union dans d'autres ; ou parfois l'information de flexion n'apparaît pas.

Les difficultés de l'orthographe française ne tiennent pas seulement à la complexité des règles traditionnelles et à la multitude des exceptions, elles résultent également, de manière plus insidieuse, de l'état d'« incertitude » qui caractérise une partie du lexique. La norme n'est pas si bien fixée qu'on se plaît à le croire : un certain nombre de variantes existent, qui désignent indirectement les failles de réécriture et invitent par là même à s'interroger sur la logique du système et les fondements de l'usage.

(M. Mathieu-Colas, 1990 : 104)

Nous avons donc adopté un système suffisamment flexible pour pouvoir prendre en charge les différents cas de figure.

Pour les noms, le codage morphologique des unités polylexicales repose sur les mêmes principes fondamentaux que celui des noms simples, mais la complexité des formes implique quelques ajustements. La difficulté résulte principalement dans le fait que la flexion peut être terminales et/ou internes, multiples ou être absente :

- absence de flexion : le mot n'a qu'une forme, par exemple le nom *beaux-parents* (nms) n'existe qu'au pluriel ; le *bien-être*, uniquement au singulier ;
- flexion terminale : le mot peut se rencontrer au singulier ou au pluriel, mais la flexion fonctionne comme celle des mots simples : une *demi-heure*, des *demi-heureS* ;
- flexion interne : le nom peut avoir une ou plusieurs flexions internes : un *arrière-grand-père*, des *arrière-grandS-pèreS* ; une *assiette à soupe*, des *assietteS à soupe*. La difficulté réside dans la gestion des variantes flexionnelles puisqu'un même élément peut donner lieu à deux flexions distinctes : *match aller matchS* ou *matchES aller* ou alors un même élément peut être fléchi ou non fléchi : un *haricot vert sans fil*, des *haricotS vertS sans fil(S)* ;
- flexion interne et terminale : une même unité peut relever simultanément de deux types de flexion : un *social-démocrate*, des *social-démocrateS*, des *sociAUX-démocrateS*.

Le codage des adjectifs composés est encore plus complexe que celui des noms, puisqu'à la flexion en genre vient s'ajouter la mise au pluriel. Par ailleurs, certaines unités intègrent des déterminants ou des pronoms qui développent leur propre paradigme (*content de SOI, de MOI, de TOI, de LUI*, etc.). On retrouve les trois types de flexion précédemment identifiés : absence de flexion, flexion terminale et flexion interne :

- absence de flexion : les pluriels obligatoires ne donnent pas lieu à un code flexionnel. Cela ne concerne que quelques mots : *bons amis, comme frère et sœur, comme larrons en foire, en boules de loto, mari et femme, sonnantes et trébuchantes* ;
- flexion terminale : de nombreux adjectifs composés se fléchissent de la même manière que des mots simples : invariabilité (féminin et pluriel identiques au masculin singulier : à *but lucratif*) ; flexion terminale (à *moitié vide, -es ; semi-public, -cs, -que, -ques*) ; variantes (*anti-grève, pl. anti-grève ou anti-grèveS ; semi-nasal, pl. semi-nasaLS ou semi-nasaUX*) ;
- flexion interne : la flexion interne peut concerner n'importe quel élément : *ÂPRE au gain* (flexion sur le 1<sup>er</sup> élément) ; *plus MORT que VIF* (flexion sur le 2<sup>e</sup> et 3<sup>e</sup> éléments) ; *NOURRI, LOGÉ, BLANCHI NÉCESSAIRE* (flexion sur tous les éléments), etc. Par ailleurs, plusieurs cas de flexion peuvent se présenter :

- un même élément peut donner lieu à deux flexions distinctes : *frais émoulu* ➤ *fém. FRAIS émoulue ou FRAÎCHE émoulue*,
- un même élément peut être fléchi ou non fléchi : *taillé en pointe* ➤ *fém. sing. tailléE en pointe, masc. plur. tailléS en pointe ou en pointeS, fém. plur. tailléES en pointe ou en pointeS* ;
- flexion interne et terminale : une même unité peut cumuler les deux types de flexion : *en épingle à cheveux* ➤ *pl. en ÉPINGLE à cheveux ou en ÉPINGLES à cheveux*.

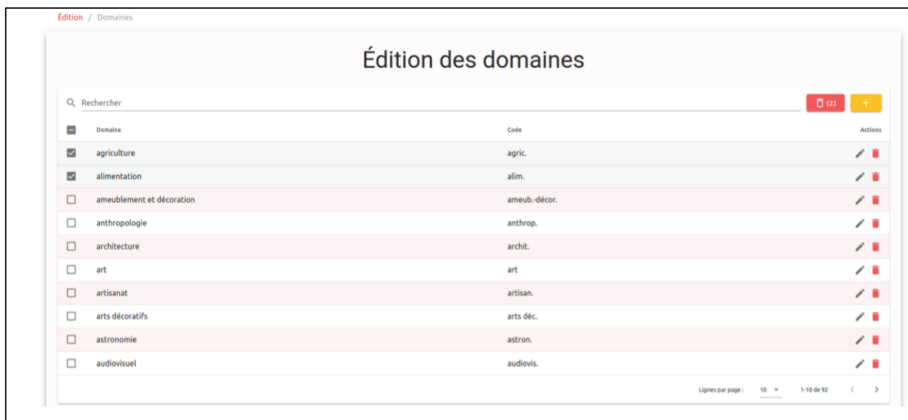
### 3.4. Implémentation des domaines et des verbes pronominaux

Parmi les dernières fonctionnalités que nous avons ajoutées à Morfetik : l'implémentation des domaines et des formes pronominales.

À partir de l'interface éditeur, il est maintenant possible d'intégrer à la description d'une entrée son domaine, à l'aide d'une liste proposée :

**Figure 7**

*Édition des domaines (image prise sur le site Morfetik)*



Nous ne codons les indications de domaine que pour les termes scientifiques ou techniques. Il ne s'agit pas ici d'établir un système conceptuel hiérarchisé, comme celui qu'utilisent les bibliothèques pour le catalogage (par exemple la classification Dewey). Les indications de domaine ont pour fonction essentielle de mieux décrire le lexique (parmi d'autres descripteurs) et non de rendre compte de réalités extra-linguistiques.

Pour l'utilisateur de la ressource, le codage des domaines permet d'identifier automatiquement le vocabulaire relatif à thème donné : il suffit d'extraire toutes les unités lexicales associées à un domaine spécifique. Par exemple, une requête portant sur les termes de la physique dans Morfetik donnera instantanément la réponse suivante : *nombre atomique, neutrino muonique, mesure magnétique, moment angulaire, intelligence fluide, dilatation cubique, échange adiabatique, résistance mécanique*, etc. Il devient alors possible d'envisager, à partir de l'ensemble des données enregistrées dans les dictionnaires, la constitution automatique de lexiques spécialisés ou, plus généralement, relatifs à un domaine d'activité particulier.

Enfin, les formes pronominales ont également été implémentées. Au niveau de l'interface d'édition, les verbes sont donc maintenant répartis dans trois tables différentes : les verbes n'ayant pas d'emploi pronominal, les verbes pouvant avoir un usage pronominal et les verbes ayant un usage uniquement pronominal :

**Figure 8**

Usage pronominal du verbe *aimer* (image prise sur le site Morfetik)

aimer		Verbe (vrb)		Ind-pr	1	S		
<b>Infinif</b>		<b>Lemme</b>		<b>Notes</b>				
aimer		aimer						
<b>Indicatif présent</b>				<b>Indicatif imparfait</b>		<b>Passé simple</b>		
<b>J' (Je)</b>	(m) aime	<b>J' (Je)</b>	(m') aimais	<b>J' (Je)</b>	(m') aimai			
<b>Tu</b>	(t') aimes	<b>Tu</b>	(t') aimais	<b>Tu</b>	(t') aimas <b>Rare</b>			
<b>Il / Elle / On</b>	(s) aime	<b>Il / Elle / On</b>	(s) aimait	<b>Il / Elle / On</b>	(s) aima			
<b>Nous</b>	(nous) aimons	<b>Nous</b>	(nous) aimions	<b>Nous</b>	(nous) aimâmes			
<b>Vous</b>	(vous) aimez	<b>Vous</b>	(vous) aimiez	<b>Vous</b>	(vous) aimâtes <b>Rare</b>			

### 3.6. Matrices morphologiques

Enfin, Morfetik propose des matrices morphologiques, alors que d'autres ressources se contentent de décrire les différentes formes liées à un lemme. Ces matrices sont particulièrement utiles car elles permettent d'étendre la couverture des dictionnaires de manière dynamique, notamment pour les parties du discours lexicales qui constituent des classes ouvertes. Sur corpus, ces matrices permettent une reconnaissance dynamique de formes inconnues, sans avoir à décrire explicitement les formes effectives. Cela est possible grâce aux informations contenues dans les matrices morphologiques, qui capturent les variations morphologiques d'un mot ou d'une forme linguistique.

## 4. Pour ne pas conclure

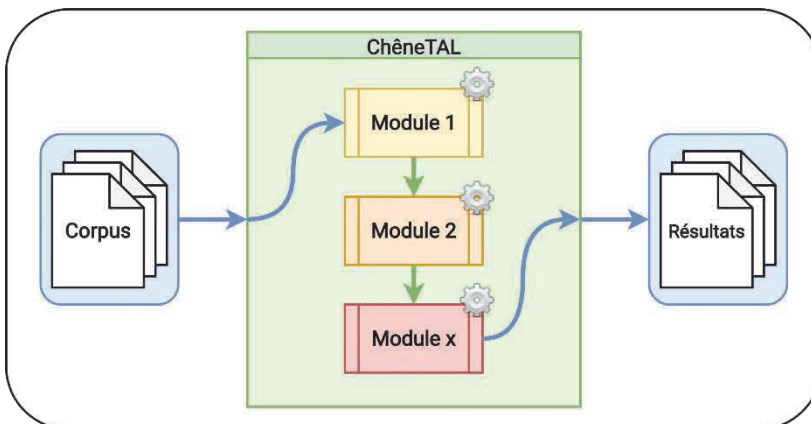
Dans le cadre d'applications centrées sur les ressources lexicales, l'enjeu est la mise en œuvre de ressources linguistiquement fondées, rigoureuses et actualisables. Le schéma traditionnel d'enrichissement de textes bruts suppose un ensemble de ressources linguistiques de bonne qualité pour l'analyse de la langue naturelle. La plus évidente de ces ressources est un ensemble de formes fléchies, associées à un certain nombre d'informations morphologiques, syntaxiques ou sémantiques, comme Morfetik.

Dans la continuité de ce travail, nous avons développé une plateforme de TAL, nommée ChêneTal, dans laquelle Morfetik est intégrée. Bien souvent, la manipulation des outils TAL est relativement complexe, et peu d'outils conviviaux ont été développés et peuvent être mis entre les mains de non-spécialistes. La plateforme ChêneTal a ainsi été conçue pour permettre la mise en place de chaînes hétérogènes de TAL en intégrant des logiciels existants en gestion et manipulation de corpus (Néoveille, SDMC) avec des modèles plus récents d'Intelligence Artificielle et en proposant une interface simple et intuitive pour les chercheurs de la communauté en TAL et pour ceux en Linguistique/Sciences Humaines et Sociales non spécialistes en informatique.

Cette plateforme intègre divers outils, dont Morfetik, qui offrent à l'utilisateur des possibilités de traiter des corpus et des procédures de recherche :

**Figure 9**

*Schéma simplifié du fonctionnement de ChêneTAL<sup>[1][5][6]</sup>*



L'utilisateur de ChêneTAL choisit un corpus (fourni par la plateforme ou importé par l'utilisateur) et définit une chaîne de traitement, c'est-à-dire une suite ordonnée de modules de traitement paramétrés par l'utilisateur. Ces modules sont des outils TAL développés au sein de l'équipe RCLN du LIPN<sup>11</sup> comme Néouvelle (outil de détection et de suivi des néologismes dans la presse en ligne) et SDMC (outil de fouille de données pour l'analyse de corpus et l'extraction de motifs). Une fois que le traitement est complété, ChêneTAL affiche l'ensemble des résultats retournés par les modules de traitement dans une interface intuitive :

**Figure 10**

*Interface de visualisation des résultats de ChêneTAL (image prise sur le site ChêneTAL)*

The screenshot shows the ChêneTAL web interface. At the top, there is a navigation bar with 'CORPUS', 'PIPELINE', 'PREVALIDATION', 'VALIDATION', 'LOGIN', 'FRANÇAIS', and a user profile 'JOHN DOE FALSE'. On the left, a sidebar lists 'Documents source' with folders for 'Fichier1', 'Fichier2', 'Fichier3', 'Néouvelle', 'Résultats de Néouvelle', and 'SDMC'. The main content area displays the title 'Nuits secrètes d'Aulnoye-Aymeries: Deux nouveaux noms et la répartition des scènes!' and a list of search results. The first result is highlighted in orange and contains the text: 'Nuits secrètes d'Aulnoye-Aymeries: Deux nouveaux noms et la répartition des scènes! L'actualité en Nord - Pas-de-Calais'. The second result is highlighted in blue and contains the text: 'Nuits secrètes d'Aulnoye-Aymeries: Deux nouveaux noms et la répartition des scènes! Le Festival Les Nuits Secrètes s'enrichit de deux nouveaux noms rendus publics dès vendredi midi. Citons Mr. Olzo déjà présent il y a trois ans et THE OISEAU. Quant au Grand Parcours, c'est complet. Mr Olzo avait déjà été l'invité principal au Jardin en 2013. Il revient aux Nuits Secrètes. PHOTO ARCHIVES LA VOIX C'est le principe que d'ignorer au fil de l'eau la programmation du festival [redacted] prévu les 29, 30 et 31 juillet. Les organisateurs des Nuits Secrètes ne changent pas une formule qui marche. Dès ce vendredi midi sur le site, ils annoncent deux nouveaux noms. Mr Olzo et THE OISEAU. Un trio punk – composé de Greg Carwright, Jack Yorke et Eric Friedl – formé à Memphis dans les années 90. Il y aura des bass réalistes déchirés et de la crête dans les rues d'Aulnoye-Aymeries, cette année, le dimanche. Revenons à Mr Olzo. Oiseau rare par rapport à d'autres habitués des grandes messes comme Vitalic (programmé le dimanche). Il avait déjà séduit le public de la scène du Jardin en 2013. Et revient donc... au Jardin le samedi 30 juillet. Voilà qui réjouira les commissaires et les autres, ceux qui apprécient cet artiste sous ses autres facettes. Quentin Dupieux de son vrai nom est aussi réalisateur de films. Il s'est musicalement fait connaître à travers son single Flat Beat et sa peluche jaune qui avait fait le buzz. Il fait actuellement son grand retour en musique avec son EP Hand in Fire. Le plus abouti selon lui, sur lequel apparaît la chanteuse britannique Cheryl XOX. Ça promet! Olivier Conan, le directeur artistique, aime bien faire revenir aux Nuits Secrètes les artistes qui ont su créer une ambiance au Jardin. C'est le cas. Outre ces deux noms d'artistes rejoignant la cohorte de ceux déjà connus, Seth Sue, Alice on the roof notamment, c'est toute la répartition des scènes (lire ci-dessous) que l'organisation dévoile. Avec le Grand Parcours déjà complet. Y aurait-il des paris sur une présence Souchon-Voulzy sur ce parcours ? On peut l'imaginer... Ça y est, la répartition des artistes programmés sur le festival est tombée ! Vendredi 29 juillet Grande scène

Morfetik est donc une plateforme en constante évolution qui vise à se développer progressivement pour améliorer le traitement des données textuelles.

## Références citées

Buvet, P.-A., Cartier, E., Issac, F., Mathieu-Colas, M., Mejri, S. & Madiouni, Y. (2009). Morfetik, ressource lexicale pour le TAL. *16ème Conférence sur le Traitement Automatique des Langues Naturelles, Actes du colloque TALN 2009, Juin 2009, Senlis (France)*, 1–10.

<sup>11</sup> Laboratoire d'Informatique de Paris Nord, LIPN, UMR CNRS 7030, Université Sorbonne Paris Nord.

- Cartier, E. (2011). Néologie et description linguistique pour le TAL. *Langages* 3(183), 105–117.
- Cartier, E. (2019). Néoveille, plateforme de repérage et de suivi des néologismes en corpus dynamique. *Neologica : revue internationale de la néologie* 13, 23–54.
- Cartier, E. & Sablayrolles, J.-F. (2009). Néologismes, dictionnaires et informatique. *Cahiers de Lexicologie* 2008-2(93), 175–192.
- Clément, L., Lang, B., & Sagot, B. (2004). Morphology based automatic acquisition of large-coverage lexica. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, 1841–1844.
- Courtois, B. (1990). Un système de dictionnaires électroniques pour les mots simples du français. *Langue française* 87, 11–22.
- Grezka, A. (2017). Morfetik, un dictionnaire morphologique : illustration avec le lexique de la perception. Dans E. Biardzka, M. Dańko, G. Komur-Thillooy & F. Marsac (éds), *La Perception en langue et en discours, Echo des études romanes* (89–99). Jihočeská univerzita.
- Grezka, A. (2020). Variabilité et traitement automatique des langues. *Linguisticae Investigationes* 43(2), 280–299.
- Grezka, A., Cartier, E. & Mathieu-Colas, M. (2015). Dictionnaires morphologiques du français contemporain : présentation de Morfetik, éléments d'un modèle pour le TAL, *Actes du colloque TALN 2015, 22–25 juin 2015, Université de Caen Basse-Normandie, Caen (France)*, 466–472.
- Gross, G. (1996). *Les expressions figées en français noms composés et autres locutions*. Ophrys.
- Hathout, N., Sajous, F. & Calderone, B. (2014). GLÀFF, a Large Versatile French Lexicon. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1007–1012.
- Mathieu-Colas, M. (1990). Orthographe et informatique : établissement d'un dictionnaire électronique des variantes graphiques. *Langue française* 87, 104–111.
- Mathieu-Colas, M. (2009). Morfetik : une ressource lexicale pour le TAL. *Cahiers de Lexicologie* 94, 137–146.
- Mathieu-Colas, M. (2014). Flexion des noms composés : principes de codage. Dans Z. Gavriilidou & A. Revithiadou (éds), *Studies dedicated to Professor Emeritus Anna Anastasiadis-Symeonidis of the Aristotle University of Thessaloniki*, 196–210.
- Sagot, B. (2010). The *Lefff*, a freely available and large-coverage morphological and syntactic lexicon for French. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, 2744–2751.
- Sajous, F., Hathout, N. & Calderone, B. (2013). GLÀFF, un Gros Lexique À tout Faire du Français. *Actes de la conférence Traitement Automatique des Langues Naturelles (TALN 2013)*, 285–298.


Sajous, F., Hathout, N. & Calderone, B. (2014). Ne jetons pas le Wiktionnaire avec l'origine du Web ! Études et réalisations fondées sur le dictionnaire collaboratif. *Actes du 4<sup>e</sup> Congrès Mondial de Linguistique Française (CMLF 2014)*, 663–680.





**Georges Kleiber**

Université de Strasbourg  
France

 <https://orcid.org/0000-0002-4808-1108>

# Sémiotique des démonstratifs cataphoriques

## Semiotics of cataphoric demonstratives

### Abstract

The studies of anaphoric expressions have shown that we can and must link the stable results acquired over the last forty years on their functioning to more recent socio-cognitive, taxonomic, multimodal, etc. perspectives and examine and explain two aspects of cataphoric demonstratives (CDs): on the one hand, their constructional scheme and, on the other, the formal and semantic nature of their subsequent or postcedent.

We propose here to continue this analysis of CDs, this time looking at their referential operation as such and asking the question: how do CDs make us reach the intended referent? What precisely is the interpretative procedure involved?

Our analyses show that rather than directly contrasting CDs with anaphoric demonstratives, it is more relevant and, therefore, more fruitful to compare them directly with gestural demonstratives. As we have seen, the two uses of demonstratives correspond to the same cognitive situation, that of introducing a new referent into discourse memory. This cognitive kinship gives rise to a common semiotic hybridity and, consequently, a referential operation, which we have detailed in nine points, that turns out to be largely identical, the differences observed being directly due to the difference in the nature of the index.

### Keywords

Cataphoric demonstratives, cataphor, anaphora, endophoric, exophoric, subsequent, ontological constraint, constructional constraint, adjacency, referent, referential identification

*À la mémoire de Gaston Gross, dont l'optimisme,  
sagacité et la robuste ludo-inventivité créatrice  
ont fait rayonner la linguistique française bien  
au-delà des frontières de l'Hexagone*

## 1. Introduction : Comment mettre l'avant après ?

Nous avons abordé, il y a une vingtaine d'années déjà, les démonstratifs cataphoriques, mais uniquement ceux de structure *Ce + N + relative* (Kleiber, 2004a, 2004b, 2005, 2006), spécifiques comme dans (1) ou génériques comme dans (2) :

- (1) *Tu te souviens de ce prof qui ne donnait que de bonnes notes ?*
- (2) *... il commençait à sentir cet accablement que vous cause la répétition de la même vie, lorsqu'aucun intérêt ne la dirige et qu'aucune espérance ne la soutient* (Flaubert, *Madame Bovary*, 150)

Nous n'avions guère touché aux démonstratifs cataphoriques du type de (3)–(7), que l'on peut appeler *standard*<sup>1</sup>, dans la mesure où leur cataphoricité, contrairement à celle des SN du type de (1)–(2), ne prête guère à discussion :

- (3) *« (...) et je vois ceci : un homme étendu par terre, sur le dos, les yeux ouverts et fixés sur le ciel, un autre homme, debout devant lui, et lui parlant par gestes seulement, l'homme à terre lui répondant des yeux seulement, tous les deux ayant l'air animé d'une prodigieuse bienveillance. »* (Baudelaire, *Les Paradis artificiels*, 1860 : 329, cité par Theissen, 2008)
- (4) *Vous ne pouvez imaginer ce supplice : ne rien avoir eu de la vie et ne rien attendre de la mort* (Mauriac, *Le nœud de vipères*, 212, cité par Kesik, 1989 : 133)
- (5) *L'un des arguments invoqués dans les milieux politiques français aurait été celui-ci : « impossible d'examiner une paix apportée à la république par un membre de la -maison de Bourbon »* (Martin du Gard, *Les Thibault*, cité dans Kleiber, 1994 : 184)
- (6) *Clarence déclare ceci : « De faux juges sont proposés à l'administration du monde »* (Maillard C. & M., 1977 : 152 ; cité par Kesik, 1989 : 36)

---

<sup>1</sup> Standard, parce qu'il y a unanimité pour leur reconnaître le statut de cataphorique, alors que ce n'est pas le cas des démonstratifs de type (1) et (2).

(7) *Sur un morceau minuscule ... je pus lire ce mot : PAX* (Mauriac, *Le nœud de vipères*, 210, cité par Kesik, 1989 : 133)

La parution récente d'un numéro de *Langue Française*<sup>2</sup> consacré à l'anaphore et la tenue, l'an dernier, d'un colloque sur la cataphore<sup>3</sup>, nous ont conduit à faire un « retour » sur les démonstratifs cataphoriques standard du type de (3)–(7), que nous noterons désormais DC. Le numéro de *Langue Française*, tout particulièrement par l'intermédiaire de l'introduction- présentation faite par Cornish (2021b) et de l'article de De Mulder (2021) sur l'adjectif démonstratif *ce* anaphorique, a montré que l'on pouvait et que l'on devait lier les résultats stables acquis au cours de ces quarante dernières années sur les expressions anaphoriques à des perspectives plus récentes<sup>4</sup>, socio-cognitives, taxinomiques, multimodales, etc. Le colloque de Strasbourg a montré que c'était une telle voie qu'il fallait suivre, pour les constructions cataphoriques en général<sup>5</sup> et pour les DC en particulier. Nous avons, pour notre part (Kleiber, à paraître a), choisi, lors de ce colloque, d'examiner et d'expliquer deux aspects des DC de (3)–(7), d'un côté, leur schéma constructionnel et, de l'autre, la nature formelle et sémantique de leur *subséquent* ou *postcédent*<sup>6</sup>.

Nous nous proposons ici de poursuivre cette analyse des DC, en nous penchant cette fois-ci sur leur fonctionnement référentiel proprement dit : comment s'effectue exactement la saisie cataphorique de leur référent ? Ou, dit autrement, comment nous font-ils atteindre le référent visé ? Quelle est précisément la procédure interprétative mise en jeu ?

Notre enquête comportera quatre parties. La première, très rapide, rappellera pourquoi les DC sont habituellement, aussi bien dans les typologies anciennes que récentes, rangés avec les démonstratifs anaphoriques dans la catégorie des

<sup>2</sup> Numéro 210 intitulé *L'anaphore revisitée* et dirigé par Cornish (2021a).

<sup>3</sup> Intitulé *Retour(s) sur la cataphore* (Strasbourg, 2–3 juin 2022).

<sup>4</sup> Développées, entre autres, par Cornish (2007, 2011, 2017 et 2021b), Vanderbauwhede (2012a, 2012b), Diessel et Coventry (2020), Doran et Ward (2019), Maes, Krahmer et Peeters (2022), Peeters, Krahmer et Maes (2021), Peeters et Özyürek (2016), Sidnell et Enfield (2017), etc.

<sup>5</sup> L'entreprise est plus que légitime, dans la mesure où la cataphore n'a pas connu un développement continu semblable à celui qu'a connu l'anaphore. On en est généralement resté, en ce qui concerne le français du moins, aux résultats mis en relief par les recherches effectuées durant la période 1980-2010, desquelles émerge avant tout l'ouvrage de Kesik (1989), qui est la seule monographie consacrée à la cataphore. À côté de Kesik (1989), il faut citer avant tout les travaux « pionniers » de Maillard (1974, 1989) et les analyses soit classificatoires, soit ponctuelles, de Fraser et Joly (1979, 1980), Härmä (1987), Danell (1990), Dixon (2003), Kesik (1991, 2014), Himmelmann (1996), Diessel (1999), Perdicoyanni-Paléologou (2001), Theissen (2008), Achard (2010) et Maass (2010), etc.

<sup>6</sup> Le terme de *postcédent* est emprunté à Maes, Krahmer et Peeters (2022),

démonstratifs *endophoriques*. Nous montrerons dans la deuxième partie qu'une telle classification, pour justifiée qu'elle puisse être si l'on prend uniquement en compte le critère du passage par le texte (soit antérieur ou subséquent), présente néanmoins un inconvénient majeur, celui de masquer l'asymétrie cognitive qu'il y a entre les DC et les démonstratifs anaphoriques. Cette asymétrie cognitive conduit à abandonner la piste classique de l'opposition intra-endophorique entre DC et démonstratifs anaphoriques et pousse à emprunter une autre voie comparative, qui conduit aux démonstratifs ostensifs. La comparaison que nous opérerons, dans les troisième et quatrième parties, entre les démonstratifs ostensifs et les DC, en même temps qu'elle établira la proximité cognitive des deux types de démonstratifs, permettra de mettre au jour, de façon plus précise et plus juste qu'auparavant, quel est le réel fonctionnement des DC.

## **2. Rappel : les DC cataphoriques = des expressions endophoriques « d'après »**

L'approche typologique standard des démonstratifs est, à un premier niveau de classification, « localiste », le critère déterminant étant le lieu, textuel ou non textuel, où se trouve le référent visé. De là, la vulgate qui divise, à un premier niveau, les démonstratifs en deux classes, les démonstratifs *endophoriques* (ou *diaphoriques* chez Maillard, 1974), où le référent est identifié par le truchement du texte, et les démonstratifs *exophoriques*<sup>7</sup>, où l'identification du référent se fait « en dehors du texte ». Cette prise en compte première du lieu, discursif ou non discursif, qui mène au référent visé par un SN démonstratif conduit à ranger les DC avec les démonstratifs anaphoriques sous la bannière des démonstratifs endophoriques, la différence entre les deux se faisant par la place qu'occupent dans le texte, par rapport à la position du SN démonstratif, l'expression ou les expressions qui fournissent le référent visé par le démonstratif. Si ces éléments discursifs figurent avant, on a affaire à un démonstratif anaphorique et cette ou ces expressions discursivement « antérieures » constitueront l'antécédent du démonstratif. S'ils sont, au contraire, postérieurs, on est en face d'un démonstratif cataphorique et on parlera de *subséquent* (ou *postcédent*) pour ces unités identificatrices d'*après*.

---

<sup>7</sup> Pour l'opposition *endophore* / *exophore*, cf. Halliday & Hasan (1976). Maillard (1974) oppose, pour sa part, les *aphoriques* (le texte n'est pas impliqué) aux *diaphoriques* (le texte est impliqué).

Une telle présentation donne à penser que les expressions cataphoriques (et donc les DC) ne sont que le symétrique « aval » ou « d'après » des expressions anaphoriques<sup>8</sup> (et donc des démonstratifs anaphoriques). Au lieu de l'ordre anaphorique normal, qui consiste à présenter d'abord dans le discours le référent Y et ensuite seulement l'expression X qui renvoie à Y, en situation de cataphore, on a l'ordre inverse : on anticipe en donnant par avance l'expression X dont l'identification référentielle se fait, postérieurement dans le discours, par l'expression Y. Wilmet (1986 : 175) formule comme suit l'opposition : « anaphorique, elle (l'endophore) livre la solution avant le problème, cataphorique, elle pose une énigme qu'elle résout après coup ». C'est dire, pour reprendre l'image de la charue et des bœufs de Wilmet (1986), qu'avec une cataphore on met la charrue « anaphorique » avant les bœufs « antécédents ».

Les travaux des dernières décennies du vingtième siècle, dont, comme déjà mentionnée, la monographie de Kesik (1989), ont toutefois montré qu'une telle position était beaucoup trop forte et qu'il fallait renoncer à la thèse de la cataphore comme « anaphore inverse », parce que les expressions cataphoriques présentent une série de contraintes que ne connaissent pas les expressions anaphoriques correspondantes. Pour ce qui est des DC, Maillard (1974, 1989), Kesik (1989, 2014), Theissen (2008), entre autres, ont relevé qu'ils étaient sujets à différentes contraintes auxquelles n'étaient pas soumis les démonstratifs anaphoriques.

Nous n'en rappellerons ici que deux d'entre-elles<sup>9</sup>, une contrainte ontologique et une contrainte constructionnelle. On observe, pour ce qui est de la première, que les DC ne peuvent renvoyer à des entités du 1<sup>er</sup> ordre (Lyons, 1980), alors que semblable contrainte n'affecte nullement les démonstratifs anaphoriques, comme le montrent (8)–(9) :

(8) *\*Il a vu ce chien : le setter irlandais*

*? Puis arriva ce train : le TER de 8h45*

(9) *Le setter irlandais est doux et affectueux. Ce chien n'aboie quasiment jamais*  
*Le TER de 8h45 entre en gare. Ce train dessert Rouffach et Merxheim*

La seconde contrainte fait ressortir que le subséquent est généralement contigu au DC, dont il n'est séparé à l'écrit que par les deux points (Theissen, 2008) :

<sup>8</sup> Cf., par exemple, Lerat (1983 : 79), qui considère la cataphore comme une anaphore « inverse ».

<sup>9</sup> D'autres contraintes portent sur la forme du subséquent qui, mis à part les DC du type de (7), doit être complexe – on parle alors de DC *résomptif* (Maillard, 1974, 1989) –, sur le type de démonstratif et sur le statut informationnel des énoncés comportant un DC, qui doit être d'allure thématique et non catégorique, etc.

(10) *Je ne dirai que ceci : « Lisez plutôt Proust ! »*

Cette adjacence peut être, certes, « interrompue », mais uniquement, du côté du DC et par des éléments appositifs ou incisives qui suivent le DC (Theissen, 2008) :

(11) *Je ne dirai que ceci, si vous le permettez : « Lisez plutôt Proust ! »*

Un tel élément intermédiaire n'est en effet pas possible du côté du subséquent, c'est-à-dire après les deux points et avant le subséquent. Il ne peut y avoir d'élément intermédiaire entre les deux points et le subséquent, comme le monter (12), où le placement après les deux points d'un élément *W* qui ne contribue pas à l'identification du référent rend la construction cataphorique bancale :

(12) *\*Je ne dirai que ceci : W, « Lisez plutôt Proust ! »*

Ces contraintes donnent lieu à un résultat stable pour les DC : il ne s'agit pas d'anaphores démonstratives inverses. Ce constat ne remet toutefois pas en cause le caractère endophorique des DC. L'abandon de la vue « symétrique » donne lieu, aujourd'hui, à une consensuelle version faible qui, tout en refusant de voir dans les DC l'inverse des démonstratifs anaphoriques, maintient le critère du « lieu de résidence » pour les deux et continue, par conséquent, de les réunir sous une même bannière, celle des démonstratifs endophoriques, c'est-à-dire des démonstratifs dont le référent est donné par le texte.

On ne saurait nier la légitimité de ce critère « localiste » qui, d'une part, regroupe les DC et les démonstratifs anaphoriques sous le chapeau subsumateur des démonstratifs endophoriques et qui, d'autre part, les différencie par le lieu, avant ou après dans le texte, du renvoi endophorique effectué. En amont pour les anaphoriques et en aval pour les DC, la différence de localisation étant responsable de la différence de contraintes observée. On ne saurait pas non plus nier qu'une telle classification en termes de lieu de résidence a des avantages typologiques intra- et interlinguistiques certains. Mais elle présente aussi un inconvénient et un inconvénient qui nous semble capital pour l'analyse des DC : elle conduit à les définir prioritairement par leur opposition aux démonstratifs anaphoriques et, *ipso facto*, écarte par avance toute comparaison directe avec les démonstratifs exophoriques. Or, comme le montrera notre deuxième partie, un tel rapprochement s'impose si l'on entend rendre compte de leur réel fonctionnement référentiel.

### 3. Asymétrie cognitive

On peut légitimement se demander ce qui peut bien justifier un rapprochement des DC avec les démonstratifs exophoriques. Ce qui pousse à les sortir du chapeau subsumateur endophorique pour les confronter directement à des démonstratifs exophoriques est que la différence entre les démonstratifs anaphoriques et les DC n'est pas uniquement une différence de place (antécédent/subséquent), mais aussi et avant tout une différence de situation cognitive entre les deux types de démonstratifs, différence directement due à la différence de localisation discursive qui les caractérise. Parler, comme on le fait couramment, de référent déjà présent dans le texte pour les anaphoriques et de référent qui ne vient qu'après dans le texte pour les cataphoriques empêche de voir que la différence de « place » dans le discours entraîne une asymétrie cognitive cruciale entre les deux types de procédure référentielle.

Voyons d'abord quelle est la situation cognitive du démonstratif anaphorique. La première chose à noter, souvent rappelée, mais qui n'est généralement pas mise en avant dans les présentations de l'opposition anaphore / cataphore, c'est que, contrairement à ce que peuvent donner à croire les définitions courantes de l'anaphore, le référent auquel renvoie une expression anaphorique n'est pas lui-même présent dans le texte ou dans le discours. Si tel était le cas, il faudrait – c'est bien connu – se méfier d'un *chien* anaphorique ! La seconde observation, qu'il faut aussi rappeler, car elle est également souvent oubliée dans les analyses des processus référentiels, c'est que, pour trouver le référent d'une expression anaphorique, l'interlocuteur ou lecteur ne revient pas en arrière dans le texte, il ne remonte pas la linéarité (spatiale ou temporelle) du texte ou du discours. On peut en faire l'expérience soi-même. Quand on lit un texte et qu'on rencontre une anaphore, on ne revient en arrière que si et seulement si on n'a pas compris quel était le référent de l'anaphore. Il en va ainsi pour nos démonstratifs anaphoriques. Au moment où l'interlocuteur ou lecteur entend ou lit une occurrence de démonstratif anaphorique, il n'a pas besoin de revenir en arrière dans le texte ou dans le discours pour chercher et trouver le « bon » référent. Du coup, la question à laquelle il faut répondre est : où trouve-t-il alors le référent introduit par l'expression-antécédent ? On peut aussi formuler cette interrogation autrement en se demandant : dans ces conditions, à quoi sert encore l'antécédent ? La réponse à ces questions oblige à postuler une entité cognitive qui n'est pas la chaîne discursive, mais qui résulte de cette chaîne discursive. Cette entité cognitive est l'espace mémoriel construit par le texte ou discours antérieur au SN démonstratif anaphorique, qui est en cours au moment de l'apparition du SN anaphorique. Le

discours/texte ne disparaît pas pour autant, puisque c'est lui qui crée et fait évoluer en progressant cet espace cognitif. Le référent d'un anaphorique n'est donc pas trouvé dans le texte, mais dans l'espace cognitif ou « univers de discours » (Himmelmann, 1996) créé par le texte antérieur, espace ou univers discursif valide, c'est-à-dire en cours, au moment d'apparition de l'expression anaphorique. Nous reprenons à Berrendonner (1983) et à Reichler-Béguelin (1988) le terme de *mémoire discursive* qui nous semble approprié pour parler de cet espace évolutif cognitif que crée et modifie selon sa progression le texte/discours. C'est dire qu'un démonstratif anaphorique renvoie à un référent déjà installé ou introduit dans la mémoire discursive, c'est-à-dire dans la mémoire construite par le texte ou discours et qui est en vigueur au moment de l'énonciation du démonstratif.

Soit à présent la situation cognitive de la cataphore démonstrative. On ne peut évidemment y voir le symétrique aval de celle de l'anaphore démonstrative, tout simplement, parce qu'au moment où apparaît le DC, le référent de ce démonstratif ne peut faire partie de la mémoire discursive en cours, puisqu'il n'y a pas encore été introduit et qu'il ne va l'être qu'après l'occurrence du démonstratif. Au moment où survient dans le discours un DC, il n'y a pas de référent disponible dans la mémoire discursive en cours.

Si on compare cette situation à celle des démonstratifs anaphoriques, on voit clairement que la situation cognitive n'est pas la même. Lorsque survient un démonstratif anaphorique, le référent visé fait déjà partie de la mémoire discursive, le démonstratif ne pouvant participer à son introduction, puisque celle-ci a déjà été opérée par le discours antérieur. Le démonstratif anaphorique a pour rôle de renvoyer à un référent déjà introduit (directement ou indirectement) dans la mémoire discursive. Lorsqu'on a affaire à un DC, celui-ci, au contraire, participe lui-même à l'introduction du référent visé dans la mémoire discursive, puisqu'il annonce dans le discours un référent qui sera livré par le discours qui suit. L'introduction d'un référent par un DC comporte ainsi deux phases introductrices dans la mémoire discursive : une première, incomplète ou non saturée, celle opérée par le DC, qui signale qu'un référent va être donné dans le discours qui suit, et la seconde qui comble l'attente suscitée par la première en livrant sub-séquentiellement le référent en question. Un DC a donc un rôle différent de celui d'un démonstratif anaphorique, puisqu'il participe lui-même à l'introduction de son référent en mémoire discursive, alors que ce n'est guère le cas d'un démonstratif anaphorique.

Cette différence de rôle cognitif entraîne aussi une différence de procédure référentielle, le renvoi à un référent déjà disponible en mémoire discursive par un démonstratif anaphorique ne pouvant se faire de la même manière que l'in-



roduction biphasique d'un nouveau référent par un DC. Même si les deux présentent des traits communs qui découlent directement de leur statut de démonstratif et même si, également, les deux impliquent le discours – ce qui légitime leur réunion sous le chapeau d'*endophoriques* – le fait qu'ils effectuent des tâches radicalement différentes, l'un renvoyant à un référent déjà présent dans la mémoire discursive, l'autre œuvrant au contraire à y installer un référent nouveau, ce fait a nécessairement pour conséquence des opérations de recouvrement référentiel également différentes.

Il en découle une autre conséquence, sur la démarche à suivre : pour répondre à la question centrale de leur fonctionnement – *comment se fait cette introduction biphasique du référent en mémoire discursive ?* – il convient d'abandonner la piste classique qui consiste à analyser les DC en opposition aux démonstratifs anaphoriques et d'emprunter une autre voie, celle qui ouvre à une comparaison des DC avec, s'il en existe, des démonstratifs non endophoriques, donc des exophoriques, qui répondent au schéma cognitif des DC, celui de participer à l'introduction d'un référent en mémoire discursive.

#### **4. Changement de cap : comparaison avec les démonstratifs ostensifs**

Il n'est guère besoin de chercher longtemps pour trouver de tels démonstratifs exophoriques. Les démonstratifs ostensifs ou gestuels, c'est-à-dire les démonstratifs qui font intervenir une ostension dans la procédure référentielle, ont pour effet d'installer un nouveau référent dans la mémoire discursive. Dans un énoncé tel que (13) :

(13) *Prenez ce livre !* (dans la situation où un libraire désigne avec le doigt à un client un livre précis sur une étagère comportant d'autres livres)

au moment où survient le démonstratif, le référent n'est pas encore présent dans la mémoire discursive. Il ne l'est que postérieurement à l'énonciation du démonstratif et à l'accomplissement du geste d'ostension. On se retrouve donc dans la même situation cognitive que celle des DC : un référent « nouveau », dans le sens de non encore disponible dans la mémoire discursive, s'y trouve introduit *via* un SN démonstratif. Un tel constat, étant donné la différence gestuelle / non gestuelle de livraison du référent qui est manifeste entre les deux types de

démonstratifs, invite à voir de plus près comment fonctionnent exactement ces démonstratifs ostensifs. Plusieurs aspects et caractéristiques sont à prendre en compte. Nous en distinguerons neuf qui nous serviront ensuite de points de comparaison avec les DC.

#### 4.1. Un signe sémiotiquement hybride : segment linguistique + ostension

En premier lieu, particularité déjà signalée, un démonstratif ostensif n'est pas seulement constitué du segment linguistique que représente le SN démonstratif. On a affaire à un signe sémiotiquement hybride, multimodal, puisqu'en plus de la partie linguistique proprement dite que représente le SN démonstratif, il comporte une partie non linguistique, le geste d'ostension.

#### 4.2. Concomitance entre l'énonciation du SN démonstratif et l'ostension

La deuxième caractéristique a trait à la relation entre l'énonciation du SN démonstratif et le geste effectué par le locuteur. Le point crucial est que le geste doit être concomitant à l'énonciation du SN démonstratif. C'est dire que, normalement, il ne peut avoir lieu ni avant ni après, mais il doit accompagner l'énonciation même du démonstratif. On retrouve là, point que nous ne développerons pas ici, la *token-réflexivité* que l'on reconnaît bien souvent aux démonstratifs : le geste concomitant constitue le premier maillon de la chaîne indexicale qui part de l'occurrence du démonstratif et conduit au référent.

#### 4.3. Index gestuel

Troisièmement, on rappellera qu'une ostension est un *index*. L'ostension donne le référent visé de manière indexicale et non pas symbolique. L'interlocuteur doit suivre la direction du geste pour trouver le référent visé. Il ne faut donc pas que cette ostension soit équivoque : si notre libraire de ci-dessus désigne avec le doigt tendu un bouquin, mais que sa tête et donc son regard sont manifestement dirigés vers un autre livre situé dans une direction opposée à celle indiquée par le doigt, il peut y avoir de sérieux problèmes quant à la trouvaille du « bon » livre désigné<sup>10</sup>.

<sup>10</sup> Un strabisme divergent peut ainsi être la cause d'embarras référentiels.

#### 4.4. Espace physique comme espace de « monstration »

Le quatrième point touche à la « sortie » du langage qu'entraîne le geste concomitant à un démonstratif ostensif. L'ostension nous fait quitter l'espace linguistique et nous invite à diriger notre regard sur un endroit de l'espace où se trouve un objet qui est soit le référent lui-même (ostension directe), soit un objet qui conduit au « bon » référent (ostension indirecte ou *différée* selon Quine, 1971).

#### 4.5. Entre le geste et l'objet pointé : une relation de contiguïté

Cinquième caractéristique : la relation entre l'index ostensif et le référent « pointé » est une relation de contiguïté spatiale. C'est dire que, sur la direction tracée par l'ostension, il ne peut y avoir entre le regard et l'objet auquel on veut référer par le geste un objet intermédiaire, qui coupe en somme la ligne de monstration, puisque ce serait cet élément qui deviendrait alors l'objet visé, surtout avec des démonstratifs du type « chosiste » comme *ceci*<sup>11</sup>.

#### 4.6. Nature « spatiale » de l'objet « pointé »

Le sixième aspect qui caractérise les démonstratifs ostensifs concerne la restriction ontologique qui pèse sur le type d'entité qui se trouve « pointée ». Comme il s'agit d'une monstration dans l'espace l'objet ainsi spatialement visé, que ce soit le référent en cas d'ostension directe ou seulement un objet qui mène au référent en cas d'ostension différée, cet objet doit être lui-même de nature « spatiale ». On ne peut montrer que du « montrable », ce qui exclut par avance les entités abstraites.

#### 4.7. Ostension directe et ostension « différée »

Notre septième point revient sur la distinction, signalée dans notre point numéroté quatre, entre ostension directe et ostension indirecte ou différée. Trois situations sont en fait possibles, l'une qui relève de la monstration directe, les deux autres de la monstration indirecte. Dans la première, le pointage se fait directement sur le référent du démonstratif gestuel. Celui-ci est donc spatialement

---

<sup>11</sup> Ce point nécessiterait des explications supplémentaires.

présent lui-même dans la situation d'énonciation, à l'endroit que délimite dans cet espace le geste de monstration.

Le premier cas de monstration différée est proche de l'ostension directe, dans la mesure où le geste, même s'il ne désigne pas effectivement le référent, est quand même directement lié à lui, dans la mesure où l'objet sur lequel il pointe est une « représentation » ou « icône » du référent visé (cf. photo, image, peinture, etc., qui représente iconiquement le référent). Le geste conduit dans ce cas au référent par l'intermédiaire d'une icône de ce référent, comme l'illustre l'exemple (14), où le geste qui accompagne l'énonciation du SN *ce grand barbu* désigne sur une photo<sup>12</sup> la représentation photographique de Paul :

(14) *Ce grand barbu, c'est Paul*

Dans le second cas de monstration différée, souvent donné comme étant le seul cas de représentation indirecte<sup>13</sup>, l'ostension ne pointe pas sur une icône du référent visé, mais sur un objet qui permet de remonter au référent visé, moyennant une relation préétablie entre l'objet désigné et le référent, comme la relation métonymique *livre* → *auteur* qui autorise, par exemple, l'emploi du démonstratif *cet auteur* pour renvoyer à l'auteur du livre sur lequel pointe le locuteur :

(15) *Cet auteur génial* (avec ostension sur un livre)

#### 4.8. Postériorité

Notre huitième et avant-dernière précision concerne la relation temporelle entre la perception par l'interlocuteur du démonstratif ostensif et celle de l'objet (soit le référent, soit un objet qui mène au référent) sur lequel pointe le geste. Il s'agit d'une relation de postériorité temporelle, même si, de prime abord, étant donné l'absence d'un support spatial la représentant, elle ne s'impose pas immédiatement, puisqu'on peut avoir l'impression que le geste est perçu en même temps que l'objet qu'il désigne. Or, elle est bien réelle pourtant : l'interlocuteur ne peut percevoir le « bon » référent qu'après avoir perçu la direction indiquée par le geste montreur.

<sup>12</sup> L'ostension, dans ce cas, se fait dans l'espace (portrait, peinture, etc.) de représentation et non plus dans l'espace d'énonciation.

<sup>13</sup> Parce que, d'un point de vue psychologique, comme le rappelle le célèbre tableau *La trahison des images* de Magritte (cf. *ceci n'est pas une pipe*), on ne sépare pas l'icône de l'objet qu'elle représente, dans la mesure où on a l'impression qu'il n'y a qu'un référent en jeu, l'objet représenté.

#### 4.9. Un interlocuteur qui, pour « voir », doit d'abord « regarder »

La neuvième et dernière caractéristique a trait à l'état dans lequel se trouve l'interlocuteur : le point à souligner est que l'interlocuteur est actif et non passif, puisque le démonstratif ostensif l'invite à « regarder » dans la direction qu'il indique pour trouver quel est l'objet « montré ». Pour qu'il « voie » ou « trouve » cet objet, il faut bien évidemment qu'il regarde<sup>14</sup> d'abord dans la direction indiquée par l'ostension.

Nous pouvons à présent retourner aux DC pour voir si l'on retrouve dans leur fonctionnement des caractéristiques similaires à celles que nous venons de mettre en relief pour les démonstratifs gestuels. Si oui, le rapprochement, opéré ci-dessus sur la base de leur situation cognitive commune, se trouvera justifié.

## 5. Retour aux DC

### 5.1. Les DC sont-ils sémiotiquement hybrides ?

Avec le premier point, la dualité multimodale du démonstratif gestuel, l'affaire semble plutôt mal engagée, dans la mesure où les DC ne semblent pas correspondre à un signe sémiotiquement hybride semblable à celui formé par le segment du SN démonstratif et l'ostension concomitante. De prime abord, on ne retrouve en effet pas, du côté des DC, l'équivalent de l'ostension caractéristique des démonstratifs gestuels. Mais, en y regardant de plus près, on voit bien qu'un DC n'est pas uniquement constitué du constituant segmental que représente le SN démonstratif. Il s'avère aussi sémiotiquement hybride, puisque le SN démonstratif, de nature segmentale, s'accompagne d'un élément suprasegmental prosodique, donc concomitant, qui remplit la fonction du « geste » concomitant de l'emploi ostensif. L'expression du SN démonstratif a en effet une intonation qui n'est pas clôturante comme une finale de phrase, mais qui reste sur la même hauteur, en suspens, et qui, par cet arrêt-ouverture discursif, marque, de manière non symbolique, mais indexicale, comme le geste indique la direction qu'il faut suivre pour trouver l'objet visé, que c'est dans le texte qui va suivre qu'on trouvera

---

<sup>14</sup> Rappelons ici la différence significative entre *regarder* et *voir* relativement à l'impératif. On peut demander à quelqu'un de regarder (cf. *Regarde !*), mais non de voir (cf. *\*Vois !*).

le référent à introduire dans la mémoire discursive. À l'écrit, cette intonation d'attente se trouve marquée par la ponctuation « suspensivo-ouvrante » que constituent les « deux points ».

## 5.2. Il y a également concomitance

Le deuxième point mis en avant pour les démonstratifs gestuels se trouve également vérifié par les DC, puisque l'élément suprasegmental caractéristique d'un DC est, tout comme le geste dans le cas des démonstratifs ostensifs, concomitant à l'énonciation même du SN démonstratif. Cette contrainte de concomitance se révèle encore plus forte pour les DC que pour les démonstratifs gestuels, puisque, par définition même, un élément suprasegmental ne saurait apparaître ni avant ni après son support segmental.

## 5.3. Élément suprasegmental des DC est un index

Troisièmement, l'élément suprasegmental des DC est de même nature sémiotique que l'ostension des démonstratifs gestuels : c'est un signe indexical et non un signe symbolique ou iconique, mais étant donné sa nature discursive, il ne peut pointer que sur un élément présent dans le même espace, c'est-à-dire dans l'espace du discours. C'est là que démonstratifs ostensifs et DC se séparent crucialement : l'espace de « monstration » n'est pas le même. Alors que la nature spatiale du geste accompagnant un démonstratif ostensif fait que son champ de monstration ne peut être que l'espace physique, la nature discursive de l'élément suprasegmental d'un DC entraîne un pointage qui ne peut s'exercer que dans l'espace de cet index suprasegmental, c'est-à-dire dans l'espace du discours.

Cette différence d'espace indexical a deux conséquences majeures. En premier lieu, étant donné que l'espace discursif est unidimensionnel – le langage étant linéaire – le pointage qu'effectue un DC s'avère également unidimensionnel. Il ne connaît pas de variabilité « directionnelle » semblable à celle que peut connaître un geste dans l'espace physique. Et, deuxièmement, comme il s'agit obligatoirement d'un référent nouveau à introduire, il ne peut porter que sur l'espace discursif à venir.

On voit ainsi que le rôle du constituant suprasegmental d'un DC, par ses caractères intonatifs et prosodiques propres, est de signaler à l'interlocuteur qu'il doit en quelque sorte « dresser l'oreille » ou « ouvrir les yeux », que le référent que

réclame tout démonstratif va être donné, directement ou indirectement<sup>15</sup>, dans le discours qui suit. On peut, toute proportion gardée, comparer son comportement à celui d'un « chien d'arrêt » qui, au cours d'une partie de chasse, s'immobilise pour signaler au chasseur la présence d'un gibier non encore entrevu.

#### 5.4. Une « entrée » dans le discours

Le quatrième point de notre comparaison avec les démonstratifs ostensifs révèle une autre différence, importante : alors que l'ostension des démonstratifs gestuels entraîne une « sortie » du discours pour que l'on trouve le référent visé, l'index suprasegmental des DC, signale l'inverse, à savoir que le référent du DC va faire son « entrée » dans le discours. Si le geste des démonstratifs ostensifs invite l'interlocuteur à regarder en dehors du discours, l'index prosodique des DC demande, au contraire, d'attendre la suite du discours qui livrera, encore une fois de manière directe ou indirecte, le référent attendu.

#### 5.5. Une relation de contiguïté discursive stricte

Un autre point de similitude – ce sera notre cinquième observation comparative – concerne la relation entre l'index suprasegmental d'un DC et le subséquent discursif qui livre le référent : elle est identique à celle qui unit le geste d'un démonstratif ostensif et l'élément spatial sur lequel il « pointe », puisque, comme déjà signalé ci-dessus, il s'agit nécessairement d'une relation de contiguïté stricte. Il ne peut pas y avoir d'élément intermédiaire entre les deux points (ou la suspension à l'oral), comme le rappelle l'énoncé (12) repris sous (16) et le subséquent, de même que dans le cas du démonstratif ostensif il ne peut pas y avoir sur la ligne de « monstration » d'objet intercalé avant l'objet « pointé » :

(16) \**Je ne dirai que ceci : W, « Lisez plutôt Proust ! »*

Il est à remarquer que cette contrainte de contiguïté commune aux démonstratifs ostensifs et aux DC a l'air plus nette avec les seconds qu'avec les premiers, parce que la longueur de l'ostension qui accompagne un démonstratif gestuel est variable, l'objet « pointé » pouvant être plus ou moins proche du locuteur « pointeur », alors qu'une telle variation n'est pas possible entre l'occurrence du DC et

---

<sup>15</sup> Nous reviendrons sur cette différence ci-dessous.

le subséquent. La position de celui-ci n'est pas susceptible d'être plus ou moins proche, tout simplement parce que l'index suprasegmental ne peut pointer que sur l'espace discursif qui lui est directement adjacent. Il ne peut pas y avoir de « vide » discursif variable avant le subséquent, comme il peut y avoir un « vide » spatial variable avant l'objet « pointé ». SN démonstratif et subséquent sont en quelque sorte toujours à la même longueur l'un de l'autre, car toujours immédiatement ou strictement contigus.

## 5.6. Contrainte ontologique : exclusion des entités du 1<sup>er</sup> ordre

On pourrait penser, à première vue, que le sixième point mis en avant dans notre caractérisation des démonstratifs ostensifs suppose à un rapprochement des DC avec ce type de démonstratifs exophoriques, puisque, du point de vue ontologique, DC et démonstratifs ostensifs sont en quelque sorte en distribution complémentaire en ce qui concerne la nature de l'objet sur lequel ils « pointent ». Les DC ont pour référent des entités « abstraites », et non, comme rappelé ci-dessus avec (8), que nous reprenons ici sous (17), des entités du 1<sup>er</sup> ordre, qui constituent, au contraire, la cible prédestinée des gestuels :

- (17) \**Il a vu ce chien : le setter irlandais*  
 ? *Puis arriva ce train : le TER de 8h45*

La raison d'une telle différence n'a rien de surprenant et ne remet nullement en cause le rapprochement opéré entre les deux types de « montreurs ». Elle découle directement de la différence de « pointeur » à l'œuvre dans les deux types de référence indexicale. Si l'index est gestuel, c'est-à-dire spatial, comme dans le cas du démonstratif ostensif, l'entité désignée<sup>16</sup> ne peut être, évidemment, que de statut spatial également. Si, cependant, il est d'ordre discursif, le subséquent ne saurait être, lui aussi, que d'ordre discursif. Dit autrement, un pointeur discursif ne peut pointer que sur du discursif, tout comme un pointeur spatial ne peut pointer que sur du spatial<sup>17</sup>.

<sup>16</sup> Rappelons encore une fois, pour éviter toute équivoque, qu'en cas d'ostension indirecte l'entité « montrée » n'est pas le référent lui-même.

<sup>17</sup> La réponse n'est toutefois pas complète. Il conviendrait de prolonger cette analyse en expliquant pourquoi un élément discursif comme le nom d'une entité spatiale ne représente pas une



### 5.7. DC directs et DC indirects

Nous avons rappelé, en septième place, que les démonstratifs gestuels pouvaient donner lieu, non seulement à une ostension directe, mais également à une ostension « différée ». Il en va ainsi également du côté des DC, même si des dénominations dissociatives dans la littérature contribuent à masquer cette similitude. On ne reconnaît généralement comme démonstratif cataphorique que le cas où, semblable à ce qui se passe pour les démonstratifs anaphoriques, le DC ne renvoie au référent que par l'intermédiaire d'un constituant linguistique (le subséquent ou postcédent). La référence n'est donc pas directe, puisque le référent n'est pas lui-même présent dans le discours. C'est le cas de nos exemples (3)–(6) où nous n'avons à chaque fois qu'une « représentation » du référent visé. Dans (3) repris sous (18) :

- (18) « (...) et je vois ceci : un homme étendu par terre, sur le dos, les yeux ouverts et fixés sur le ciel, un autre homme, debout devant lui, et lui parlant par gestes seulement, l'homme à terre lui répondant des yeux seulement, tous les deux ayant l'air animé d'une prodigieuse bienveillance. »

le subséquent de *ceci* n'est évidemment qu'une « représentation » de la scène ou situation visée et non la scène ou situation elle-même<sup>18</sup>.

Il y a toutefois des cas où le « référent » anticipé par un DC est bien présent lui-même et où on a donc bien affaire, comme avec un démonstratif ostensif direct, à la présence effective du référent dans l'espace discursif de monstration. Il s'agit du cas de (7), noté ici (19) :

- (19) *Sur un morceau minuscule ... je pus lire ce mot : PAX*

où le subséquent de *ce mot*, à savoir le mot *Pax*, étant lui-même un objet discursif, figure directement dans le discours. De tels cas, comme déjà noté *supra*, sont bien souvent reconnus dans la littérature depuis Lyons (1980), comme relevant, non de la cataphore, mais de la deixis et plus précisément de la *deixis discursive*<sup>19</sup>. Nous ne discuterons pas ici des avantages ou désavantages des choix terminolo-

---

entité « discursive » pertinente pour donner lieu à un l'emploi d'un DC. Nous avons répondu à cette « énigme » dans Kleiber (à paraître b).

<sup>18</sup> Ce point est à développer en relation avec la contrainte ontologique qui exclut les entités du 1<sup>er</sup> ordre. Cf. Kleiber (à paraître b) pour une analyse de cette question.

<sup>19</sup> Qui s'applique aussi aux cas où l'expression anaphorique renvoie directement à une unité discursive.

giques opérés par les uns et les autres. Le point qui nous semble primordial est que les DC peuvent, tout comme les démonstratifs ostensifs, « montrer » directement ou indirectement.

Ce que nous soulignerons, en revanche, c'est que la différence d'espace de renvoi (discours/DC) / (espace physique/démonstratifs ostensifs) a pour conséquence une inversion d'emploi : pour les démonstratifs ostensifs, l'emploi « normal » ou courant est constitué par l'ostension directe, alors qu'avec les DC, c'est le contraire. Comment expliquer une telle inversion ? Cette dissymétrie a pour fondement la nature même du langage : le langage, chose bien connue, sert avant tout à parler d'autre chose que de lui-même et ce n'est que dans des circonstances bien précises, donc limitées, qu'il œuvre « directement », c'est-à-dire métalinguistiquement<sup>20</sup>. Ceci explique pourquoi l'emploi premier des DC n'est pas l'emploi direct, comme dans (19), où il y a effectivement une ostension directe sur le mot *PAX* lui-même, mais l'emploi indirect, où l'index suprasegmental ne pointe que sur une « représentation » du référent. Dans (18), comme déjà signalé ci-dessus, l'élément que « touche » le constituant suprasegmental de *ceci* n'est pas la « situation » elle-même, mais qu'une « représentation » de celle-ci.

Une précision supplémentaire s'impose. Les exemples de discours direct illustrés ci-dessus par (6) ne sont pas à ranger, comme on peut être tenté de le faire, avec les DC directs du type de (19). Le discours direct n'est en fait jamais ... direct, dans le sens où « direct » indiquerait qu'il s'agit du référent même, c'est-à-dire du discours effectivement tenu. Les spécialistes du discours direct ont montré avec des arguments convaincants (cf. Authier-Revuz, 2021), qu'un discours direct n'était qu'un discours ... *rapporté* et – c'est ce qui est à l'origine de l'appellation *discours direct* – un discours qui se trouve rapporté de manière iconique, c'est-à-dire un discours qui mime un discours. Le discours direct de (6), repris ici sous (20), n'est ainsi qu'une présentation iconique du discours attribué à Clarence :

(20) *Clarence déclare ceci* : « *De faux juges sont proposés à l'administration du monde* ».

## 5.8. Relation de postériorité

Notre huitième et avant-dernier point de comparaison concerne la relation temporelle entre l'énonciation du SN démonstratif et la perception de l'objet visé par l'index concomitant au SN. Il est presque inutile de rappeler que, tout comme

<sup>20</sup> D'où l'appellation de *cataphore métalinguistique* chez Kesik (1989).

les démonstratifs ostensifs, les DC « roulent » sur une relation de postériorité, puisqu'aussi bien les définitions et présentations standard des expressions cataphoriques que les étiquettes de *subséquent* ou de *postcédent*, employées pour dénommer le segment qui livre le référent d'un DC, mettent en avant la condition de postériorité. Ce que l'on peut souligner, cependant, c'est que cette postériorité, parce que, pour le discours, elle se double à l'écrit d'une représentation spatiale iconique linéaire, paraît plus évidente avec les DC qu'avec les démonstratifs ostensifs.

### 5.9. Actif / Passif

La neuvième caractéristique des démonstratifs gestuels est également manifeste avec les DC. L'interlocuteur ou lecteur qui entend ou rencontre un DC se révèle également actif et non passif, même si cela est moins évident que dans le cas des démonstratifs gestuels. Un DC oral invite l'interlocuteur à « écouter » ce qui va suivre pour trouver le référent à apparier avec le DC entendu, de même qu'un DC écrit invite le lecteur à « regarder »<sup>21</sup> ce qui va suivre dans le discours pour découvrir le subséquent annoncé. De là, provient l'expressivité<sup>22</sup> – ou, si l'on veut, le « suspens » – inhérent aux DC, qui apparaît clairement lorsqu'on compare une structure de discours direct comme (21) à la structure correspondante avec DC (22) :

(21) *Il m'a dit* : « P »

(22) *Il m'a dit ceci* : « P »

## 6. Conclusion

Tous les problèmes que posent les DC ne sont évidemment pas réglés. Il nous semble toutefois que nous avons bouclé la boucle que nous avons tracée dans notre introduction : celle de montrer que, plutôt que d'opposer directement –

---

<sup>21</sup> On retrouve là l'image du « chien d'arrêt » qui invite le chasseur à « regarder », et non évidemment à « voir », pour trouver le gibier flairé ou entrevu par le chien.

<sup>22</sup> Cf. à ce propos le récent numéro de *Langages* (dirigé par Bordas, 2022a, n°228) consacré à *La notion d'expressivité*, sa présentation (Bordas, 2022b) et l'article *Expressivité de la cataphore* de Bordas (à paraître).

analyse classique exposée dans notre première partie – les DC aux démonstratifs anaphoriques, il était plus pertinent et, partant, plus fructueux, de les comparer directement aux démonstratifs gestuels. Les deux emplois de démonstratifs correspondent en effet, comme nous l'avons montré dans la deuxième partie, à la même situation cognitive, celle de l'introduction d'un référent nouveau dans la mémoire discursive. De cette parenté cognitive découle une hybridité sémiotique commune et, partant, comme l'ont montré nos troisième et quatrième parties, un fonctionnement référentiel, que nous avons détaillé en neuf points, qui s'avère grandement identique, les différences observées étant directement dues à la différence de nature de l'index.

## Références citées

- Achard, M. (2010). Le sens cataphorique de *ceci, cela* et *ça*. Dans F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada, S. Prévost (éds.), *Congrès Mondial de Linguistique Française*. Institut de Linguistique Française Sémantique.
- Authier-Revuz, J. (2021). *La représentation du discours autre. Principes pour une description*. De Gruyter.
- Berrendonner, A. (1983). Connecteurs pragmatiques et anaphore. *Cahiers de linguistique française* 5, 215–246.
- Bordas, E. (éd.) (2022a). La notion d'expressivité. *Langages* 228, 120.
- Bordas, E. (2022b). La notion d'expressivité. Présentation. *Langages* 228, 7–24.
- Bordas, E. (2023). Expressivité de la cataphore. *Linguisticae Investigationes* 46(1), 133–146.
- Cornish, F. (2001). Modal' *that* as determiner and pronoun: The primacy of the cognitive-interactive dimension. *English Language and Linguistics* 5(2), 297–315.
- Cornish, F. (2007). English Demonstratives: Discourse deixis and anaphora. A discourse-pragmatic account. Dans R. A. Nilsen, N. A. Amfo & K. Borthen (éds), *Interpreting Utterances: Pragmatics and its interfaces. Essays in honour of Thorstein Fretheim* (137–156). Novus Press.
- Cornish, F. (2011). Strict anadeixis, discourse deixis and text structuring. *Language Sciences* 33(5), 753–67.
- Cornish, F. (2017). SN démonstratifs et anadeixis : sens « spatial », ou valeurs tributaires d'une stratégie pragmatique potentielle ?. *Journal of French Language Studies* 27(2), 215–240.
- Cornish, F. (éd) (2021a). L'anaphore revisitée. *Langue Française* 210.

- Cornish, F. (2021b). Lanaphore revisitée : fonctionnements discursifs et interactionnels. *Langue Française* 210, 9–20.
- Danell, K.-J. (1990). Notes sur la concurrence entre *ceci* et *cela* en français moderne. *Studia Neophilologica* 62, 195–212.
- De Mulder, W. (2021). Le déterminant démonstratif *ce* : d'un marqueur *token*-réflexif à une instruction contribuant à la construction de référents. *Langue Française* 210, 41–58.
- Diessel, H. (1999). *Demonstratives: Form, Function, and Grammaticalization*. John Benjamins.
- Diessel, H. & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology* 11, 31–58.
- Dixon, R. M. W. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language* 27(1), 61–112.
- Doran, R. B. & Ward, G. (2019). A taxonomy of uses of demonstratives. Dans J. Gundel & B. Abbott (éds), *The Oxford Handbook of reference* (236–259). Oxford University Press.
- Fraser, T. & Joly, A. (1979). Le système de la deixis. Esquisse d'une théorie d'expression en anglais. *Modalités linguistiques* 1, 97–157.
- Fraser, T. & Joly, A. (1980). Le système de la deixis. Esquisse d'une théorie d'expression en anglais. *Modalités linguistiques* 2, 22–51.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. Routledge.
- Härmä, J. (1987). La cataphore pronominale français. *Neophilologica Fennica* 45, 53–79.
- Himmelman, N. P. (1996). Demonstratives in narrative discourse: A taxonomy of universal uses. Dans B. A. Fox (éd.), *Studies in anaphora* (205–54). John Benjamins Publishing.
- Kesik, M. (1989). *La cataphore*. PUF.
- Kesik, M. (1991). Contre la phore, tout contre.... *Travaux de linguistique* 22, 127–129.
- Kesik, M. (2014). Cataphore / anaphore : complémentarité référentielle, sémantique et syntaxique. Dans R. Daval, P. Frath, E. Hilgert & S. Palma (éds), *Les théories du sens et de la référence. Hommage à Georges Kleiber* (567–578). EPURE.
- Kleiber, G. (2004a). Sémantique, référence et discours : le cas des démonstratifs cataphoriques spécifiques. Dans A. Auchlin *et al.* (éds), *Structures et discours. Mélanges offerts à Eddy Roulet* (231–245). Éditions Nota bene.
- Kleiber, G. (2004b). Anticipation, mémoire et démonstratifs cataphoriques. Dans R. Sock & B. Vaxelaire (éds), *L'anticipation à l'horizon du Présent* (221–236). Pierre Mardaga Éditeur.
- Kleiber, G. (2005). Des démonstratifs bien énigmatiques : les démonstratifs cataphoriques génériques. Dans C. Dobrovie-Sorin (éd.), *Noms nus et généricité* (65–95). Presses Universitaires de Vincennes.


- Kleiber, G. (2006). Référence et prédication : le cas des démonstratifs cataphoriques. Dans G. Engwall (éd.), *Construction, acquisition et communication. Études linguistiques de discours contemporain, Romanica Stockholmiensia 23* (123–148). Acta Universitatis Stockholmiensis.
- Kleiber, G. (2007). Des démonstratifs *mémoriels* aux démonstratifs de *point de vue*. Dans L. Begioni & C. Muller (éds), *Problèmes de sémantique et de syntaxe. Hommage à André Rousseau* (373–392). Éditions du conseil Scientifique de l'Université Charles de Gaulle-Lille 3.
- Kleiber, G. (2023a). Cataphores démonstratives : construction et subséquent. *Linguisticae Investigationes* 46(1), 41–61.
- Kleiber, G. (2023b). L'énigme ontologique des démonstratifs cataphoriques. *Acta Linguistica Hafniensi* 55(1), 1–15.
- Lerat, P. (1983). *Sémantique descriptive*. Hachette.
- Lyons, J. (1980). *Sémantique linguistique*. Larousse.
- Maass, C. (2010). *Diskursdeixis im Französischen. Eine korpusbasierte Studie zu Semantik und Pragmatik diskursdeiktischer Verweise*. De Gruyter.
- Maes, A., Krahmer, E. & Peeters, D. (2022). Understanding Demonstrative Reference in text : a new taxonomy based on a new corpus. *Language and Cognition, mars 2022*, 1–23.
- Maillard, M. (1974). Essai de typologie des substituts diaphoriques. *Langue Française* 21, 55–71.
- Maillard, M. (1989). *Comment ÇA fonctionne*. Thèse d'État, Université de Paris 10–Nanterre.
- Peeters, D., Krahmer, E. & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review* 28(2), 409–433.
- Peeters, D. & Özyürek, A. (2016). *This and that revisited: A social and multimodal approach to spatial demonstratives*. *Frontiers in Psychology* 7, 222.
- Perdicoyanni-Paléologou, H. (2001). Le concept d'anaphore, de cataphore et de deixis en linguistique française. *Revue Québécoise de linguistique* 29(2), 55–77.
- Quine, W. V. O. (1971). The Inscrutability of Reference. Dans D. Steinberg & L. A. Jakobovits (éds), *Semantics: an Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (142–156). Cambridge University Press.
- Reichler-Béguelin, M.-J. (1988). Anaphore, cataphore et mémoire discursive. *Pratiques* 57, 15–43.
- Sidnell, J. & Enfield, N. J. (2017). Deixis and the interactional foundations of reference. Dans Y. Huang (éd.), *The Oxford Handbook of Pragmatics* (217–239). Oxford University Press.

- Theissen, A. (2008). *Ceci* en emploi cataphorique. Dans O. Bertrand, S. Prévost, M. Charolles, J. François & C. Schnedecker (éds), *Discours, diachronie, stylistique du français* (159–173). Peter Lang.
- Vanderbauwhede, G. (2012a). *Le déterminant démonstratif en français et en néerlandais*. Peter Lang.
- Vanderbauwhede, G. (2012b). La référence démonstrative dans les corpus écrits : théorisation et analyse de données empiriques. *Congrès mondial de linguistique française CMLF*.
- Wilmet, M. (1986). *La détermination nominale*. PUF.



*Denis Le Pesant*

Professeur émérite, Université Paris Nanterre  
France

 <https://orcid.org/0000-0003-2391-4887>

## Gaston Gross, héritier et continuateur de Maurice Gross et de Zellig Harris

**Gaston Gross, heir and successor of Maurice Gross and Zellig Harris**

### Abstract

This article intends to situate as precisely as possible the work of Gaston Gross and his team within what can be called “The School of Distributional and Transformational Grammars”, whose main representatives are Zellig S. Harris, Maurice Gross, Jean Dubois and Françoise Dubois-Charlier, and Gaston Gross. After discussing the collaboration of Maurice Gross with Zellig Harris in the process of developing the theory of Distributional and Transformational Grammars, we will highlight what Gaston Gross borrows from Maurice Gross and what he innovates, even going so far as to address an aspect of Harris’ original project that had been left out by Maurice Gross.

### Keywords

Gaston Gross, history of Distributional and Transformational Grammars, Zellig S. Harris, Maurice Gross, Jean Dubois, Françoise Dubois-Charlier

## 1. Introduction

Gaston Gross (1939–2022) fit ses études supérieures à Strasbourg, où il reçut l’influence du linguiste et statisticien Charles Muller (1909–2015). Il travaille ensuite au LADL (Laboratoire d’Automatique Documentaire et Linguistique, CNRS et Université Paris 7) dirigé par Maurice Gross depuis 1970. Sa thèse de doctorat (1987) traite des constructions converses du français (ex. *donner vs recevoir* une gifle, un conseil, etc.); elle montre (cf. G. Gross, 1989) que la



diathèse de passivation n'est pas la propriété exclusive de certains verbes. Cette thèse inaugure une longue série de travaux de Gaston Gross qui visent à mettre en évidence l'étendue considérable du vocabulaire des *verbes supports* de noms. Recruté la même année comme professeur à l'Université Paris 13 (Villetaneuse), il y fonde, en collaboration avec Mathieu-Colas, le LLI (Laboratoire de Linguistique Informatique, CNRS et Université Paris 13). Les travaux de cette équipe, qu'il va diriger pendant une quinzaine d'années, figurent dans la continuité de ceux du LADL<sup>1</sup>. Le LLI, bien doté en locaux et en personnel, va être, comme c'était aussi le cas du LADL, très tourné vers l'international : un nombre impressionnant de thèses furent préparées et soutenues à Villetaneuse, des dizaines d'universitaires du monde entier furent invités dans les locaux de l'université, et Gaston Gross multiplia les coopérations avec des universités de tous pays<sup>2</sup>.

Cet article entend situer le plus précisément possible les travaux de Gaston Gross et de son équipe au sein de ce qu'on peut appeler « L'École des Grammaires Distributionnelles et Transformationnelles », dont les principaux représentants sont Zellig Harris, Maurice Gross, Jean Dubois & Françoise Dubois-Charlier, et Gaston Gross.

Esquissons à grands traits la filiation du linguiste américain Zellig S. Harris (1909–1992), qui pose les fondements d'une théorie générale des Grammaires Distributionnelles et Transformationnelles. Maurice Gross, avec ses collaborateurs, enrichit la théorie en lexicalisant les grammaires et en les rendant, de ce fait, disponibles pour le TAL (Traitement Automatique du Langage). Maurice Gross aura à son tour deux héritiers et continuateurs.

Le premier héritier et continuateur de Maurice Gross est Jean Dubois (1920–2015), qui intervient puissamment dès les années 1960, dans un environnement

---

<sup>1</sup> Le fait que les travaux du LLI s'inscrivent dans la continuité de ceux du LADL était une évidence pour la direction du CNRS des années 1990. Entre 1993 et 1999 environ, le LADL, basé jusqu'alors à l'Université Paris 7 (Jussieu), et le LLI, qui venait d'être fondé, furent administrativement fusionnés au sein de l'URA n°1576, basée à l'Université Paris 13 (Villetaneuse). En témoignent au moins les faits suivants : a) dans le n° 59 de *L'Information Grammaticale* (année 1993), deux auteurs revendiquent l'appartenance à l'URA n° 1576 en ajoutant la mention « LADL & LLI » ; b) dans le n° 136 de *Langages* (année 1999), un des auteurs revendique l'appartenance à l'URA n°1576 en ajoutant également la mention « LADL & LLI ». Cela étant dit, cette fusion fut l'occasion d'une scission scientifique entre les deux équipes, la plupart des chercheurs du LADL ayant choisi, à la suite de Maurice Gross, de se regrouper au sein de l'Institut Gaspard Monge localisé à l'Université de Marne-la-Vallée plutôt que de rejoindre Villetaneuse.

<sup>2</sup> L'auteur de ces lignes, après avoir soutenu sa thèse au LADL sous la direction de Maurice Gross et Jean-Roger Vergnaud, a eu le privilège de faire partie du LLI de 1994 à 2001 et d'y être Habilité à Diriger des Recherches sous le patronage de Gaston Gross.

intellectuel français favorable aux travaux linguistiques, pour faire connaître les syntacticiens américains<sup>3</sup>, dont Harris qu'il prise particulièrement. D'autre part, sans faire partie du personnel du LADL<sup>4</sup>, il participe aux activités du laboratoire et, avec son épouse Françoise Dubois-Charlier (1941–2016), publie au début des années 1990 l'extraordinaire dictionnaire informatisé *Les Verbes Français* (LVF) (cf. Dubois & Dubois-Charlier, 1993a) qui vérifie la thèse harrienne de l'homomorphisme de la syntaxe et de la sémantique. Nous ne nous étendons pas dans cet article sur l'œuvre de Dubois et Dubois-Charlier parce que cette dernière n'a exercé aucune influence sur le parcours de Gaston Gross<sup>5</sup>.

Le deuxième héritier et continuateur de Maurice Gross est Gaston Gross qui, avec ses collaborateurs du LLI, s'attache à compléter les résultats du LADL dans les domaines alors encore inexplorés des noms non prédicatifs et des prédicats connecteurs, et à approfondir l'étude des verbes supports et des expressions figées du français. Après avoir évoqué la collaboration de Maurice Gross avec Zellig Harris dans le processus d'élaboration de la théorie des Grammaires Distributionnelles et Transformationnelles, nous mettrons en évidence ce que Gaston Gross emprunte à Maurice Gross et ce en quoi il innove, allant même jusqu'à aborder un aspect du projet initial de Harris qui avait été laissé de côté par le LADL.

## **2. Un épisode important du processus d'élaboration de la théorie des Grammaires Distributionnelles et Transformationnelles : la collaboration de Zellig Harris et de Maurice Gross dans les années 1960**

Polytechnicien (promotion 1955), puis ingénieur de l'armement affecté à un Centre d'Etudes pour la Traduction Automatique dépendant à la fois de l'armée

---

<sup>3</sup> Sur le rôle de « passeur » de Jean Dubois, on se reportera à l'article de Léon (2020).

<sup>4</sup> Jean Dubois est alors professeur à l'Université de Nanterre, où il fonde le département de linguistique. Son action s'exerce également dans l'édition scientifique : il est le fondateur et animateur principal de deux revues encore aujourd'hui très influentes, *Langages* et *Langue Française*. Sur l'œuvre de Dubois et Dubois-Charlier, nous renvoyons le lecteur au n° 80 (2020) de la revue *Linx* sur *L'héritage de Jean Dubois et Françoise Dubois-Charlier* (cf. Leonarduzzi, 2020).

<sup>5</sup> Elle n'a exercé aucune influence sur Gaston Gross à ce point que ce dernier n'a jamais, à notre connaissance, polémique contre le choix surprenant de Dubois et Dubois-Charlier dans *LVF* (1993) de ne pas attribuer à une multitude d'emplois verbaux leur nature de verbe support (cf. Le Pesant, 2020, § 6.3).

et du CNRS, Maurice Gross est jusque 1960 un mathématicien de haut niveau au fait des progrès les plus récents de l'Informatique Théorique. Les recherches sur la traduction automatique étant progressivement abandonnées tant aux USA qu'en Europe à la suite du rapport Bar Hillel (1959), il s'en détourne pour se concentrer sur l'informatique fondamentale et se spécialiser dans la théorie des grammaires formelles. En 1961–1962, un premier séjour aux USA conduit M. Gross au MIT, dans un cercle dont Chomsky est le centre<sup>6</sup>. C'est là qu'il fait la connaissance de Marcel-Paul Schützenberger, co-auteur du théorème de Chomsky-Schützenberger, qui deviendra son ami le plus proche<sup>7</sup>. Retourné à Paris, il intègre l'Institut Blaise Pascal, centre de calcul du CNRS de l'époque, et prépare, avec le mathématicien Lentin, un livre intitulé *Notions sur le grammaires formelles* (cf. M. Gross & Lentin, 1967).

En 1964–1965, sur la recommandation de Schützenberger, Maurice Gross est invité par Zellig Harris à Philadelphie, qui, comme ç'avait été le cas peu de temps auparavant pour Chomsky, a besoin d'un mathématicien susceptible de l'aider à formaliser ses grammaires. Comme l'écrivent Chevalier et Encrevé (2006 : 259) : « Pour Harris, qui a envie de formaliser un peu plus ce qu'il fait, Gross sera l'ange providentiel ; pour Gross, ce sera le contact permanent avec un grammairien éprouvé ». C'est pendant ce séjour d'un an que Gross non seulement devient, selon ses dires, un véritable linguiste (grâce aux « cours particuliers » de Harris), mais aussi conçoit le projet du *lexique-grammaire*. La collaboration des deux chercheurs se poursuit jusqu'à la fondation du LADL en 1970, puis la publication par Gross de *Méthodes en syntaxe* en 1975 et celle de *Notes du cours de syntaxe* en 1976, qui est la traduction-transcription par Gross lui-même du cours professé à Paris en anglais par Harris à l'université de Paris-Vincennes en 1973–1974.

---

<sup>6</sup> Gross ne se contente pas d'apprendre des Américains. D'après ce qu'il déclare à Chevalier & Encrevé (2006 : 256), il fait un jour au MIT un exposé sur Tesnière (1959). Or, c'est sur la lecture de Tesnière que se fonde un des premiers articles de Gross (Gross : 1962), où on lit ces lignes : « We will turn to the dependency grammars as defined in Hays (...). The linguistic conception originated by Tesnière differs from that of Immediate Constituent Analysis; here the morphemes are connected in terms of the intuitive notions of governor and dependent ». L'analyse en constituants immédiats à laquelle Chomsky (1957) adhère est ainsi opposée aux *grammaires de dépendance*, qui ont la préférence non seulement de Tesnière et de Hays, mais aussi de Harris et de Gross. Cela étant dit, rien ne permet, à notre connaissance, d'en déduire qu'Harris et Tesnière ont pu s'influencer mutuellement. On reviendra rapidement sur ce point *infra* au § 2.2.

<sup>7</sup> Schützenberger (1920–1996) fut une sorte de génie universel : membre de l'Académie des sciences, d'abord médecin psychiatre puis épidémiologiste, statisticien, professeur de mathématiques aux facultés des sciences de Poitiers puis de Paris, rangé par Lichnerowicz dans un article de la revue *La Recherche* (1996) parmi les trois ou quatre fondateurs de l'informatique théorique, et par-dessus tout selon Perrot (1996), mathématicien algébriste.

C'est aussi dans cette période que paraît la traduction en français, par Catherine Fuchs (Harris, 1971), de *Mathematical Structures of Language* (cf. Harris, 1968).

Le dépassement de Harris par Maurice Gross fut d'ordre méthodologique<sup>8</sup> : il entreprit de lexicaliser les règles des grammaires de Harris. Le projet de construction du lexique-grammaire est issu du constat que les grammaires de Zellig Harris ont l'avantage d'être axiomatisées (cf. la préface de *Notes du cours de syntaxe*, M. Gross, 1976), mais que les règles ont le défaut d'être privées de leur *extension lexicale*. L'extension lexicale d'une règle, c'est l'énumération de la totalité des faits qu'elle « prédit ». Soit la distribution « N0 qq V Prép. à N1qq N2 (que P<subj> / de V-er) », qui est celle par exemple du verbe *conseiller*, et érigeons-la en règle de construction syntaxique. La définition *en extension* de cette règle équivaut à l'énumération des verbes qui ont la même distribution que *conseiller* ; il y en a plus de 40 (*commander, demander, ordonner, recommander*, etc.), dont la parenté sémantique est évidente.

En effet, la lexicalisation des règles met en évidence le phénomène d'*homomorphisme* de la syntaxe et de la sémantique. Plusieurs passages de *Méthodes en syntaxe* (1975) évoquent ce phénomène, ainsi que l'article essentiel de Maurice Gross intitulé, de façon caractéristique, *Les bases empiriques de la notion de prédicat sémantique* (Gross, 1981 : 20–21)<sup>9</sup>. Et c'est à Dubois et Dubois-Charlier (1993b), grâce à leur classification syntactico-sémantique de quelque 26000 entrées verbales, que revint le mérite de démontrer la validité du principe d'homomorphisme<sup>10</sup>.

### 3. Gaston Gross, héritier de Maurice Gross

Dans cette section, nous étudions les deux principaux aspects théoriques que Gaston Gross emprunte, avec son équipe du LLI, aux travaux du LADL.

<sup>8</sup> Sur la personne et l'œuvre de Maurice Gross, on se reportera au bel article de Lamiroy (2003).

<sup>9</sup> Les questions de sémantique furent aussi la principale préoccupation de Harris à la fin de sa vie (cf. Harris, 1988).

<sup>10</sup> Citons les premières lignes de *LVF* (1993b) : « La *classification des verbes français* repose sur l'hypothèse qu'il y a adéquation entre les schèmes syntaxiques de la langue et l'interprétation sémantique qu'en font les locuteurs de cette langue : à la différence syntaxique entre *venir à Paris* et *venir de Paris* correspond une différence sémantique entre la destination et l'origine (Dubois & Dubois-Charlier, 1993 (b), p. III, *Objectifs et méthode*).

### 3.1. Référence au modèle logique frégéen de la proposition et choix du cadre des *grammaires de dépendance*

En logique, le modèle de la proposition qui a prévalu jusqu'à la fin du 19<sup>e</sup> siècle remonte à l'antiquité (Aristote, Porphyre). Il est fondé sur une théorie du jugement, qui oppose le *Sujet* (c'est-à-dire le thème de la proposition) au *Prédicat* (c'est-à-dire l'énoncé de ce qu'on a à dire à propos du thème). Exemple :

(1)a    Sujet      Prédicat (ou Attribut)  
           *Socrate*    *interroge Cratyle*

C'est ce modèle logique qui inspire les *grammaires de constituants immédiats*. Ces dernières instituent la structure fondamentale suivante, homologue de (1)a :

(1)b    Groupe sujet    Groupe verbal  
           *Socrate*            *interroge Cratyle*

Ce modèle grammatical est dominant dans la tradition scolaire et il caractérise aussi, on l'a dit, les *grammaires de Chomsky*.

Aux *grammaires de constituants immédiats* s'opposent les *grammaires de dépendance* (*Dependency Grammars*), fondées sur les travaux du mathématicien logicien Frege (1848–1925), qui propose un modèle nouveau de la proposition en termes de relation prédicat-arguments. Le *prédicat* (ou *fonction*) est l'élément « insaturé » (ungesättigt, ou « incomplet, en besoin de supplément ») de la proposition, le supplément étant l'ensemble des *arguments* (ou *variables*) qui saturent le prédicat<sup>11</sup>. Par exemple, dans ce cadre théorique homologue de celui qui caractérise les analyses (1)a et (1)b, le prédicat *interroger* prend pour arguments *Socrate* et *Cratyle* dans la structure suivante :

(2)    Prédicat      argument 1      argument 2  
           *interroger*    *Socrate*            *Cratyle*

C'est le modèle des *grammaires de dépendance* que Maurice Gross emprunte tant à Tesnière (1959), auteur d'une théorie de la *valence*, qu'à Z. Harris<sup>12</sup>. Les gram-

<sup>11</sup> Cf. l'article de Frege *Funktion und Begriff* (1891), trad. dans *Ecrits logiques et philosophiques*, Frege (1971 : 80–101).

<sup>12</sup> M. Gross indique à Chevalier & Encrevé (2006 : 256) avoir lu Tesnière à Paris avant son séjour de 1962 au MIT (sur son exposé sur Tesnière au MIT et son article de 1962, Cf. *supra* la note 6). Il est donc probable que sa découverte des écrits de Harris à cette époque n'a fait que confirmer son

maires de Gaston Gross et de ses collaborateurs sont aussi des *grammaires de dépendance*, qui ont pour caractéristique la plus évidente d'éliminer la notion de Groupe Verbal.

### 3.2. Distributions et Transformations

Les grammaires distributionnelles et transformationnelles stipulent que la bonne formation d'une phrase dépend d'une part de règles syntaxiques de construction (la structure argumentale) et de variations de forme (les transformations), d'autre part de conditions sémantiques sur la sélection lexicale. L'application des règles de construction et de transformation produit un vaste ensemble de variantes que Maurice Gross (1990 : 44) appelle *classe d'équivalences* laissant invariante la structure argumentale. Tous ces principes sont adoptés par Gaston Gross, même si les expressions de « classe d'équivalences » et de « structure argumentale » ne font pas partie de la terminologie du LLI.

Par exemple le verbe *lire*, dans un de ses emplois, est un verbe transitif direct à argument sujet humain agent d'un processus, et à argument complément nominal appartenant à la classe des documents écrits. La réalisation de cette structure correspond à l'invariant qui vient d'être évoqué. Des transformations telles que la passivation, les différents types de pronominalisation, de permutation, de réduction, de nominalisation, etc., ainsi que la plupart des opérations de morphologie dérivationnelle et d'affixation, et les adjonctions de verbes auxiliaires (notamment les verbes-supports de noms et d'adjectifs), permettent de construire une vaste classe d'équivalences dont voici un court extrait non structuré :

(3)

Quelqu'un **lit** un/des roman(s)

Marie lit des romans, elle les lit, elle en lit un, elle lit ; qui lit quoi ? qu'est-ce qu'elle lit ? elle en lit combien, etc.

elle le lit, ce roman ; ce roman, elle ne l'a jamais (re)lu ; etc. ; elle commence à le lire, elle continue à lire, etc.

combien de romans ont été lus par Marie ? il en a été lu beaucoup par elle, etc.

ces magazines ont été lus par elle, ils n'ont pas encore été lus, ils sont illisibles, etc.

il a été fait une lecture (approfondie) de ces romans par Marie ; elle en a (fait,

---

attirance pour le modèle des *grammaires de dépendance*. Sur la question controversée de savoir si le slavisant Tesnière a pu lui-même, dans les années 1930, emprunter sa théorie de la valence aux membres du Cercle de Vienne ou à des linguistes ou logiciens d'Europe orientale et d'Union Soviétique, ou s'il l'a conçue en tant que novateur isolé, on se reportera à Sériot (2004).

accompli, commencé ...) une lecture approfondie etc. ;  
 (voici) les romans qu'elle a lus ; etc.  
 elle fait lire ce roman à ses élèves de 4<sup>ème</sup>, elle les laisse lire ce roman, etc.  
 etc.

(M. Gross, 1990 : 44)

## 4. Gaston Gross, continuateur de Maurice Gross et de Zellig Harris

On aborde ici quatre domaines dans lesquels Gaston Gross innove dans le processus d'élaboration de la théorie des Grammaires Distributionnelles et Transformationnelles, dans la continuité des œuvres de Maurice Gross et de Zellig Harris.

### 4.1. Ouverture de la théorie au traitement des noms non-prédicatifs : les *classes d'objets*

Maurice Gross (1990 : 49) énumère une liste de travaux de chercheurs du LADL ayant étudié les noms prédicatifs, c'est-à-dire les noms pourvus d'une structure argumentale qui, pour se réaliser en phrases, ont besoin de s'associer à un ou plusieurs verbes auxiliaires appelés *verbes supports*. Ainsi sont évoqués entre autres les noms à V-sup *faire* (Jacqueline Giry-Schneider), *donner-recevoir* (Gaston Gross), *avoir* (Jacques Labelle), *être en* (Dominique de Négroni), *avoir, prendre, perdre* (Robert Vivès). À la suite de ce passage, il observe :

(4)

Seuls semblent échapper à cette description les substantifs **concrets** (e.g. *chaise, notaire, homme des bois, pomme de terre*), qui ne donnent pas lieu à des phrases élémentaires syntaxiquement significatives, et présentent d'autres problèmes de représentation.

(M. Gross, 1990 : 49)

Maurice Gross indique donc en substance que le traitement des noms non prédicatifs constitue un défi pour l'analyse distributionnelle et transformationnelle.

Or ce défi est relevé avec succès par Gaston Gross et ses collaborateurs. Pour faire entrer le traitement des noms non prédicatifs dans le cadre de la théorie,

il suffit de les considérer pour ce qu'ils sont au sein de toute proposition, à savoir nécessairement des arguments, et de déterminer de quels *prédicats* verbaux et adjectivaux ils sont spécifiquement les arguments. Soit par exemple la classe intuitive des noms d'*aliments*. Le LLI propose de la définir selon le critère de leurs *prédicats appropriés*, c'est-à-dire les classes de prédicats qui peuvent de façon spécifique les prendre comme arguments, tels les verbes de *préparation culinaire* (ex. *assaisonner, cuisiner, frire*) et de *consommation* (ex. *avaler, déguster, digérer, manger*), ainsi que des adjectifs comme *avarié, délicieux, immangeable, indigeste*. Ces classes de prédicats appropriés peuvent à leur tour être regroupées par catégories sémantiques, comme on le constate par les quelques exemples qui viennent d'être donnés.

Gaston Gross et Mathieu-Colas convinrent d'appeler *classes d'objets* ces classes d'arguments et le développement de cette méthodologie devint rapidement la marque du LLI : la théorie des grammaires distributionnelles et transformationnelles ouvrait la voie, d'une façon simple et intuitive, à des travaux de classification sémantique pour la multitude des noms concrets : humains, animaux, lieux, outils, métiers, matériaux, artefacts, etc.

Du point de vue théorique, le modèle des *classes d'objets* constitue un enrichissement majeur apporté par Gaston Gross au programme des *grammaires distributionnelles et transformationnelles*. On trouvera chez Le Pesant et Mathieu-Colas (1998) une présentation d'ensemble de la question des classes d'objets. Dans le même numéro de la revue *Langages*, Mathieu-Colas (1998) développe de façon extensive l'exemple de la classe des noms de *voies* de communication.

#### **4.2. Ouverture de la théorie au traitement des prédicats circonstanciels prépositionnels, conjonctifs et adverbiaux : Gaston Gross héritier et continuateur de Zellig Harris**

Dans ses ouvrages sur l'expression de la finalité et de la cause, Gaston Gross aura été le premier à reprendre à son compte la proposition de Harris d'intégrer la plupart des prépositions, conjonctions et adverbes connecteurs (circonstanciels) dans la catégorie des prédicats. C'est un travail qui n'avait pas été réalisé par le LADL, bien qu'il fasse partie du programme élaboré par Harris, tel qu'il figure par exemple dans Harris (1976). Par exemple, dans un passage qui traite des prédicats à deux arguments non prédicatifs (Harris utilise le terme d'*Opérateur de type  $O_m$* ), cette construction est exemplifiée ainsi :



- (5)
- (a) manger : Max mange du poisson
  - (b) porter : Max porte des chapeaux
  - (c) près : Max est près de la maison
  - (d) père : Max est le père de Franck
- (Harris, 1976 : 64)

On constate que Harris verse dans la même catégorie (celle des prédicats à deux arguments nominaux) la locution relationnelle *être le père de*, les verbes transitifs directs *manger* et *porter*, et la préposition *près de* accompagnée du verbe support *être*. Dans un autre passage du même ouvrage, Harris (1976 : 76–77) traite de l'Opérateur  $O_{on}$ , dont il dit que, dans de nombreuses analyses, « il prend la forme d'une préposition (...), il relie alors son premier argument (qui est lui-même un « opérateur ») à des noms non prédicatifs de temps, de lieu, de quantité, de manière, ou d'autres propriétés répandues ». Suivent quelques exemples dont ceux-ci :

- (6)
- (a) throughout a period (pendant une période)
  - (b) since morning (depuis le matin)
  - (c) near the road (près de la route)
- (Harris, 1976 : 77)

Ces analyses ont inspiré à Gaston Gross les ouvrages suivants, en collaboration : sur la *finalité* (2004) co-écrit avec Prandi et sur la *sémantique de la cause* (2009) co-écrit avec Pauna et Valetopoulos. Enfin, Gaston Gross s'étend sur l'analyse des prédicats connecteurs circonstanciels dans les Chapitres 15 et 16 de son *Manuel d'analyse linguistique* (G. Gross, 2012).

#### 4.3. Intuition d'une extension très large du vocabulaire des verbes supports

L'intérêt de Gaston Gross pour la question des verbes supports a déjà été mentionné plus haut. Après sa thèse sur les verbes supports de conversion de type *donner-recevoir*, il n'a jamais cessé de réfléchir sur les verbes supports et ses publications dans ce domaine sont nombreuses (cf. entre autres G. Gross 1989, 1993, 2017, ainsi que G. Gross & de Pontonx, 2004). Rappelons que les verbes supports sont des verbes auxiliaires qui permettent aux adjectifs qualificatifs et aux noms prédicatifs de se réaliser en phrases, comme le font les verbes prédicatifs. Par

exemple, les verbes auxiliaires *être* et *rester* sont des verbes supports d'adjectifs, et les verbes *faire*, *donner*, *prendre* sont des verbes supports de noms dans les phrases, respectivement, « Pierre (*est*, *reste*) malade », « Marie *fait* un voyage », « elle *donne* des conseils à Paul », « elle *prend* la décision de partir ».

Ce phénomène est reconnu depuis longtemps tant dans les grammaires de Harris que dans celles de Chomsky, mais reste sous-évalué quant à son importance numérique. Les recherches de Gaston Gross sur les verbes supports, restées inachevées, montrent qu'ils sont en français au nombre de 300 au moins. À la fin de sa vie, il espérait avoir le temps de rédiger un grand ouvrage sur ce domaine et, au cours d'une conversation tenue en juin 2022, il nous avait confié son intuition qu'on devrait dénombrer les verbes supports plutôt par milliers que par centaines. Pour donner une ombre de vraisemblance à cette intuition, nous allons prendre deux exemples.

Soit d'abord l'exemple des verbes *commencer*, *continuer*, *terminer*, *finir* dont la nature de verbes supports aspectuels des noms n'est contestée, croyons-nous, par personne. Exemples :

(7)a

Les techniciens *font* le **nettoyage** d'un local

Les techniciens (*commencent*, *continuent*, *terminent*, *finissent*) le **nettoyage** d'un local

Or il y a près de trois fois plus de verbes auxiliaires de ce type, comme le montre la liste complémentaire suivante :

(7)b

attaquer, débiter, démarrer, entamer, recommencer, reprendre, se mettre à ; poursuivre ; achever, arrêter, cesser interrompre, terminer, parachever, mettre fin à

Il est à noter que nombre de ces verbes auxiliaires ont des dérivés nominaux. On devrait en conclure l'existence d'une catégorie de « noms-supports ». Exemples :

(8)

J'ai assisté au **nettoyage** d'un local

J'ai assisté au/à la (*commencement*, *début*, *démarrage*, *reprise* ; *continuation*, *poursuite* ; *achèvement*, *arrêt*, *interruption*, *fin*) du **nettoyage** d'un local

Pour conforter l'intuition de Gaston Gross que les mots auxiliaires doivent se compter par milliers plutôt que par centaines, prenons maintenant l'exemple

des noms de sentiment et énumérons une partie de leurs verbes supports appropriés. Les plus généraux sont bien connus : *avoir*, *éprouver* et *ressentir*, ainsi que les locutions verbales *être plein de* et *être dans*. Mais il en existe encore beaucoup d'autres (au moins une quarantaine), appropriés à telle ou telle classe de noms de sentiment comme :

(9)a

*déborder de*, *être au comble de* (joie)

*être pétri de* (ambition, orgueil)

(se) *plonger dans* (mélancolie), *nager dans* (bonheur); *s'abîmer*, *sombrer dans* (désespoir);

*concevoir*, *nourrir*, *bercer*, *caresser* (ambition, espoir); *porter*, *vouer à qq* (amour, admiration)

*connaître*, *endurer*, *subir* (déception, souffrance); *être en proie à* (tristesse, mélancolie, dépression)

*brûler de*, *griller de* (désir, envie, impatience)

*y avoir entre x et y* (amitié, inimitié); (amitié) *lier x à y*; (inimitié) *opposer x à y*  
etc.

De plus, certains verbes supports, passivables pour la plupart, succèdent aux noms prédicatifs qu'ils supportent :

(9)b

(espoir, inquiétude, joie, tristesse) *habiter*, *prendre*, *gagner*, *envahir*, *submerger qq*;  
*s'emparer de qq*

(admiration, amour, joie) *transporter qq*

(douleur, regret, remords) *consumer*, *démanger*, *frapper*, *torturer*, *déchirer*, *écraser*,  
*êtreindre*, *oppresser*, *tarauder*, *tenailler*, *travailler qq*

(doute, soupçon) *effleurer*, *saisir*, *assaillir qq*

etc.

#### 4.4. Un critère supplémentaire pour la classification des noms prédicatifs : la prise en compte des verbes supports appropriés

Nous venons d'utiliser, avec l'exemple des verbes supports des noms de sentiment, la notion de *verbe support approprié*. Elle se révèle être une notion particulièrement utile dans l'entreprise de classification des noms prédicatifs. L'apport de Gaston Gross dans ce domaine est également considérable : son travail

montre qu'au sein de l'ensemble des propriétés syntaxiques des noms prédicatifs, la prise en compte des verbes supports joue un rôle crucial pour l'établissement de subdivisions fines dans les classes sémantiques, comme le suggèrent les exemples (9) ci-dessus. Prenons un dernier exemple d'application de la méthode. Soient les noms de *crimes et délits*. Leurs verbes supports appropriés généraux sont *commettre* et *se rendre coupable de*; parmi les noms de *délits*, les noms de *crimes* se distinguent par le fait de pouvoir bénéficier du verbe support approprié *perpétrer*.

## 5. Conclusion

Nous n'avons pas rendu compte, dans cet article, de la totalité des apports de Gaston Gross à la théorie des Grammaires Distributionnelles et Transformationnelles. Par exemple, nous n'avons pas évoqué ses travaux sur le figement (cf. notamment M. Gross, 1996) qui se situent dans la continuité de ceux, pionniers, de Maurice Gross (cf. M. Gross, 1990). Dans la dernière année de sa vie, Gaston Gross s'intéressait aux *phrases préfabriquées des interactions verbales* qui font l'objet d'une partie du numéro 114 de la revue dont il avait été longtemps le directeur, les *Cahiers de Lexicologie* (cf. Dostie & Tutin (éds), 2019), et quelques semaines avant sa mort, intervenue le 13 octobre 2022, il avait commencé une recherche à ce sujet en collaboration avec l'auteur de ces lignes. Regret inconsolable d'un collègue génial!

## Références citées

- Bar Hillel, Y. (1959). *Report on the State of Machine Translation in the United States and Great Britain States*. Hebrew University, Jerusalem.
- Chevalier, J.-C. & Encrevé, P. (2006). *Combats pour la linguistique*. ENS Éditions.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co.
- Dostie, G. & Tutin, A. (éds) (2019). Les phrases préfabriquées : sens, fonctions, usages. *Cahiers de lexicologie 114*. Classiques Garnier.
- Dubois, J. & Dubois-Charlier, F. (1993a). *Dictionnaire électronique des verbes français* (LVF). Tableau Excel, 26.610 entrées. Téléchargeable sur : [modyco.fr/fr/Ressources/](http://modyco.fr/fr/Ressources/)

- ldlvf.html (consulté le 12/06/2022). Version XML sur : <http://rali.iro.umontreal.ca/rali/?q=fr/lvf> (consulté le 09/09/2022).
- Dubois, J. & Dubois-Charlier, F. (1993b). *Les verbes français* (LVF). Larousse (épuisé). Téléchargeable sur : [modyco.fr/fr/Ressources/ldlvf.html](http://modyco.fr/fr/Ressources/ldlvf.html) (consulté le 12/06/2022).
- Frege, G. (1971). *Écrits logiques et philosophiques* (trad. de Claude Imbert). Éditions du Seuil.
- Gross, G. (1989). *Les constructions converses du français*. Librairie Droz.
- Gross, G. (1993). Trois applications de la notion de verbe support. *L'information grammaticale* 59, 6–22. Paris.
- Gross, G. (1996). *Les expressions figées en français. Noms composés et autres locutions*. Ophrys.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses Universitaires du Septentrion.
- Gross, G. (2017). Petit historique de la notion de verbes supports. *Cahiers de Lexicologie* 111, 121–144. Classiques Garnier.
- Gross, G. & de Pontonx, S. (éds) (2004). Les verbes supports, nouvel état des lieux. *Special Issue of Lingvisticae Investigationes* 27(2). John Benjamins.
- Gross, G. & Prandi, M. (2004). *La finalité. Fondements conceptuels et genèse linguistique*. Duculot.
- Gross, G., Pauna, R. & Valetopoulos, F. (2009). *Sémantique de la cause*. Peeters.
- Gross, M. (1962). On the equivalence of models of language used in the fields of mechanical translation and information retrieval. *NATO Summer School, Advanced Study Institute on Automatic Translation of Languages*, 15–31.
- Gross, M. (1975). *Méthodes en syntaxe*. Klincksieck.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 8–50. Larousse.
- Gross, M. (1990). « Sur la notion harrissienne de transformation et son application au français ». *Les grammaires de Harris et leurs questions*. *Langages* 99, 39–56. Larousse.
- Gross, M. & Lentin, A. (1967). *Notions sur les grammaires formelles*. Gauthier-Villars.
- Harris, Z. (1968). *Mathematical Structures of Language*. Wiley-interscience.
- Harris, Z. (1971). *Structures mathématiques du langage*. Dunod.
- Harris, Z. (1976). *Notes du cours de syntaxe*. Éditions du Seuil.
- Harris, Z. (1988). *Language and Information*. Columbia University Press.
- Lamiroy, B. (2003). Maurice Gross (1934–2001). *Travaux de linguistique* 46, 145–158.
- Leonarduzzi, L. (éd.). (2020). *L'héritage de Jean Dubois et Françoise Dubois-Charlier*. LINX 80.
- Léon, J. (2020). Jean Dubois, un passeur ? LINX 80.
- Le Pesant, D. (2020). Place de l'œuvre de Jean Dubois et de Françoise Dubois-Charlier dans l'histoire de la linguistique française. LINX 80.

- Le Pesant, D. & Mathieu-Colas, M. (1998). Introduction aux classes d'objets. *Langages* 131, 6–33.
- Lichnerowicz, A. (1996). *Marcel-Paul Schützenberger (1920–1996)*. *La Recherche* 291.
- Mathieu-Colas, M. (1998). Illustration d'une classe d'objets : les voies de communication. *Langages*, 131, 77–90.
- Perrot, J.-F. (1996). Marcel-Paul Schützenberger (1920–1996). *Hermès, La Revue* 20, 259– 260.
- Sériot, P. (2004). L'affaire du petit drame : filiation franco-russe ou communauté de pensée ? (Tesnière et Dmitrievskij). *Slavica Occitania* 17, 93–118.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.



**Igor Mel'čuk**

Observatoire de linguistique Sens-Texte  
Université de Montréal, Canada

 <https://orcid.org/0000-0002-4520-0554>

## Russian Free Phrasal Indefinite Pronouns

### Abstract

The paper describes Russian phrasal indefinite pronouns: both 1) phraseologized (= collocational), such as *koe-kto* ≈ 'someone', *Bog znaet kto* 'God knows who', *kto ugodno* 'whoever' or *kto by to ni bylo* 'no matter who', and 2) free, such as {*Ja vstretil*} *nikto iz tvoix družej ne dogadaetsja kogo* '{I met} none of your friends will guess whom' (a.k.a. *syntactic amalgams*). Three lexical entries are presented: for the indefinite pronominal lexeme *кто* ≈ '-body; -one', for the premodifying indefiniteness type marker idiom 'БОГ ЗНАЕТ' 'God knows', and for the postmodifying indefiniteness type marker particle *-то*<sup>4</sup> ≈ 'some'. Formal representations of collocational and free phrasal indefinite pronouns at three levels of linguistic representation (semantic, deep-syntactic and surface-syntactic) are given, as well as rules for constructing both types of pronouns.

### Keywords

Russian language, lexicology, indefinite pronouns, collocational and free phrasal pronouns, syntactic amalgams

1 Statement of the Paper's Goals .....	2
2 Russian Phraseologized (= Collocational) Indefinite Pronouns .....	4
2.1 The Structure and the Linguistic Nature of Russian Collocational Indefinite Pronouns .....	4
2.2 Three Lexical Entries for the Components of Russian Collocational Indefinite Pronouns .....	10
3 Russian Freely Constructed Phrasal Indefinite Pronouns .....	17
4 Not to Be Confounded .....	20
5 Conclusions .....	22
Acknowledgments .....	23
Appendix: Some Vital Notions .....	23
Abbreviations and Notations .....	25
References .....	26
Notes .....	27

To the dear memory of Gaston Gross

## 1 Statement of the Paper's Goals

The intended object of the present paper is the family of Russian expressions of the following form (boldfaced):

- (1) *Ja segodnja vstretil **daže Petina mama ne dogadaetsja kogo***  
 'I met today even Pete's mom won't guess whom.'

The boldfaced phrase is a phrasal indefinite pronoun, similar to such phrasal indefinite pronouns as КОЕ-КТО lit. 'some who' ≈ 'somebody' or КТО УГОДНО lit. 'who you want' ≈ 'anybody'.

☞ Vital notions, on the first mention, are printed in Helvetica; they are discussed in the body of the paper or briefly explained in the Appendix, p. 23.

Calling such an expression as *daže Petina mama ne dogadaetsja kogo* a pronoun is, of course, an *abus de langage*, or a "poetic licence": the expression in question is a pronominal phrase rather than simply a pronoun. However, this loose use of the term can be accepted as a convenient abbreviation, since in the given context it cannot entail any confusion.

Now, there is a striking difference between phrasal indefinite pronouns of the above two types:

- The indefinite pronouns of the *koe-kto* and *kto ugodno* type are phraseologized, that is, non-free, phrases, and their number is finite (as will be shown, about 360: see Table 1, p. 5); they must be listed and described in the lexicon.
- The indefinite pronouns of the *daže Petina mama ne dogadaetsja kogo* type are free phrases, and their number is infinite; they cannot be presented in the lexicon and must be specified by rules.

Therefore, the final goal of this paper is twofold: to propose

- 1) the formal representations of Russian free phrasal indefinite pronouns on the semantic [Sem-], deep-syntactic [DSynt-] and surface-syntactic [SSynt-] levels
- and
- 2) the formal rules relating these representations, that is, the rules that describe the production of these pronouns by the Speaker.



NB English freely produced phrasal indefinite pronouns of this type were brought to linguists' attention by George Lakoff (1974, pp. 321–324) under the name of *syntactic amalgams*.

But there is a complication: the syntactic head of a free-phrase pronoun in question is a pronominal lexeme, such as *кто* 'who', *что* 'what', *где* 'where', etc. Traditionally, these lexemes are called “interrogative pronouns,” yet in the role that interests us here, they are by no means interrogative: their meaning and their syntactic behavior are drastically different from those of interrogative pronouns. Informally, they are also known as WH-words (in Russian, as K-words)—a convenient name, which, however, does not reflect their notional status. Strictly speaking, what are they indeed? Unfortunately, the available literature does not give a satisfactory formal enough answer. Hence, the paper's preliminary goal is:

|| To propose a formalized description, even if a cursory and sketchy one, of the closed system of Russian indefinite pronouns—both the indefinite pronominal lexemes and the indefinite pronominal phrasemes.

The indefinite pronominal phrasemes are discussed in Section 2, and Section 3 is dedicated to the freely built phrasal indefinite pronouns.

Two remarks before I proceed.

- The literature on indefinite pronouns is too rich for a reasonable overview, but fortunately there is the fundamental monograph (Haspelmath, 1997), which offers a detailed and multifaceted characterization of indefinite pronouns in the world's languages; what follows is based to a great degree on this work. As far as Russian is concerned, one finds in the studies (Padučeva, 1985, pp. 209–220, 2015–2017, and 2018) a rigorous description of the main Russian indefinite pronouns, and I largely use Padučeva's results. A precise and compact description of Russian phrasal indefinite pronouns is offered in (Iomdin, 2010). Finally, recently the set of Russian indefinite pronominal lexemes has been semantically and combinatorially treated as a particular lexicographic type in (Apresjan, V. & Iomdin, B., 2022), so that the present paper is a contribution to the research endeavor launched by this article.
- What follows is but a pilot study—a very first step towards the declared goals. Therefore, everything is fairly approximative, especially the semantic characterization of pronominal expressions.

## 2 Russian Phraseologized (= Collocational) Indefinite Pronouns

Russian phraseologized indefinite pronouns will be dealt with in two stages: their structure and linguistic nature (2.1) and their lexicographic treatment (2.2).

### 2.1 The Structure and the Linguistic Nature of Russian Collocational Indefinite Pronouns

The Russian language has a (more or less) closed set of indefinite pronominal expressions having the following composition:

PREMODIFYING LEXICAL UNIT ← **pronominal-auxiliary** → INDEFINITE PRONOMINAL LEXEME

OR

INDEFINITE PRONOMINAL LEXEME ← **pronominal-auxiliary** → POSTMODIFYING LEXICAL UNIT

Thus, Russian phraseologized indefinite pronouns are two-component phrases, which are collocations (Mel'čuk, 2023b, pp. 112–136). The base of such a collocation is an indefinite pronominal lexeme, and the collocate is a lexical unit [LU] that is a marker of the indefiniteness type [IT-marker];<sup>1</sup> see Table 1, next page.

Let it be emphasized:

A Russian phraseologized indefinite pronoun such as КОЕ-КТО ≈ 'some-body', КТО-ТО ≈ 'some-body' or КТО-НИБУД' ≈ 'some-body' is not a wordform (as is often tacitly admitted), but a phrase; more specifically, it is a collocation, whose base is the nominal pronoun КТО 'who' and whose collocates are the particles КОЕ- ≈ 'some-', -ТО<sup>4</sup> ≈ 'some-' and -НИБУД' ≈ 'some-'.

To see this, consider the linear separation of the complex pronoun's components by a preposition (*koe s kem* lit. 'some with body'), the "intra-word inflection" (*k+ogo-to* lit. 'who+m+body', *s k+em-nibud'* lit. 'with who+m+body') and the parallelism with such obvious cases of phrasal pronouns as *Bog znaet kto* 'God knows who' or *kto by ... ni (prixodil)* lit. 'who ... wouldn't (come)' ≈ 'whoever (comes)'.

**Table 1**

☞ Idioms are shown by top corners, as in 'HIT THE HAY' 'go to bed'.

Premodifying LU	Indefinite pronominal lexeme	Postmodifying LU
1. ABY ≈ 'any-'	1. ČEJ 'whose'	1. 'BY ... NI' ≈ 'no matter'
2. 'BOG VEST' lit. 'God knows', ≈ 'some-'	2. ČTO 'what'	2. 'BY TO NI BYLO' ≈ '-ever'
3. 'BOG (EGO) ZNAET' lit. 'God (it) knows', ≈ 'some-'	3. GDE 'where'	3. -LIBO ≈ 'some-'
4. ČĚRT-TE lit. 'devil to you', ≈ 'any-'	4. KAK 'how'	4. 'NI POPADJA' ≈ 'no matter'
5. 'ČĚRT (EGO) ZNAET' lit. 'devil (it) knows', ≈ 'some-'	5. KAKOJ 'what a ...'	5. 'NI POPALO' ≈ 'no matter'
6. 'FIG (EGO) ZNAET' lit. 'fig sign (it) knows', ≈ 'some-'	6. KOGDA 'when'	6. -NIBUD' ≈ 'some-'
7. KOE- ≈ 'some-'	7. KTO 'who'	7. NADO ≈ 'ne should'
8. 'KTO (EGO) ZNAET' lit. 'who (it) knows', ≈ 'some-'	8. KUDA 'where to'	8. POPALO ≈ 'no matter'
9. MALO 'few'	9. OTKUDA 'where from'	9. PRIDĚTJA ≈ 'no matter'
10. 'MALO LI' ≈ 'any-'	10. POČEMU 'why'	10. -TO <sup>4</sup> ≈ 'some-'
11. NEVEST'/NIVEST' lit. 'not known', ≈ 'any-'	11. SKOL'KO 'how many/much'	11. UGODNO ≈ 'any-'
12. NEZNAMO 'not known', ≈ 'any-'	12. ZAČEM 'what for'	12. XOČEŠ' lit. 'you want', ≈ 'any-'
13. 'PĚS (EGO) ZNAET' lit. 'dog (it) knows', ≈ 'some-'		
14. REDKO lit. 'rarely', = 'few'		
15. 'ŠUT (EGO) ZNAET' lit. 'buffoon (it) knows', ≈ 'some-'		
16. XOT' lit. 'at least', ≈ 'any-'		
17. 'XREN (EGO) ZNAET' lit. 'horseradish (it) knows', ≈ 'some-'		
18. 'X.. (EGO) ZNAET' lit. 'penis (it) knows', ≈ 'some-'		

According to Table 1, Russian has  $\approx 360$  indefinite pronominal collocations, or phrasal indefinite pronouns.

**NB** The number of Russian phrasal indefinite pronouns is not exactly  $12 \times (18 + 12) = 360$ , specified by the table, for two reasons. On the one hand, some combinations of premodifiers/postmodifiers with *SKOL'KO* 'how much/many', *POČEMU* 'why' and *ZAČEM* 'what for' are impossible: \**aby skol'ko/počemu/začem* or \**počemu/začem ugodno/xočes'*, etc.; some other combinations are frozen as idioms, see Comment 2 below. On the other hand, Table 1 does not include several restrictedly used LUs, for instance:

—the indefinite pronominal lexemes *NASKOL'KO* 'by how many/much', which can be premodified only by 'BOG (ČĚRT, FIG, KTO, PĚS, ŠUT, XREN, X...) (EGO) ZNAET', and *POČĚM* 'what is the price of', which accepts far from all pre- and postmodifiers;

—the IT-markers, such as: 'LEŠIJ (EGO) ZNAET' 'the wood goblin (it) knows', which is rather infrequent; 'BOG POSLAL/POŠLĚT' 'God sent/will send', which combines mainly with ČTO 'what'; 'GLAZA GLJADJAT' lit. 'eyes look', which combines only with KUDA 'to where'; and NI + V<sub>PERF.FUT.2.SG</sub> (*kogo ni vybereš'* 'whoever you will elect'), NI + V<sub>IMPER.2.SG</sub> (*kogo ni vyberi* 'whoever elect<sub>IMPER</sub>') or NI + V<sub>INF</sub> (*kogo ni vybrat'* 'whoever to elect'), which introduce additional semantic and grammatical complications.

## Comments

1) The four major semantic/functional types of indefinite pronouns have been established in (Haspelmath, 1997, pp. 130–141):

- the 'dunno' type (Eng. *somebody*, Rus. *kto-to*);
- the 'want/pleases' type (Eng. *anybody*, Rus. *kto ugodno*);
- the 'it may be' type (Eng. *whoever*, Rus. *kto by to ni bylo*);
- the 'no matter' type (Eng. *no matter who*, Rus. *kto popalo*).

2) The English glosses supplied in Table 1 are more than approximate: the meanings of the corresponding Russian pronouns are complex and very different from those of the English indefinite pronouns. As said above, the accurate semantic description of Russian indefinite pronouns is not attempted in this paper: strictly speaking, such a description concerns the IT-markers rather than the pronominal lexemes themselves, cf. (Levontina & Šmel'ev, 2005 and 2018); see also (Padučeva, 1985, pp. 209–220, 2015–2017 and 2018).

3) Some Russian indefinite pronominal collocations are homonymous with idioms; for instance:

—The collocation ČTO-TO 'something' is homonymous with two (invariable) idioms:

'ČTO-TO'<sup>1</sup> 'approximately' (*čto-to okolo desjati tonn* lit. 'approximately about 10 tons');

'ČTO-TO'<sup>2</sup> 'for unknown reasons' (*Čto-to ty zagrustil* lit. 'For.unknown.reasons you became.sad').

☞ The period is used in the glosses to unite several English words that correspond to one Russian word.

- The collocation GDE-NIBUD' 'somewhere' co-exists with the idiom 'GDE-NIBUD' 'approximately' (*gde-nibud' okolo poludnja* lit. 'approximately about noon').
- The collocation XOT' KUDA 'anywhere' co-exists with the idiom 'XOT' KUDA' 'excellent' (*Kvartira byla xot' kuda* 'The apartment was excellent').
- The expression 'KOGDA-TO' functions only as an idiom meaning 'at a remote moment in the past or the future'.
- The expression 'ČERT-TE SKOL'KO' also functions only as an idiom meaning 'far too many/ much'.

This complication is especially important when considering meanings of indefinite pronominal collocations.

4) The postmodifying IT-marker '(BY) ... NI' represents the constructions shown in (2):

- (2) a. **Kto by k Ivanu ni** *prixodil, vse uxodjat dovol'nye*  
 'No matter who is coming to see Ivan, everyone leaves satisfied.'
- b. **Kogo tol'ko Ivan ni** *prinimaet, vse uxodjat dovol'nye*  
 'No matter whom only Ivan is receiving, everyone leaves satisfied.'

Note a possible contamination of the IT-marker 'BY ... NI' with the IT-marker 'BY TO NI BYLO', if the indefinite pronoun is КТО 'who':

- (3) a. 'BY ... NI': *Kto by to ni byl, my ego najdëm* 'No matter who this is, we'll find him'.  
 Here, the verb *byl* can be replaced with any verb, provided it fits semantically (*okazalsja* 'turned out; found oneself', *učinil* 'committed', etc.), and TO, which is a demonstrative pronoun, with ÈТО 'this'.
- b. 'BY TO NI BYLO': *Pust' prixodit kto by to ni bylo* 'Let no matter who come'.  
 Here nothing can be replaced.

Note also that the combination of an indefinite pronominal lexeme with the postmodifier 'TOL'KO ... NE' is not an indefinite, but an exclamatory pronoun:<sup>2</sup>

- (4) *Kto tol'ko k Ivanu ne prixodit!* lit. 'Who only to Ivan doesn't come!' =  
 'How many different people come to see Ivan!'

The meaning of an exclamatory pronoun includes the component ‘many different ...’; such a pronoun can be only used in exclamatory utterances and, signaling an emotional state of the Speaker, it is a non-descriptive lexeme—unlike all indefinite pronouns, which are descriptive lexical expressions.

5) The premodifying IT-markers of the form ‘BOG (EGO) ZNAET’ are described in (Mel'čuk, 2012 and 2021, pp. 309–332). A meticulous semantic and combinatorial description of several postmodifying IT-markers and of the premodifying marker ABY is found in (Levontina & Šmel'ev, 2005 and 2018), already mentioned above.

Russian indefinite pronominal collocations will be characterized in three steps: the base of the collocation, its collocater, and the deep-/surface-syntactic relation [D/S-SyntRel] between them.

### The Base of a Russian Indefinite Pronominal Collocation

The base of a Russian indefinite pronominal collocation has been called above *indefinite pronominal lexeme*. It is a polysem of an interrogative pronoun. In this way, twelve Russian indefinite pronominal collocations are directly related to Russian interrogative pronominal lexemes (in fact, to all Russian interrogative pronominal lexemes, except KOTORYJ ‘which’).

**NB** As indicated in (Haspelmath, 1997, pp. 26–27), such state of affairs—that is, indefinite pronouns being formally based on interrogative pronouns—is typical of the majority of the world's languages (63 out of 100 languages in Haspelmath's sample).

A Russian indefinite pronominal lexeme has a meaning corresponding to the one—central—component of an interrogative pronoun's meaning: namely, the meaning of one of the so-called basic ontological categories, a notion that goes back to Aristotle.<sup>3</sup>

ČEJ ‘whose’	: [related to a]	‘person’	KTO ‘who’	:	‘person’
ČTO ‘what’	: [non-human]	‘entity’	KUDA ‘where to’	: [moving to a]	‘place’
GDE ‘where’	: [in a]	‘place’	OTKUDA ‘where from’	: [moving from a]	‘place’
KAK ‘how’	: [in a]	‘manner’	POČEMU ‘why’	: [for a]	‘cause’
KAKOJ ‘what a ...’	: [having a]	‘property’	SKOL'KO ‘how many/much’	:	‘quantity’
KOGDA ‘when’	: [at the]	‘time [of] ...’	ZAČEM ‘what for’	: [with the]	‘goal [of] ...’

For instance, KTO-TO means ‘a person whose identity is unknown to the Enunciator’ (≈ ‘somebody’), and KOE-GDE, ‘in several places whose identity is known to the Enunciator, but unknown to the Addressee’ (≈ ‘in some places’).

A Russian indefinite pronominal lexeme has a theoretically interesting property: it cannot be used without its obligatory collocater, that is, without an IT-marker. In other words, an indefinite pronominal lexeme appears in a sen-

tence only as the base of one out of  $\approx 30$  collocations. These lexemes constitute a subclass of degenerate lexemes (Mel'čuk, 2023b, p. 45), that is, of lexemes functioning exclusively within phrasemes; this particular subclass is naturally called collocate-bound lexemes. All 12 Russian indefinite pronominal lexemes are collocate-bound.

### The Collocate of a Russian Indefinite Pronominal Collocation

Premodifying and postmodifying LUs appearing as collocates in Russian indefinite pronominal collocations are special particles—lexemes and idioms; as said above, they are markers of indefiniteness type, that is, they express particular “indefiniteness” meanings. The exact meaning and the combinatorial properties of an IT-marker must be described in detail in its own lexical entry. In the entry of its base—that is, in the entry of an indefinite pronominal lexeme—the corresponding IT-markers are only approximately specified by non-standard lexical functions [LFs]. These specifications are supposed to be sufficient (in the process of sentence production) just for the first rough choice by the Speaker; the finergrained selection is performed based on the IT-marker's own lexical entry.

### The SSyntRel Between the Components of a Russian Indefinite Pronominal Collocation

In the deep-syntactic structure, an IT-marker, being a modifier of the indefinite pronominal lexeme, is subordinated to it by the **ATTR(ibutive)** deep-syntactic relation, as all modifiers are. In the surface-syntactic structure, a special **pronominal-auxiliary** surface-syntactic relation [SSyntRel] is proposed to express this syntactic link (Mel'čuk, 2012, p. 12).<sup>4</sup> The **pronominal-auxiliary** SSyntRel covers  $\approx 360$  indefinite pronominal collocations plus an open set of freely constructed phrasal indefinite pronouns. It determines the linear position of the IT-marker with respect to the pronominal lexeme:

—If anteposed, this marker is placed closer to the pronominal lexeme than restrictive particles, but further than monosyllabic prepositions: *tol'ko ne aby* <**Bog znaet**> *s kem* ‘only not with anybody (with God knows whom)’ vs. *\*aby* <**Bog znaet**> *tol'ko ne s kem*.<sup>5</sup>

—If postposed, the collocational IT-marker follows the pronoun immediately.

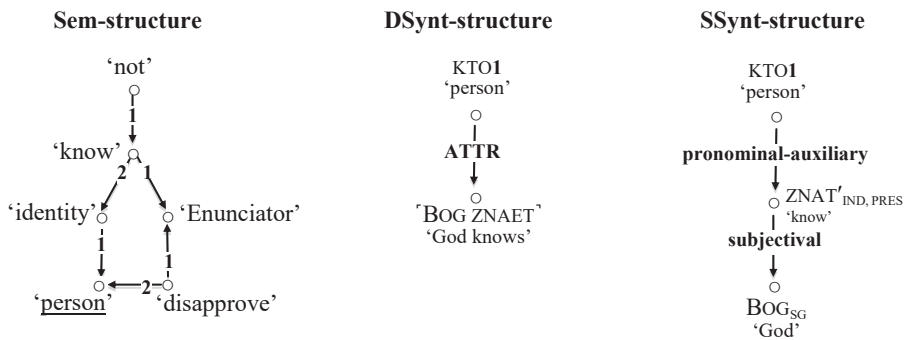
The ante-/post-position of the indefiniteness type marker is ensured by the indication in its syntactics (i.e., in its lexical entry): «anteposed» or «postposed».

Now the ground is ready to show the formal representation of Russian indefinite pronominal collocations. As an illustrative specimen the collocation *Bog znaet kto* ‘God knows who’ is taken, its base being the indefinite pronominal

lexeme  $\kappa\tau\omicron 1$ , and as the collocate, the idiom 'BOG ZNAET'. This collocation is quite typical of indefinite pronominal collocations—“seen one, seen them all.”

- (5) The structures of the Russian collocational indefinite pronoun 'Bog znaet' kto at three levels of linguistic representation.

☞ The underscoring of a semanteme in a Sem-structure indicates its communicatively dominant role.



## 2.2 Lexical Entries for Three Components of Russian Collocational Indefinite Pronouns

To throw more light on Russian collocational indefinite pronouns, three lexical entries are cited in full: for an indefinite pronominal lexeme ( $\kappa\tau\omicron 1$ ), for a pre-modifying indefiniteness type marker (the idiom 'BOG ZNAET'), and for a post-modifying IT-marker (the particle -TO<sup>4</sup>).

**$\kappa\tau\omicron 1$** , indefinite; masc, sg; collocate-bound lexeme.<sup>6</sup>

### Signified

'čelovek' = 'person'

### Syntactics

#### Obligatory Dependents

$\kappa\tau\omicron 1$ -**pronominal-auxiliary**→LF( $\kappa\tau\omicron 1$ ) [ $\kappa\tau\omicron 1$  is used only with one of its IT-markers]



**Lexical Functions**

specified, unknown to the Enunciator	: [~]-to <sup>4</sup> ; coll. 'Bog vest' <sup>17</sup> [~]
specified, unknown to, and disapproved of by, the Enunciator	: coll. 'Bog, čěrt (fig, kto, pės, šut, rude xren, vulg. x..) (ego) znaet' [~]; coll. čěrt-te, coll. nevest'/nivist', substand. neznamo [~]
specified, (several,) known to the Enun- ciator, unknown to the Addressee	: koe-[~]
specified, known to, and approved by, the Enunciator	: [~] nado
non-specified	: [~]-nibud'
non-specified, arbitrary	: [~] 'by to ni bylo'; [~]-libo; [~] ugodno; coll. [~] xočěš'; coll. xot' <sup>8</sup> [~]
non-specified, arbitrary, disapproved of by the Enunciator	: aby [~], [~] pridětsja
non-specified, arbitrary, strongly disapproved of by the Enunciator	: [~] 'ni popadja', 'ni popalo', popalo
non-specified, many, utterly variegated	: [~] '(by) ... ni' L <sub>(v)</sub>   L <sub>(v)</sub> →KTO1
non-specified, many, disapproved of by the Enunciator	: 'malo li' [~]
non-specified, few	: malo, redko [~]

**Examples**

- (i) *Ivan xot' kogo vyvedet iz sebja* 'Ivan can drive anybody mad'.  
(ii) a. *S kem Ivan tol'ko ni vstrečaetsja, on nnavitsja vsem*  
lit. 'With whomever Ivan only meets, he pleases everybody'.  
b. *S kem by Ivan ni vstrečalsja, on nnavitsja vsem*  
lit. 'With whomever Ivan would meet, he pleases everybody'.  
(iii) *Ivan mog rasskazyvat' istorii o kom ugodno* 'Ivan could tell stories about anybody'.

**Comments**

The formulations of the above non-standard LFs are based on the data from Padučeva's work; here are the abbreviations adopted.

Specified [X]: concretely referential X, that is, X referring to a particular entity present in the real world.

Non-specified [X]: non-referential X.

**NB** A non-referential LU X cannot be used in an “strongly affirmative” context (i.e., in an affirmation about a specified—referential—situation): \**Kto-nibud' prišël včera* ‘Somebody came yesterday’. vs. *Kto-nibud', vozmožno, prišël včera* ‘Somebody probably came yesterday’. | *Kto-nibud' pridët zavtra* ‘Somebody will.come tomorrow’. | *Kto-nibud' prišël včera?* ‘Did somebody come yesterday?’. The problem of the correlation between referentiality of an expression and its use in an affirmative context must be resolved at the semantic level: what is at stake here is the combinability of meanings rather than that of words.

[X] unknown to the Enunciator: X such that the Enunciator does not know X's identity.

Arbitrary [X]: X that is a freely choosable element of a particular set.

To put the proposed description of the indefinite pronominal lexeme  $\kappa\tau\omicron\iota$  in an appropriate context, it is worth sketching the structure of the polysemous vocable  $\kappa\tau\omicron$  ‘who’, of which  $\kappa\tau\omicron\iota$  is a lexeme.

**KTO** ‘who’, noun, pronominal.

### Signifier

{ $\kappa\tau\omicron$ } = /któ/ *kto*, /kavó/ *kogo*, /kamú/ *komu*, /kěm/ *kem*, /kóm/ *kom*

$\kappa\tau\omicron\iota$  ‘čelovek’ = ‘person’, indefinite; collocate-bound lexeme.

- (6) a. *Tuda kto ugodno možet prijti*  
 to.there person any can come  
 ‘Any person can go there.’  
 b. *Tuda kto -nibud' možet prijti*  
 to.there person some can come  
 ‘Some person [= ‘somebody’] can go there.’

$\kappa\tau\omicron\iota\text{a}$  ‘čelovek←P’ = ‘person←P’, pseudo-relative.

☞ The expression ‘person←P’ = ‘person.who P’ stands for ‘the person who does/undergoes P’; P represents the syntactic head of a pseudo-relative clause. In other words, the lexeme  $\kappa\tau\omicron\iota\text{a}$  can be used only in a pseudo-relative clause.

- (7) a. *Kto prišël, možet sadit'sja*  
 person.who came can sit.down  
 ‘The person who came can sit down.’  
 b. *Ja bespokojus' o kom ty rasskazyval včera*  
 I am.worried about person.who you spoke.about yesterday  
 ‘I am worried about the person you spoke about yesterday.’

- (8) *Vot komu ja dal knigu*  
 I.indicate.here to.person.who I gave book  
 ‘This is the person to whom I gave the book.’

кто2b ‘čelovek, č’ja identičnost’ ...’ = ‘person whose identity ...’, indefinite; collocate-bound lexeme.

- (9) *Ja vstretil ty ne ugadaeš’ kogo*  
 I met you not will.guess person.whose.identity  
 ‘I met you won’t guess whom.’ = ‘I met the person whose identity you won’t guess.’

- (10) *Tak vot komu ja dal knigu!*  
 so I.indicate.here to.person.whose.identity I gave book  
 ‘Now I know the identity of the person to whom I gave the book!’

кто2c ‘neopredelennyj čelovek’ = ‘a non-specified person’, indefinite, **colloquial**. Completely synonymous with *kto-nibud’*.

- (11) *Esi kogo uvidiš’, pozdorovajsja* ‘If you meet somebody, greet him.’  
 if a.person you.see greet

кто3 ‘identičnost’ človeka, kotoryj ...’ = ‘identity of person who ...’, interrogative.

- (12) *Kto prišël?* ‘Who came?’  
 identity.of.person.who? came

- (13) *Ja znaju, kto prišël* ‘I know who came.’  
 I know identity.of.person.who came

кто4: relative pronoun (that is, an empty substitute pronoun used to introduce a relative clause modifying a designation of human being; equivalent to WHICH/THAT/WHO). See (Iordanskaja & Melčuk, 2020).

- (14) *Nado imet’ delo s ljud’mi, kto ponimaet v torgovle*  
 ‘One must deal with people who understand commerce.’

## Comments

1) The order—or, more precisely, the numbering—of the lexemes inside the vocable *кто* follows their growing semantic complexity (however, the semantically empty pronoun *кто4* is placed the last): ‘person’ < ‘person.who’ < ‘person whose identity...’ < ‘a non-specified person’ < ‘identity of person who ...’

2) All the lexemes of this vocable have the same signifier, so that it is indicated just once, for the whole vocable.

3) *кто2a*: a pseudo-relative pronoun. A pronoun of this class is underlying the formation of a pseudo-relative clause (Mel'čuk, 2021, pp. 249–252):

- (15) a. *Maša poljubila, kogo ty privël*  
 ‘Masha fell.in.love with.whom you had.brought {to her}’.
- b. *Maše nravitsja, čto ty prinës*  
 ‘Masha likes what you have.brought {to her}’.
- c. *Maša xočet žit', gde žizn' deševle*  
 ‘Masha wants to.live where life is cheaper’.
- d. *Kobyлина, čej èžik, ... govorit: ...* [M. Zoščenko]  
 ‘Kobyлина, whose scrub brush it is, ... says: ...’
- e. *Kotorye bez deneg – ne ezđut s damami* [M. Zoščenko]  
 ‘Which are without money don't go out with no ladies’.

4) *кто2b*: a genuine indefinite pronominal lexeme. Its meaning involves the identity of a person rather than the person himself: one can naturally say *Ja znaju, kto prišël, no ja eë ne znaju* ‘I know who came, but I don't know her [i.e., we never met]’.

*Vot kto ...!* ‘Now I know who ...!’ is a non-standard collocation of *кто2b* (and of all indefinite pronominal lexemes), where *vot* ‘I.indicate.here’ expresses the meaning ‘now I know’.

5) *кто2c*: another indefinite pronominal lexeme, whose meaning—‘a non-specified person’—is different from the meaning of *кто2b*—‘person whose identity ...’. Such indefinite pronominal lexemes are currently called “bare interrogatives” (for instance, Haspelmath, 1997, p. 170), but this name is misleading.

6) *кто3*: an interrogative pronominal lexeme. The meaning of *кто3*—‘Communicate<sub>IMPER</sub> the identity of the person!’ consists of two components:

- ‘identity of the person who ...’, expressed by a form of *кто*;
- ‘communicate<sub>IMPER</sub>!’, expressed by the interrogative prosody (Bryzgunovas’s “Intonational Contour 2”).

The vocable *кто* specifies the general schema for the vocables of all Russian indefinite pronominal lexemes.

‘**BOG (EGO) ZNAET**’, lit. ‘God (it) knows’, idiom, particle; **colloquial**.

### Signified

‘Bog (ego) znaet’  $\underline{X}$  = ‘concretely-referential  $\underline{X}$  whose identity the Enunciator does not know and which he disapproves of’.

### Signifier

/bóxznait/, /bóxjivòznàit/

### Syntactics

Anteposed to X; X = L<sub>(pronominal, indefinite)</sub> [X is one of 12 Russian indefinite pronominal lexemes].

### Examples

- (i) *Deti tam zanimajutsja Bog znaet čem* ‘The kids are out there doing God knows what’.
- (ii) *On priexal Bog znaet otkuda* ‘He came God knows from where’.
- (iii) *Èti foto sobiralis’ dlja Bog ego znaet kakix reklamnyx celej*  
‘These photos have been collected for God it knows what advertising purposes’.

### Comment

This idiom (as all other idioms of this type with ZNAET) may include the expletive pronoun ON ‘he/she/it’ in the accusative, syntactically depending on ZNAT’ ‘know’.<sup>9</sup> Two cases, semantically equivalent, are possible:

- either the pronoun ON is used in a non-referential way, that is, in the singular of the neuter gender, as *ego* ‘it’ (16a);
- or it is coreferential with the SSynt-subject of the clause that contains the idiom, and then ON agrees with this subject in number and gender (16b):

- (16) a. *Maša z анимаetsja Bog ego znaet čem*  
lit. ‘Masha is doing God it knows what’.
- b. *Maša z анимаetsja Bog eë znaet čem*  
lit. ‘Masha is doing God her knows what’.

The addition of the expletive ON enhances the colloquial character of the idiom.

**-TO<sup>4</sup>**, particle.

**Signified**

'X-to' = 'concretely-referential X whose identity the Enunciator does not know'.

**Signifier**

/tõ/ [unstressed]

**Syntactics**

Immediately postposed to X; X = L<sub>(pronominal, indefinite)</sub> [X is one of the 12 Russian indefinite pronominal lexemes, see Table 1].

**Examples**

- (i) *Kakim-to xazaram kakoj-to Oleg za čto-to otmstil počemu-to* [A. Galič]  
lit. 'Some Oleg took some revenge for something on God knows what Khazars for some reason'.<sup>10</sup>

**Comment**

Russian has two lexemes TO, homographous with -TO<sup>4</sup>: TO<sup>1</sup> 'now ..., now ...', a repeated coordinating conjunction, and TO<sup>2</sup> ≈ 'then', a collocate of several subordinating conjunctions; both can be stressed:

TO<sup>1</sup>, as in *Ivan to plakal, to smejalsja, to smolkal* 'Ivan now was crying, now was laughing, now was silent'.

TO<sup>2</sup>, as in *Esli Ivan doma, to my ostanemsja* 'If Ivan is home, then we'll stay'.

Two fully homonymous unstressed particles are -TO<sup>3</sup> ≈ 'as for ...' and -TO<sup>5</sup> 'The Enunciator does not want to be more precise':

-TO<sup>3</sup>, as in *Ivan-to sumeet* 'As for Ivan, he'll manage'.

-TO<sup>5</sup>, as in *Ivan pobyval tam-to i tam-to* 'Ivan was to several places, it does not matter exactly where'.

Thus, the preliminary goals (p. 3) are reached; now I can switch to the final goal: the description of freely constructed Russian phrasal indefinite pronouns.

### 3 Russian Freely Constructed Phrasal Indefinite Pronouns

Russian freely constructed phrasal indefinite pronouns (a.k.a. *syntactic amalgams*) form an open set of two-component lexical expressions—free phrases—of the following form:

PREMODIFYING QUASI-CLAUSE ← **pronominal-auxiliary** – INDEFINITE PRONOMINAL LEXEME

The 12 Russian indefinite pronominal lexemes have been characterized in Section 2.

A quasi-clause premodifying an indefinite pronominal lexeme constitutes a marker of indefiniteness type (= an IT-marker). Informally, it is a freely constructed clause which has a Main Verb, but where the verb's subject or direct object (depending on the verb) has been “amputated”; this amputated element corresponds to the meaning ‘identity of [the entity designated by the indefinite pronominal lexeme]’. Thus, the resulting quasi-clause boldfaced in (17a) does not have its subject, and the quasi-clauses in (17b) lack their direct object:

- (17) a. *Ko mne podošël **tebe uže Ø<sup>BYT'</sup> jasno kto***  
 to me came to.you already be clear who  
 ‘It is already clear to you who came over to me.’
- b-i. *Ko mne podošël **ja ne rešus' skazat' daže mame kto***  
 to me came I not will.dare tell even to.Mom who
- b-ii. *Ko mne podošël **ja ne rešus' skazat' kto daže mame***  
 to me came I not will.dare tell who even to.Mom  
 ‘I won't dare to tell who came over to me even to Mom.’

The phrase *tebe uže jasno kto* means ‘a person whose identity is already clear to you’, where ‘X's identity’ = ‘information that allows for (at least, partial) X's identification’. Similarly, the phrase *ja ne rešus' skazat' daže mame kto* means ‘a person such that I won't dare to tell his identity even to Mom’. The wordform *kto* in the sentences of (17) belongs to the lexeme КТО<sub>2b</sub>, which has the meaning ‘person whose identity ...’; the Main Verb of the sentence semantically bears on ‘person’, while the Main Verb of the quasi-clause that constitutes the IT-marker depending on the indefinite pronominal lexeme bears on ‘identity’. The КТО<sub>2b</sub> ‘person whose identity ...’ lexeme is thus indeed an “amalgam.” (Lakoff's term is very convenient.)

By their grammatical nature, the underscored phrases in (17) are phrasal indefinite pronouns, similar to such phrasal indefinite pronouns as КОЕ-КТО ≈ ‘somebody’, АБЫ КТО ≈ ‘no matter who’ or ‘МАЛО ЛИ’ КТО ≈ ‘many arbitrary people’.

But while the latter pronouns are collocations of the lexeme КТО1 and must be listed in its lexical entry, the expressions of the type *ja ne rešus' skazat' daže mame kto* are free phrases. Nevertheless, their freedom is constrained.

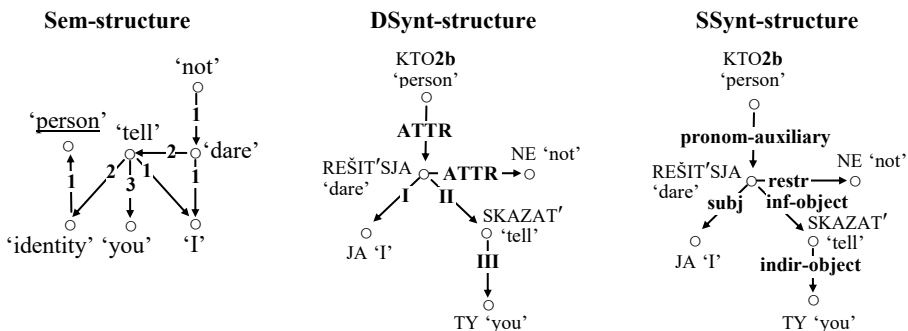
On the one hand, any of the lexical components of such a phrase can be replaced: for instance, one can say *Ix lager' okazalsja tol'ko professor Tugokaki – ty ego znaeš' – sumel soobrazit' gde* lit. 'Their camp turned out to be only Professor Tugokaki—you know him—could figure out where'; etc. Therefore, such expressions cannot be listed in a lexicon and have to be described by rules, more specifically, by semantic and deep-syntactic rules. However, the syntactic head of a free phrasal indefinite pronoun is necessarily one of the twelve indefinite pronominal lexemes. This is a lexical constraint.

On the other hand, the premodifying quasi-clause semantically bears on the semanteme 'identity [of ...]' and expresses the meaning 'X knows/doesn't know ...'. Therefore, its Main Verb has the meaning whose communicatively dominant semanteme is, at some depth of semantic decomposition, the semanteme 'know'. This is a semantic constraint.

To illustrate the formal representations of Russian indefinite pronominal free phrases at the semantic, deep-syntactic and surface-syntactic levels, the sentence *Ivan dal knigu ja ne rešus' skazat' tebe komu* 'Ivan gave the book I won't dare to tell you to whom' is chosen.

(18) Structures of the phrase *ja ne rešus' skazat' tebe komu* on three levels of linguistic representation

The structures given are partial: only the fragments relevant to my purposes here are shown, and the grammemes are not spelled out.

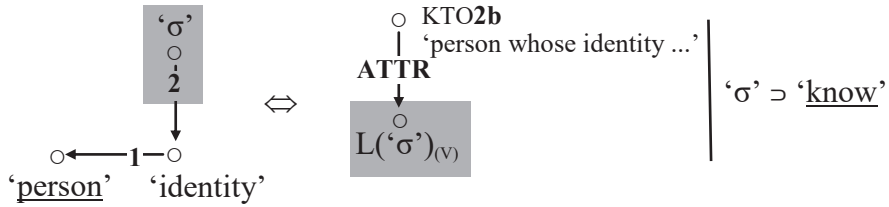


Based on the structures in (18), it is easy to formulate the corresponding rules.

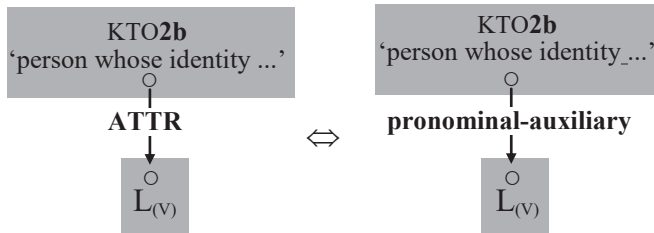
☞ The shading indicates the context of the rule: the fragments of the manipulated structures which are not affected by the rule itself, but whose presence is necessary for the rule to apply correctly.



(19) a. Semantic rule for the lexeme **кто2b**



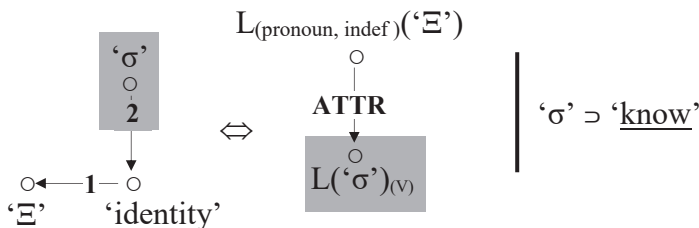
b. Deep-syntactic rule for the lexeme **кто2b**

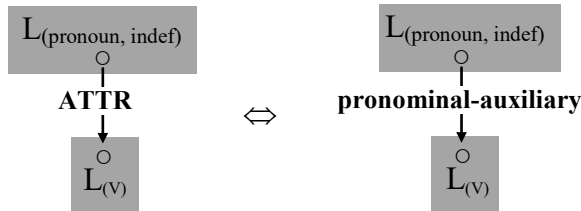


Similar Sem- and DSynt-rules must be written for the other 11 Russian indefinite pronominal lexemes (**что2b** ‘non-human entity whose identity ...’, **где2b** ‘in a place whose identity ...’, etc.). A pair of such rules—a Sem-rule and a DSynt-rule—has to be placed into the lexical entry of each corresponding indefinite pronominal lexeme.

Alternatively, one could write just one general schema for all 12 Sem-rules and one general schema for all 12 DSynt-rules specifying Russian free phrasal indefinite pronouns. These schemata are part of the grammar of Russian; namely, they must be included in “Syntax,” Section “Indefinite Pronouns.”

(20) General schemata of semantic and deep-syntactic rules for Russian free phrasal indefinite pronouns





## 4 Not to Be Confounded

In order to round off my presentation, I will cursorily characterize in parallel the following four syntactic phenomena, which are different in nature, but superficially fairly similar: they all involve the notorious WH-words. The first two of these are indefinite pronominal phrases (the object of the present paper), and the other two are pseudo-relative clauses. They can be and sometimes are confounded, so that it seems useful to insist once more on their distinctions.

### Indefinite pronominal phrases

- Collocational indefinite pronouns, such as  
{*Ivan priexal*} *Bog znaet otkuda* ‘Ivan came God knows from.where’
- Free phrasal indefinite pronouns, such as  
{*Ivan priexal*} *ty, konečno, dogadyvaeš’ja otkuda* ‘Ivan came you can.guess, of course, from.where’

### Pseudo-relative clauses

- Full pseudo-relative clauses, such as  
{*Ivan priexal*,} *otkuda nikto ne priezžal ran’she* ‘Ivan came from.where nobody had come before’
- Elliptic pseudo-relative clauses (results of the “Sluicing” transformation), such as  
{*Ivan priexal, no ja ne znaju*,} *otkuda* ‘Ivan came, but I don’t know from.where’

**Collocational indefinite pronouns** in Russian constitute, as said above, a closed set:  $\approx 360$ . They are listed in the lexicon, where they are specified by non-standard lexical functions in the lexical entries of their bases (indefinite pronominal lexemes, see Table 1, page 5). The formal description of Russian collocational indefinite pronouns is illustrated in Subsection 2.1, p. 10, (5).

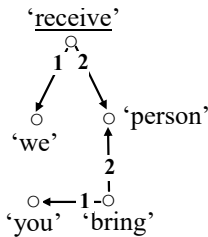
**Free phrasal indefinite pronouns** constitute an open set and are described by the general schemata of the Sem- and DSynt-rules, presented at the end of the previous section. The formal description of Russian free phrasal indefinite pronouns is illustrated in Section 3, pp. 19–20, (20).

NB The difference between collocational and free phrasal indefinite pronouns is not always obvious. Thus, for instance, why *neznamo←pron-auxil-kto* is a collocational pronoun, but *neizvestno←pron-auxil-kto* is a free phrasal pronoun? Because *NEZNAMO* is a single lexeme, while *neizvestno* is an element of a quasi-clause with the zero form of *BYT* ‘be’, cf. *Èto sdelał vsem nam Ø<sup>BYT</sup> (bylo) soveršenno neizvestno kto* lit. ‘This did to.all us is (was) absolutely unknown who’.

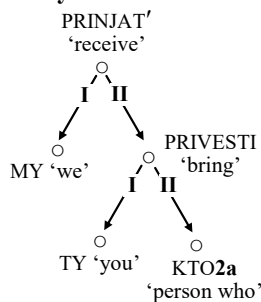
**Pseudo-relative clauses** are, strictly speaking, essentially unrelated to phrasal indefinite pronouns, but formally they are easily confounded with the latter, although A. Grosu (2006) clearly showed that they are absolutely different. Here are the structures of a Russian sentence with a pseudo-relative clause of the relevant type (the grammemes are not represented):

(21) *My primem, kogo ty privedeš’* lit. ‘We will.receive who you will.bring’.

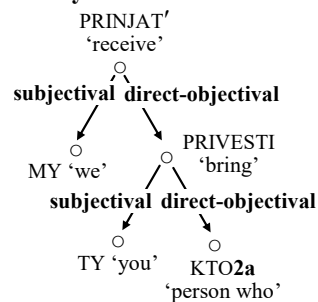
**Sem-structure**



**DSynt-structure**



**SSynt-structure**



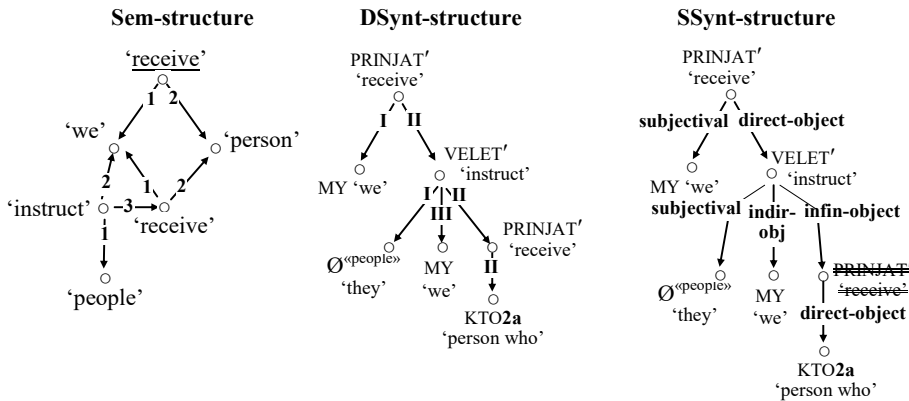
Recall that the pseudo-relative pronoun in the construction under consideration has to satisfy the requirements of the so-called matching parameter: roughly speaking, it must satisfy the government patterns of the two Main Verbs—that of the superordinate clause and that of the pseudo-relative one (Melčuk, 2021, p. 235, footnote 7).<sup>11</sup>

- Finally, **elliptic pseudo-relative clauses** represent a version of pseudo-relatives. In (22), the SemS, the DSyntS and the SSyntS of a sentence with an elliptical pseudo-relative clause are given (as above, the grammemes are not shown).

(22) *My primem, kogo nam veljat* ‘We will.receive whom they tell us {to receive}’.

☞ Ø<sup>‘people’</sup> stands for the Russian zero lexeme with the meaning ‘indefinite people’ (similar to Fr. ON or Ger. MAN).

The double strikethrough indicates the lexeme that is to be elided, that is, to remain unexpressed in the morphological string of the sentence.



## 5 Conclusions

The results of the present study may be summed up in the following points, relative to lexicology and syntax.

### Lexicology

- The characterization of a class of collocate-bound lexemes (which form a subclass of degenerate lexemes) that includes two groups of Russian indefinite pronominal lexemes. Of course, context-dependent LUs have been well known in linguistics; however, an accurate description of a linguistically important homogeneous set of lexemes that are used only as the bases of collocations was absent.
- A general schema for a lexicographic description of Russian pronominal vocables including indefinite, interrogative and pseudo-relative pronominal lexemes.

### Syntax

- The introduction of the class of pseudo-relative pronominal lexemes.
- The introduction of the **auxiliary-pronominal** SSyntRel.
- The rules that specify the open set of Russian free phrasal indefinite pronouns.

Although the present study deals exclusively with Russian, the observations proposed seem to have a more general validity for the linguistic typology.<sup>12</sup>

## Acknowledgments

The first draft of the paper has been discussed with David Beck, Lidija Jordanskaja, Sébastien Marengo and Jasmina Milićević; thank you from the depth of my heart for your constructive remarks and proposals.

## Appendix: Some Vital Notions

### collocation, base of collocation, collocate

A collocation is a compositional semantic-lexemic phraseme  $L_1$ -**synt**→ $L_2$  such that one of its components, say,  $L_1$ , is selected by the Speaker freely—for its meaning and syntactic properties, while the other, that is,  $L_2$ , is selected to express a particular meaning as a function of  $L_1$ .  $L_1$  is the **base** of the collocation, and  $L_2$  is its **collocate**. For example: *do* <sub>$L_2$</sub>  *a favor* <sub>$L_1$</sub>  vs. *give* <sub>$L_2$</sub>  *a kiss* <sub>$L_1$</sub>  vs. *take* <sub>$L_2$</sub>  *action* <sub>$L_1$</sub> .

### communicatively dominant semanteme

A communicatively dominant semanteme ' $\sigma_1$ ' in a configuration of semantemes ' $\sigma_1$ — $\sigma_2$ ' is the semanteme to which the configuration ' $\sigma_1$ — $\sigma_2$ ' can be reduced such that the meaning conveyed is simply impoverished, but not distorted; the communicative dominance of ' $\sigma_1$ ' is shown by underscoring. For example: in 'A bird is.singing' the semanteme 'sing' is communicatively dominant, since the utterance is about singing; in 'a singing bird', the semanteme 'bird' is communicatively dominant, since this utterance is about a bird.

### descriptive vs. non-descriptive lexical expressions

A descriptive lexical expression serves to **communicate** a meaning that denotes something in the real world and is spelled out in such a form that admits interrogation and negation; a non-descriptive lexical expression serves to **signal** a meaning that denotes an internal state of the Speaker in such

a form that does not admit interrogation and negation. For example: *I like it very much* vs. *Wow!* See (Iordanskaja & Mel'čuk, 1995) and (Mel'čuk, 2001, pp. 351–356).

### **Enunciator, the**

The Enunciator is the person to whom the whole given utterance **U** is attributed. The primary Enunciator is the Speaker, who produces **U**; the secondary Enunciator is the person to whom the Speaker attributes an utterance **U'** quoted inside **U** (i.e., **U'** is Indirect Speech). Thus, the primary Enunciator of the sentence *Ivan said that Masha was tired* is the Speaker, who wrote it, that is, me, Igor Mel'čuk; the secondary Enunciator, to whom I attribute the utterance *Masha was tired*, is Ivan. (On the contrast “the Speaker vs. the Enunciator,” see, in particular, Iordanskaja & Mel'čuk, 1995).

### **idiom**

An idiom is a non-compositional semantic-lexemic phraseme; for example: 'ALL THUMBS' 'very awkward' or 'HIT THE ROAD' '[to] leave'.

### **lexeme**

A lexeme is the set of all wordforms and analytical-form phrases that have a common signifier and semantically differ only by inflectional significations; for example:

I = {I, me}; SEE = {see, sees, saw, seeing, have seen, am seen, will see, ...}.

In prose, a lexeme is one word taken in one well-defined sense.

### **lexical unit**

A lexical unit [LU] is a lexeme or an idiom; every LU has its own lexical entry, and every lexical entry corresponds to an LU.

### **phraseme**

A phraseme is a phrase that cannot be produced freely, that is, in such a way that each of its lexical components is selected by the Speaker independently of its other components. Phrasemes come in four families: idioms, collocations, nominemes, and clichés.

### **polysem**

An LU  $L_1$  is a polysem of LU  $L_2$  if and only if  $L_1$  and  $L_2$  both belong to the same vocable  $L$ . For example: BREADI 'baked leavened food' and BREADII 'livelihood—as if it were breadr' are polysems of the vocable BREAD.

**pronoun/pronominal lexeme**

A pronoun is a lexeme that has 1) a very poor meaning or no meaning at all, 2) a particular syntactic behavior (e.g., cannot receive dependents) and 3) quite a special morphology (e.g., in English only personal pronouns have cases, and only demonstrative-pronominal adjectives have the plural (*this ~ these, that ~ those*). Pronouns can belong to all parts of speech: they can be nouns (I, SOMEBODY), adjectives (SOME, NONE), adverbs (HERE, THEN) and verbs (DO).

**pseudo-relative clause**

A pseudo-relative clause is a subordinate clause that formally has the structure of a relative—it is introduced by a WH-word, but is semantically equivalent to a nominal or prepositional phrase: it does not modify a noun, but is used as an actant. For example: *I am interested in **what John said*** or *He sent me **where John was staying***. (Pseudo-relatives are also known as *free*, or *headless, relatives*.) See Melčuk (2021, pp. 249–252).

**vocable**

A vocable is the set of all LUs whose signifiers are identical and whose signifieds share important enough components (these components are called semantic bridges).

**Abbreviations and Notations**

<b>coll.</b>	: colloquial	SemS	: semantic structure
DSyntS	: deep-syntactic structure	SSyntRel	: surface-syntactic relation
IT-marker	: indefiniteness type marker	SSyntS	: surface-syntactic structure
L	: a particular lexical unit	'L <sub>1</sub> L <sub>2</sub> ... L <sub>n</sub> '	: idiom L <sub>1</sub> L <sub>2</sub> ... L <sub>n</sub>
LF	: lexical function	'σ'	: the communicatively dominant semanteme in a SemS
LU	: lexical unit		

## References

- Apresjan, V. (2015). *Ustupitel'nost': mexanizmy obrazovanija i vzaimodejstvija složnyx značenij v jazyke* [Concessivity: mechanisms of formation and interaction of complex meanings in language]. *Jazyki slavjanskoj kul'tury*.
- Apresjan, V., & Iomdin, B. (2022). Russkie voprositel'nye mestoimenija kak leksiko-grafičeskij tip [Russian interrogative pronouns as a lexicographic type]. *Russian Journal of Linguistics*, 26(4), 1078–1113. <https://doi.org/10.22363/2687-0088-30714>.
- Durie, M. (1985). *A grammar of Acehnese on the basis of a dialect of North Aceh*. Foris.
- Grosu, A. (2006). An amalgam and its puzzles. In: H.-M. Gärtner, S. Beck, R. Eckardt, R. Musan, & B. Stiebels, Eds., *Between 40 and 60 puzzles for Krifka*. [https://www.tau.ac.il/~grosua/papers/Grosu\\_2006\\_An\\_Amalgam\\_And\\_Its\\_Puzzles.pdf](https://www.tau.ac.il/~grosua/papers/Grosu_2006_An_Amalgam_And_Its_Puzzles.pdf).
- Haspelmath, M. (1997). *Indefinite pronouns*. Oxford University Press.
- Iomdin, L. (2010). Sintaksičeskie frazemy: meždu leksikoj i sintaksisom [Syntactic phrasemes: between lexicon and syntax]. In: Ju. Apresjan, I. Boguslavskij, L. Iomdin & V. Sannikov, *Teoretičeskie problemy russkogo sintaksisa. Vzaimodejstvie grammatiki i slovarja* [Theoretical problems of Russian syntax. Interaction between grammar and lexicon], pp. 141–190. *Jazyki slavjanskix kul'tur*.
- Iordanskaja, L., & Mel'čuk, I. (1995). Traitement lexicographique de deux connecteurs textuels du français contemporain 'EN FAIT' vs 'EN RÉALITÉ'. In: H. Bat-Zeev Shyldkrot & L. Kupferman, Eds., *Tendances récentes en linguistique française et générale (volume dédié à David Gaatone)*, pp. 211–236. John Benjamins.
- Iordanskaja, L., & Mel'čuk, I. (2020). Mestoimenija KTO i KOTORYJ v opredelitel'nom pridatočnom [The pronouns KTO 'who' and KOTORYJ 'which' in a modifying subordinate clause]. In: L. Iomdin & I. Boguslavskij, Eds., *Ot semantičeskix kvarkov do vselennoj v alfavitnom porjadke. K 90-letiju akademika Jurija Derenikoviča Apresjana [Trudy Instituta russkogo jazyka im. V. V. Vinogradova, № 2]*, pp. 107–129. IRJaz RAN.
- Krejdlin, G. (1994). Ob odnoj lakune v sisteme russkix mestoimenij: vosklicatel'nye mestoimenija russkogo jazyka [On a lacuna in the system of Russian pronouns: exclamatory pronouns in Russian]. *Rusistika segodnja*, № 1, 56–65.
- Lakoff, G. (1974). Syntactic amalgams. *Chicago Linguistic Society: Papers from the 10<sup>th</sup> Regional Meeting*, 321–344.
- Levontina, I., & Šmelëv, A. (2005). Maloizučennye edinicy so značeniem nezadannosti kriteriev vybora v russkom jazyke [Understudied units with the meaning of unspecified choice criteria]. In: N. Arutjunova, Ed., *Logičeskij analiz jazyka. Kvantifikativnyj aspekt jazyka*, pp. 638–651. Indrik.



- Levontina, I., & Šmel'ev, A. (2018). *Aby*: korpusnoe issledovanie v aspekte sinxronii [The Russian *aby*: Corpus-driven research (synchrony and diachrony)]. *Computational Linguistics and Intellectual Technologies*, 17(24), 436–447.
- Matsuyama, T. (2015). The syntactic structure of WH-syntactic amalgams. *English Linguistics*, 32(1), 78–101.
- Mel'čuk, I. (2001). *Communicative organization in natural language. The semantic-communicative structure of sentences*. John Benjamins.
- Mel'čuk, I. (2012). Mestoimennye vyraženiia s imenem čertyxatel'nym tipa [*Ona uexala čert znaet kuda* i im podobnye v russkom jazyke [Pronominal expressions with a “blasphemous” noun of the type [*Ona uexala čert znaet kuda* and similar ones in Russian]. *Russkij Jazyk v Naučnom Osveščennii*, 2(24), 5–22.
- Mel'čuk, I. (2021). *Ten studies in dependency syntax*. De Gruyter Mouton.
- Mel'čuk, I. (2023a). Čto èto za X? ‘What is this X?’ In: S. Pereverzeva, I. Itkin & L. Xesed, Eds., *Grani estestvennogo jazyka i kinesiki. Sbornik statej k 75-letiju G. E. Krejdlina*, 238–250. Diskurs.
- Mel'čuk, I. (2023b). *General phraseology: Theory and practice*. John Benjamins.
- Padučeva, E. (1985). *Vyskazyvanie i ego sootnesënnost' s dejstvitel'nost'ju* [The utterance and its relation to reality]. Nauka.
- Padučeva, E. (2015–2017). *Rusgram: russkaja korpusnaja grammatika* [Rusgram: Russian corpus-based grammar], chapters 25, 27, 28, and 34. <http://rusgram.ru/new/>.
- Padučeva, E. (2018). Russkie mestoimenija svobodnogo vybora [Russian free choice pronouns]. *Russian Linguistics*, 42(3), 291–319.

## Notes

<sup>1</sup> (p. 4) In Haspelmath (1997, pp. 21–22) it is called simply *indefiniteness marker*, which seems infelicitous, given that the pronominal lexeme itself is indefinite.

<sup>2</sup> (p. 7) On exclamatory pronouns in Russian, see Krejdlin (1994) and Mel'čuk (2023a and 2023b, pp. 103–108).

<sup>3</sup> (p. 8) This state of affairs is found in many languages of the world, see Haspelmath (1997, pp. 29–31). On Aristotle's categories, see <https://plato.stanford.edu/entries/categories/>. Durie (1985, p. 151ff) explicitly identifies the interrogative pronouns, or WH-words (of Acehnese), which he aptly calls *epistemological classifiers*, with basic ontological categories and demonstrates their prominent role in the language.

<sup>4</sup> (p. 9) In Mel'čuk (2021, p. 91) this SSyntRel is called **WH-pronominal**, but a less Anglo-centric name seems preferable.

<sup>5</sup> (p. 9) A polysyllabic preposition precedes the premodifying IT-marker: *blagodarja ne aby komu* ‘thank.to not anybody’ or *iz-pod koe-čego* ‘from-under something’.

<sup>6</sup> (p. 10) The characteristics “noun, pronominal” (= nominal pronoun) are indicated for the whole vocable  $\kappa\tau\omicron$ , to which the lexeme  $\kappa\tau\omicron\iota$  belongs, see p. 12. The characteristics “masculine, singular” of  $\kappa\tau\omicron\iota$  are not absolute. In some special contexts,  $\kappa\tau\omicron\iota$  can appear:

- In the feminine, as, for instance, in *Ešli **prixodila**<sub>FEM</sub> kto-nibud' iz ètix ženščin<sub>(fem)</sub>...* ‘If someone of these women came...’
- In the plural, as, for instance, in *Sjuda **xodjat**<sub>3,PL</sub> Bog znaet kto (kto ni popadja)* ‘God knows who (Just anyone) are coming here.’

Since these properties are marginal and not relevant to our discussion, they are not described here.

<sup>7</sup> (p. 11) In (Mel'čuk, 2012) the expressions of the form *Bog (ččert, ...) znaet kto (čto, gde, kuda, kak, ...)* are described as idioms, which I consider incorrect now. At that time, my toolkit did not contain the notion of collocate-bound lexeme. Today I know that a phrase of the form *Bog znaet kto* is a collocation of the collocate-bound lexeme  $\kappa\tau\omicron\iota$  ‘person’ as the base with the idiom ‘ $\text{BOG (EGO) ZNAET}$ ’ [ $\underline{X}$ ]  $\approx$  ‘[ $\underline{X}$ ] whose identity the Enunciator does not know ...’ as its collocate.

<sup>8</sup> (p. 11) On the particle  $\text{XOT'}$ , see Apresjan, V. (2015, p. 240).

<sup>9</sup> (p. 15) The expletive  $\text{ON}$  is subordinated to the  $\text{ZNAT'}$  verb by the **quasi-direct-objectival** SSynt-relation (Mel'čuk, 2021, pp. 54–55).

<sup>10</sup> (p. 16) This is a parody of a line from Pushkin's famous poem *The Song of Wise Oleg*, which every Russian knows by heart from elementary school.

<sup>11</sup> (p. 21) The Russian language features an interesting violation of the matching parameter requirements. Namely, a pseudo-relative clause in the role of SSynt-subject can itself have the pseudo-relative pronoun in any morphological form:


(i) *Segodnja **prixodil komu** ty podaril knigu (s **kem** ty besedoval)*  
 today came to.whom you had.given book with whom you had.been.talking

<sup>12</sup> (p. 23) Thus, see, for instance, Matsuyama (2015) about the English phrasal indefinite pronouns. The author proposes that the indefinite pronominal lexeme (such as  $\text{WHO}$ ,  $\text{WHAT}$ , etc.) is the syntactic head of the whole expression, while the rest—the quasi-clause, in my terms—is a kind of parenthetical. However, a closer comparison of his results with my description is complicated by the fact that he uses the phrase-structure representation of the syntactic structure.



**Salah Mejri**

Université Sorbonne Paris Nord  
France

 <https://orcid.org/0000-0003-0094-6181>

# Prédicats, sens, polylexicalité et figement : un parcours heuristique\*

## **Predicates, Sense, Polylexicality, and Freezing: a heuristic path**

### **Abstract**

For Gaston Gross, a rigorous study of the language means describing all the uses of lexical units of the language within the framework of free combinatorics, by determining the argumental paradigms involved in the schemes of arguments for predicates and all the appropriate predicates for arguments. This approach is complemented by a description of frozen sequences whose internal combinatoriality is more or less restricted. The author devotes his work of 1996 and part of his textbook of 2012 to this subject. Our contribution highlights the great richness of the author's thoughts particularly from a heuristic point of view. We start from the concepts of predicate, frozenness, polylexicality and semantic opacity to show how much impact Gaston Gross's methodological suggestions have had on the development of the discipline of linguistics.

### **Keywords**

Gaston Gross, Frozenness, Polylexicality, Semantic opacity, Heuristic path

---

\* Je remercie les deux rapporteurs anonymes qui m'ont aidé à améliorer la version originale.

« Le langage est le triomphe  
de l'ingéniosité humaine »  
(A. N. Whitehead, 2004 : 54)

## 1. En guise d'introduction

Quand j'ai eu la chance et l'honneur de prendre la suite de Gaston Gross (G.G.) dans sa chaire « Syntaxe et sémantique » et dans la direction du laboratoire de linguistique informatique (LLI), j'ai pris conscience de la portée épistémologique de sa pensée. Parallèlement aux échanges que j'avais continuellement eus avec lui, je me nourrissais de la pensée profonde de Robert Martin. Ainsi ai-je évolué pendant ces trois dernières décennies entre deux grands esprits alliant la rigueur de la pensée à la précision dans la démarche, avec une ambition commune, malgré les différences de point de vue : participer à faire émerger une méthodologie adaptée pour la linguistique<sup>1</sup>. C'est cette dimension méthodologique que je voudrais souligner dans cette contribution à la mémoire de Gaston Gross qui, indépendamment des questions traitées, transcende sa réflexion. Cela s'inscrit dans la suite des discussions que j'ai eues avec lui jusqu'à la fin de ses jours.

L'un des points méthodologiques les plus cruciaux dans l'étude de tout objet est la détermination avec précision de l'élément jugé central à partir duquel l'on pourrait accéder facilement à l'ensemble des autres composantes du système, c'est-à-dire là où s'entrecroisent et se concentrent non seulement tous les ingrédients constitutifs mais également et surtout l'ensemble des mouvements de la dynamique interne. L'extrême complexité des systèmes linguistiques a conduit les linguistes à chercher les points les plus cruciaux pour accéder à l'économie générale de la langue. Selon les modes et les aléas de la sociologie de la recherche, morphologie, syntaxe, phonologie, sémantique, lexicologie, etc. sont soit privilégiées soit marginalisées. G.G. a choisi le lexique comme point d'entrée pour décrire le système de la langue.

Il est convaincu que tout passe par le lexique, confrontant ainsi l'intuition courante qu'une langue est d'abord des mots où siègent morphologie, syntaxe, sémantique, etc. Pour procéder à des descriptions systématiques de la langue, il faut relever le défi de s'attaquer à la totalité du lexique. Pour ce faire, des outils méthodologiques sont nécessaires. Pour éviter les inconvénients des approches

---

<sup>1</sup> Cf. par exemple les deux manuels : Martin, R. (2014). *Comprendre la linguistique*. PUF ; Gross G. (2012). *Manuel d'analyse linguistique*. Presses universitaires du Septentrion.

fixistes du lexique, G.G., s'inscrivant dans la lignée de Zellig S. Harris et de Maurice Gross, opte pour le Lexique-grammaire, approche qui a le mérite d'expliquer les différents emplois des mots grâce à l'ensemble des transformations syntaxiques que ces emplois admettent. Se profile derrière ce type d'analyse, dont la visée principale est l'élaboration d'un formalisme pour le traitement automatique des langues (cf. les travaux de LADL), la notion de prédicat qui met au centre de l'analyse la dimension logico-sémantique (R. Martin, 2021). Sa thèse sur les constructions converses du français (1989) illustre parfaitement comment le croisement entre lexique, syntaxe et sémantique permet de dégager des classes de substantifs ayant des caractéristiques distributionnelles et transformationnelles propres, actualisées par des verbes supports exprimant des relations converses (du type *donner/recevoir*), dont le nombre s'élève à 3500 constructions<sup>2</sup>. Depuis cette publication, G.G. a enrichi son cadre d'analyse par une dimension sémantique de plus en plus fine qui va le conduire, moyennant la notion d'emploi, aux classes d'objets (D. Le Pesant & M. Mathieu-Colas, 1998), à l'expression de la finalité et à la sémantique de la cause (2010). Il a consacré son énergie pendant les dernières années de sa vie aux verbes supports, dont il a dénombré plusieurs milliers<sup>3</sup>.

Pour retracer le cheminement qui a conduit G.G. au figement, nous rappellerons ce qu'il entend par relation prédicative et comment il établit le lien entre cette relation et les contenus sémantiques des unités lexicales. Après quoi nous essaierons de montrer comment ses travaux de pionnier sur le figement (1996) ont nourri pendant ces trois dernières décennies la réflexion linguistique sur le figement en tant que processus linguistique qui agit en profondeur dans tout le système linguistique, sur la problématique de la polylexicalité qui en découle et ce qu'elle pose comme problème pour la définition de l'unité lexicale, sur l'opacité sémantique des unités polylexicales et l'émergence d'une nouvelle globalité sémantique, sur les contraintes énonciatives de certaines séquences figées comme les pragmatèmes et les contenus culturels qu'elles charrient, et sur la saillance sémiotique de ces unités leur permettant de jouer un rôle central dans l'organisation textuelle et dans la prédication oblique comme dans le défigement ; le fil conducteur étant l'apport heuristique de la démarche de G.G.

---

<sup>2</sup> Cf. le compte rendu d'A. Bertin-Delbey (1991).

<sup>3</sup> Communication personnelle. Espérons que les résultats d'un tel travail verront le jour.

## 2. Au cœur de la réflexion de Gaston Gross, le prédicat

Nous n'enrons pas ici dans le débat relatif à la définition du prédicat. Nous nous contentons de renvoyer à l'abondante littérature sur la question (cf. entre autres M. Forsgren, K. Jonasson & H. Kronning (éds), 1998; A. Guilbert & C. Leclère (éds) 1981; J.-J. Franckel 1989; G. Gross, 1996; M. Gross, 1981; P. de Rouilhan, 1984; Zellig S. Harris, 1973; C. Muller, 2013, etc.). La liste est longue. Nous retenons seulement l'excellente synthèse faite par Claude Muller où il essaie de dégager l'ensemble des orientations théoriques des différentes conceptions. Celle qui nous intéresse ici parmi les diverses conceptions est celle qui voit dans le prédicat une fonction, c'est-à-dire une relation, établie entre des variables :  $f(x)$ . Robert Martin (2021 : 19) en fait un universel linguistique et l'enrichit avec la modalisation; ce qui donne la forme universelle suivante : «  $M(P_A)$  »,  $M$  étant la modalisation qui prend dans sa visée la relation prédicative ( $P$ ) établie entre des arguments ( $A$ ) : « La prédication est un des fondements de la grammaire universelle ».

Pour G.G., la relation prédicative représente la charpente logico-sémantique de la phrase. Qu'il s'agisse de phrase élémentaire ou complexe, la fonction prédicative assure la distribution des arguments selon un nombre et un ordre bien établi dont dépend l'interprétation finale de la phrase. La différence entre une phrase élémentaire, qui ne comporte qu'un seul prédicat, comme :

Le voisin arrose les plantes.

Arg<sub>1</sub> Prédicat Arg<sub>2</sub>

et une phrase complexe comme :

Les récoltes sont détruites à cause de la grande sécheresse.

Arg<sub>1</sub> Prédicat Arg<sub>2</sub>

est une différence de hiérarchie prédicative. Dans la seconde phrase, l'on peut opérer toutes sortes de transformations : la relation prédicative reste la même, malgré les variations formelles. L'on peut opter par exemple pour une forme verbale exprimant le prédicat de second ordre :

La grande sécheresse a causé la destruction des récoltes.

arg. arg.

on constate que le schéma argumental demeure le même.

Se profile derrière cette relation prédicative, telle qu'elle est exprimée par la forme verbale, une double prédication : une relation syntaxique qui prend en charge la distribution des éléments de la phrase (sujet-verbe-complément) et une relation sémantique assurée par le sémantisme du verbe *causer*. Les travaux que G.G. a consacrés aux prédicats non verbaux, et conséquemment aux verbes supports, sont, à notre avis, la preuve incontestable qu'il existe deux types de prédicats : un prédicat syntaxique qui se charge de l'actualisation du prédicat non verbal, assurée par le verbe support appauvrie sémantiquement – sans tout à fait perdre la totalité de son contenu sémantique –, et le prédicat sémantique véhicule par une forme lexicale non verbale.

Partant de la centralité de la prédication dans la langue, telle qu'elle est explicitée par G.G., des développements ultérieurs ont vu le jour. L'apport principal de G.G. consiste principalement dans l'abolition des frontières imposées par la tradition :

Il a libéré le prédicat de la forme verbale en l'étendant à d'autres parties du discours comme le nom, l'adjectif, l'adverbe, la préposition et la conjonction. Partant du constat que l'analyse traditionnelle focalise sur la morphologie verbale, il considère que les deux phrases suivantes ont la même structuration prédicative, la différence étant dans la différence du mode d'actualisation :

Le jury l'a félicité *parce qu'*il a battu tous les records.

Le jury l'a félicité *pour* avoir battu tous les records.

L'actualisation, étant conditionnée par des règles de bonne formation de la langue concernée, le recours à la forme conjuguée à un mode personnel (non-coréférence entre les deux sujets des verbes concernés), ou non (coréférence), est indépendant de la relation prédicative :

Il fait tout *pour* que la récolte soit sauvée.

Il fait tout *pour* sauver la récolte.

G.G. n'a pas poussé seulement les frontières des catégories grammaticales (parties du discours) de la prédication, il les a étendues à l'actualisation prédicative. Étudiant en détail la dimension aspectuelle, il insiste sur la distribution des marques du contenu aspectuel sur plusieurs éléments de la phrase autres que la forme verbale. Tel est le cas de l'expression du multiplicatif (itératif) dans la phrase suivante, assurée par le déterminant défini du prédicat nominal :

Il a fait *plusieurs* achats.

Certaines actualisations peuvent être assurées par des modifieurs :

Il a fait *des* achats *compulsifs*.

Comme l'unité d'analyse est la phrase, G.G. défend l'idée que la polysémie est un artefact et que toute analyse visant la solution de l'ambiguïté lexicale doit se faire dans un cadre homonymique, reposant sur la notion d'emplois sémantiques qui s'opèrent entre l'unité lexicale concernée et son environnement. Tel est le cas dans les emplois suivants du verbe *prendre* :

Il *prend* un livre sur l'étagère. (*prendre*<sub>1</sub> : saisir)

Il *prend* un café. (*prendre*<sub>2</sub> : boire)

Il *prend* un sandwich. (*prendre*<sub>3</sub> : manger)

L'on pourrait comprendre à partir de cette analyse que les unités lexicales renferment des contenus sémantiques selon la classe sémantique des paradigmes auxquels appartiennent les unités de son environnement immédiat. Vu sous l'angle de la virtualité de la langue, l'on peut considérer que le lexique de la langue concentre en lui-même des emplois spécifiques comportant des contenus sémantiques imposant des contraintes combinatoires spécifiques à chaque emploi. Cela reviendrait à dire que les unités lexicales, mises sous pression de la dynamique constante des emplois, encapsulent des contenus prédicatifs virtuels, actualisables selon les besoins de la production langagière. Dans cette perspective, les paraphrases définitionnelles ne seraient en fin de compte que des ensembles de prédicats qui s'activent ou se désactivent selon les exigences des énoncés dans lesquels ils se réalisent.

Ainsi pourrait-on distinguer deux grandes classes de prédicats : prédicats virtuels renfermés dans les unités lexicales représentant leurs potentialités combinatoires, et prédicats, que l'on pourrait qualifier d'énoncifs<sup>4</sup>, qui sont nécessairement actualisés dans un énoncé quelconque lors de la production langagière. Si G.G. privilégie dans son analyse ce second type de prédicat, son analyse homonymique du lexique laisse suggérer la potentialité prédicative telle qu'elle s'exprimerait dans les définitions ou les paraphrases permettant d'isoler les significations propres à chaque emploi. La trilogie méthodologique est ainsi instituée : un lexique, une combinatoire et une signification. Le lexique est le siège de la combinatoire et du sens ; la combinatoire gouverne la congruence des associations

<sup>4</sup> *Énoncif* signifie « relatif à l'énoncé » et s'oppose à *énonciatif* qui signifie relatif à l'énonciation. Ainsi obtiendrait-on le paradigme suivant : *énoncé/énonciation* ; *énoncif/énonciatif*.



syntagmatiques ; le sens émerge à la faveur du croisement des deux. Le tout est versé dans le moule prédicatif.

Cette extension de la notion de prédicat qui montre que tout est prédicatif dans la langue se trouve développée dans un autre cadre théorique. Le dernier article d'Alain Lemaréchal (2022 : 51–109), intitulé « Des prédicats à perte de vue ». (Ryle 1933) : pourquoi faire ? Ethocentrismes et tabous », pousse encore plus loin le concept de prédicat en l'appliquant à plusieurs langues, notamment celles qui ne possèdent pas de copule comme *être* ou tout autre élément équivalent et montre comment le prédicat, en plus de la forme nominale, peut s'appliquer à des adpositions, à des marques casuelles, à des indexicaux et même à des « déterminants et autres spécifieurs » (*Ibidem*, 52). Cette analyse rejoint la nôtre et nous y souscrivons.

### 3. Le sens au croisement de l'ensemble des dimensions linguistiques

On l'aura compris, la sémantique représente le point focal de sa réflexion. À un niveau très abstrait se situe la relation prédicative sans laquelle il n'y a pas de prédication possible. Cette relation crée des positions à saturer lexicalement. Ces positions impliquent les deux types de relations déjà retenus ci-dessus : les relations syntaxiques et les relations strictement sémantiques. Les premières, non moins pourvues de contenus sémantiques, reposent sur le caractère orienté de la relation prédicative. « La prédication va normalement de l'argument *a* à l'argument *b* dans la relation binaire ; elle va de l'argument *b* et *c* dans la relation ternaire » (R. Martin, 2021 : 141). En l'absence de symétrie entre les deux pôles du schéma argumental, « s'instaurent [...] universellement les fonctions argumentales de sujet, d'objet et d'objet second » (*Ibidem*). « Les langues qui possèdent un passif sont en état de d'inverser l'orientation initiale » (*Ibidem*). Se profilent derrière ces fonctions argumentales les notions de source (le sujet), de cible (l'objet), de bénéficiaire, de patient, etc. (l'objet second). Si l'on pousse l'analyse un peu plus loin, de nouvelles notions émergent à la faveur de certaines classes de prédicats comme les rôles sémantiques d'agent, de patient, d'instrument, de locatif, etc.

Une fois les positions de la relation prédicative définies en nombre et selon un ordre bien déterminé, les unités lexicales viennent les saturer, à l'exception des prédicats avalents qui trouvent leur support référentiel en dehors de la langue. Interviennent alors des contenus sémantiques plus précis. D'où

l'ensemble des interactions entre tous les éléments constitutifs de l'énoncé. C'est là que G.G. choisit de situer sa description des emplois des mots tels qu'ils sont discriminés par des critères formels. C'est pourquoi il procède à dégager les classes d'objets, c'est-à-dire les paradigmes distributionnels devant saturer les différentes positions argumentales pour chaque prédicat. Deux critères sont retenus : une invariance du sens de l'emploi corrélée à un ensemble de variations de combinaisons spécifiques à cet emploi. Ce qui définit l'emploi propre à une unité lexicale peut être ainsi ramené à ces trois éléments formalisables pour un éventuel traitement automatique : un schéma prédicatif, des paradigmes lexicaux appartenant à des classes sémantiques de plus en plus fines et des combinaisons syntaxiques spécifiques.

G.G. a étendu cette analyse à la prédication de second ordre. C'est ainsi qu'il s'est intéressé aux relations logiques comme la finalité ou la causalité (G. Gross, 2004, 2010). Dans son ouvrage sur la finalité (G.G. & M. Prandi, 2004), il introduit avec Michele Prandi une nouvelle dimension : « l'ontologie naturelle spontanée » partagée par tous les locuteurs, qui constitue

un système de présupposés soustraits en même temps au doute, à l'épreuve et même à l'expression explicite, auxquels la communauté des humains fait confiance comme le promeneur qui marche fait confiance à la solidité du terrain [...] un système de présupposés d'ordre pratique qui fondent la cohérence de notre comportement spontané avant même de fonder la cohérence de nos concepts et de nos discours

(G.G. & M. Prandi, 2004 : 13)

C'est sur cette base ontologique que s'élabore le jugement téléologique. Partant de là, les deux auteurs procèdent à la description de l'expression de la finalité en ayant recours aux mêmes choix méthodologiques, mais en y intégrant les relations transphrastiques. C'est ainsi qu'une nouvelle approche est proposée pour analyser les phrases comportant des prédicats de second ordre où les arguments sont eux-mêmes des prédicats. En mettant en parallèle les structures conceptuelles et les structures sémantiques, les deux auteurs montrent que le codage de la relation peut être complet :

*Bien qu'il se soit levé tôt, Luc a raté le train* (G.G. & M. Prandi, 2004 : 36),

où le relateur, ici de concession, prend en charge explicitement la nature de la relation qui existe entre les deux arguments prédicatifs. À l'opposé, l'on a un sous-codage qui augmente la part de l'inférence nécessaire pour interpréter le

contenu de la relation prédicative de second ordre. « L'inférence [...] relaie le codage » (G.G. & M. Prandi, 2004 :37). Le cas extrême, où seul l'inférence prend en charge le contenu prédictif, se concrétise dans la juxtaposition de deux prédictions. Ainsi « le codage linguistique ne peut prétendre au monopole de l'expression, mais présente une option, privilégiée certes, mais non exclusive » (*Ibidem*).

*Il a beaucoup plu ; les routes sont inondées.*

Symétriquement au sous-codage figure le surcodage où « le mot de liaison ne se limite pas à porter à l'expression une connexion transphrastique concevable, et donc inférable, indépendamment de l'expression, mais lui impose un profil sémantique très fin » (G.G. & M. Prandi, 204 : 88).

C'est dans la *Sémantique de la cause* que G.G. (2010 : 1) systématise sa description et met à l'épreuve sa méthodologie : son ouvrage « a pour but de décrire de façon systématique et complète les constructions causales du français », « décrire les relations causales telles qu'elles sont codées dans les langues naturelles » (*Ibidem*, 4). Pour ce faire, il ne procède pas par « des décompositions sémiques ». Il pense que « les informations que l'on déduit de la décomposition peuvent s'obtenir par l'étude de la combinatoire des éléments de la phrase ». Le cadre général est clair : « l'expression de la cause est fonction du sens des arguments sur lesquels elle opère » (*Ibidem*, 6). Le critère en jeu dans la causalité linguistique est la bonne formation des phrases, où le choix du prédicat est tributaire des classes d'arguments compatibles avec lui » (*Ibidem*, 6).

Appliquant son analyse systématiquement à des prédicats de premier ordre, il distingue les causes internes des causes externes, lesquelles relèvent d'une prédication de second ordre, objet de son ouvrage. Qu'il s'agisse d'arguments élémentaires ou d'arguments prédictifs, la méthode est toujours la même : dégager les classes sémantiques des arguments du prédicat, d'abord selon le filtre des hyperclasses, puis affiner l'analyse sémantique pour obtenir des sous-classes, les classes d'objets, que G.G. (2010 : 24–25) définit de la manière suivante : « Une classe d'objets est un ensemble sémantiquement homogène de substantifs, décrivant une position argumentale donnée et qui permet de séparer les différents emplois d'un prédicat ». Appliquée à l'expression de la cause, cette méthode permet de dégager des « relateurs de cause qui ont des arguments de nature événementielle, métaphorique, ou pas ». S'y ajoutent les causes aspectuelles, celles du *faire*, les causes des états et celles qui ont une dimension énonciative.

Cette description systématique est de nature empirique : elle est adossée à des corpus conséquents qui permettent d'établir la liste des arguments appropriés pour chaque prédicat de cause, avec mention de sa fréquence. C'est cette

notion d'appropriation qui nous servira de fil conducteur pour faire le lien entre la méthodologie adoptée par G.G. (2010 : 19) et le phénomène du figement en général, notamment un type particulier de locutions, celles qui expriment des relations logiques. Il montre par exemple que « la forme appelée *locution* n'est qu'une des variantes morphosyntaxiques des constructions de prédicats de second ordre ». Ainsi figement et variation sont-ils intimement liés dans le cadre de l'analyse prédicative. L'appropriation des emplois, conçue dans le cadre de la phrase, ne se limite pas aux éléments élémentaires constitutifs de la phrase, il comporte également la catégorie adverbiale.

Toujours en rapport avec les problèmes de sémantique, G.G. s'intéresse à ce qu'il est convenu de considérer comme un sens opaque. L'intérêt méthodologique réside dans la cohérence avec laquelle il inscrit le phénomène de l'opacité dans la combinatoire et la construction de la phrase. En d'autres termes, il résout le problème en termes de classes sémantiques ou d'unités lexicales appropriées. Ainsi fournit-il cette définition de la métaphore, dont il va se servir pour analyser les causes métaphoriques. En illustrant son analyse par l'exemple des classes d'objets de l'<argent> et des <liquides>, il considère que « le processus de la métaphore consiste en ce qu'une classe d'objets, ici celle de l'<argent>, accapare et pour ainsi dire « vole » des prédicats strictement appropriés à une autre classe d'objets, celle des <liquides> » (G.G., 2010 : 93). Cette opacité sémantique est elle-même source d'un grand nombre de collocations dont G.G. cite : *baigner dans l'argent, l'argent coule à flots, drainer de l'argent, verser de l'argent*, etc.

Une dernière remarque en rapport avec l'analyse sémantique que G.G. (2010 : 10) mentionne à propos de l'inférence maximale : la nécessité d'une interprétation sur « la base de notre expérience ». Il enchaîne à ce propos : « ce cas de figure, d'une grande fréquence dans le discours quotidien, va à l'encontre de l'idée reçue que la langue doit être étudiée en elle-même et pour elle-même, sans recours à la connaissance du monde ». La métaphore est un des outils mis au service de la catégorisation du monde.

Connaissances du monde, lexique, combinatoire syntaxique, tout concourt à la production du sens et à fixer dans l'extrême variation des emplois ceux qui sont les plus appropriés, et par conséquent ceux qui forment des assemblages syntagmatiques solidaires : c'est ce que G.G. (2010 : 37) considère comme des « co-occurents [...] spécifiques ». Lister ces co-occurents permettrait de dégager des formations polylexicales solidaires.

## 4. Le processus du figement

Son ouvrage *Les expressions figées en français* (1996) demeure un ouvrage clef dans la littérature consacrée au phénomène du figement : dans un style accessible à tous, G.G. fournit une synthèse sur la question tout en la situant dans le cadre de sa méthodologie. Pour en saisir la portée, il faut l'intégrer dans toute l'œuvre de G.G. Dès la première page, il rappelle les deux principes opposés dans les langues posés par Otto Jespersen (1971) : « La liberté combinatoire et le figement ». Et d'ajouter : « Mettre ainsi, d'entrée de jeu, le figement sur le même plan que la notion de règles était une grande innovation » (G.G., 1996 : 3).

Partant de là, G.G. essaiera de montrer que le fonctionnement des langues repose sur ces deux principes qui s'inscrivent dans un continuum et que la tâche consiste à procéder à des descriptions fines dont l'objectif est de montrer que le figement est un phénomène scalaire qu'il est possible de décrire sur les plans formel et sémantique.

Qu'il s'agisse de combinatoire libre ou de combinatoire figée, la méthode est toujours la même : l'analyse prédicative. Si l'analyse donne lieu à des classes d'objets, l'on peut opérer l'ensemble des transformations qui participent à la détermination des prédicats et à la distribution de leurs classes d'arguments. S'il n'est pas possible de dégager les classes d'objets, c'est la preuve que la séquence fait l'objet d'un figement dont il faut déterminer le degré. Appliquant cette analyse aux locutions verbales, G.G. (1996 : 80) précise qu'« elles ont, en position d'argument non pas des classes d'objets mais des 'éléments' isolés, qui ne sont pas susceptibles, d'aucun paradigme : on ne peut les remplacer ni par un synonyme, ni par un antonyme ». Aussi paradoxal que cela puisse paraître, il s'agit de décrire le figement en termes de combinatoire libre, cette dernière autorisant l'émergence de paradigmes et des reformulations correspondant à la même signification. Le figement n'est pas une question de configuration de « surface ». Ainsi donne-t-il l'exemple de ces deux suites ayant la même constitution sans toutefois avoir le même comportement combinatoire : *avoir faim* et *avoir froid* (*Ibidem*, 74–75). Si la première accepte toutes sortes de variations, la seconde les rejette : déterminations (*Luc a une faim de loup* / \**Luc a un froid de canard*), emploi du possessif (*sa faim* / \**son froid*), etc. La raison en est l'appartenance de *faim* à la classe des prédicats nominaux. D'où les emplois des verbes supports comme *avoir* pour les actualiser ; ce qui n'est pas le cas de *froid*. *Avoir froid* est donc une séquence figée.

G.G. montre dans cet ouvrage que le figement transcende tout dans les langues. C'est pourquoi il le fait sortir de ce qui est habituellement considéré comme locution pour les séquences non nominales et les noms composés consacrés par

la tradition. En définissant la locution en tant que « groupe dont les éléments ne sont pas actualisés individuellement » (*Ibidem*, 14), il étend l'analyse à toutes les parties du discours et au-delà : noms composés, déterminants, locutions verbales et locutions prépositives et conjonctives. Avec ces dernières, son analyse annonce les deux ouvrages qu'il publiera par la suite sur la finalité et la cause.

Pour ce faire, il emploie l'ensemble des critères du figement qu'il expose dans le premier chapitre dont nous retenons particulièrement la polylexicalité et l'opacité sémantique que nous développerons par la suite. Pour le reste, il applique des tests généraux comme l'insertion et la substitution paradigmatique, auxquels il ajoute les transformations spécifiques à chaque type de construction. L'analyse relativement détaillée des noms composés et des locutions verbales le conduit à l'idée de moule, de schème ou de type.

Il peut paraître incongru d'évoquer les notions de *moule*, de *schème* ou de *matrice*<sup>5</sup> quand on évoque un phénomène comme le figement qui échappe totalement à la volonté des locuteurs. Comme il s'agit d'un processus inhérent au fonctionnement des langues, venant faire le contrepois de la liberté de la production langagière régie seulement par des contraintes sous forme de « règles », poser la notion de moule, ou ses équivalents, c'est supposer qu'elle joue le rôle de pendant naturel pour la règle dans la combinatoire libre. Ainsi aurions-nous une production langagière libre régie par des règles et une production langagière figée régie par des moules. Ce point de vue a le mérite de mettre au cœur de l'analyse linguistique la dynamique de la production langagière, et par conséquent celle du système linguistique ; ce qui milite en faveur de la thèse que la langue naît dans le discours et le discours nourrit la langue. Si l'on admet un tel point de vue, l'on pourrait saisir la pertinence de la notion de moule avancée par G.G.

Il en parle dans son ouvrage à plusieurs reprises. Ces occurrences se présentent comme suit :

- « 205 moules de formation adjectivaux » (p. 43) ;
- « Nous sommes en présence d'un *moule* de production d'une partie importante des adverbes complexes ; la syntaxe en est régulière » (p. 114) ;
- « Les séquences Prép (Det) N constituent des *moules* d'adverbes complexes » (p. 116) ;
- « [...] un grand nombre d'adverbes reposent sur des *schèmes* productifs » (p. 117) ;
- « [...] ils [les adverbes complexes] peuvent être générés sur la base de structures susceptibles de former des *moules* productifs » (p. 117).

---

<sup>5</sup> Cf. également Jean-Claude Anscombe qui parle de « schèmes et de matrices » à propos du figement (2011) et Lichao Zhu (2022).

Quand il s'agit des noms composés, il emprunte à Michel Mathieu-Colas (1996) la notion de *type*. Évoquant la typologie des noms composés qu'il a établie, il rappelle qu' « elle comprend plus de 700 types et permet de rendre compte de l'ampleur du phénomène » (p. 48). Trois remarques sont à faire à propos de ces occurrences :

- Elles névoquent pas la notion de structure, très fréquente dans les analyses effectuées dans le cadre du Lexique-grammaire, notion de nature strictement syntaxique ;
- Elles sont associées systématiquement aux notions de production et de génération des séquences figées ;
- Elles ne se limitent pas à une seule partie du discours, même si leur emploi est associé dans ces citations uniquement à l'adjectif et à l'adverbe.

À la lumière de ces précisions, l'on peut supposer que G.G. cherche à appréhender quelque chose qui dépasse les structures syntaxiques et la matérialité lexicale. Mais avant de voir en quoi consiste son intuition, rappelons que dans d'autres cadres théoriques, on a essayé de préciser cette entité qui se situe en deça des énoncés libres et au-delà de l'unité polylexicale. Partant du jeu impliqué dans les séquences à signifiant pluriel entre la fixité et la variation, Jean-Claude Anscombe définit la *matrice lexicale* comme un schéma comportant des unités linguistiques fixes et des variables linguistiques :

- a) « Les unités linguistiques fixes sont des éléments grammaticaux.
- b) Les variables linguistiques représentent des éléments lexicaux.
- c) Le schéma est productif, au moins dans une certaine mesure.
- d) Les contraintes régissant l'instanciation des variables proviennent uniquement de la structure elle-même, et des relations préexistant entre les unités lexicales servant cette instanciation » (J.-C. Anscombe, 2011 : 25).

Le même auteur évoque cette matrice pour les proverbes et en parle dans ce contexte en rapport avec les « noms composés – et éventuellement d'autres phénomènes » (J.-C. Anscombe, 2011 : 25). L'on peut considérer ces précisions comme une avancée dans l'appréhension de ce genre de phénomène encore imprécis, mais dont la discrimination serait très précieuse pour expliquer ces phénomènes de génération à partir de formes sémiotiques assez peu connues jusque-là. Retenons de ce que J.-C. Anscombe avance les trois éléments suivants : jeu entre variance et invariance, productivité et opposition entre lexical et grammatical. Les autres éléments avancés plus haut concernent beaucoup plus les noms composés.

Par ailleurs, il faut rappeler les autres pistes explorées dans d'autres cadres théoriques en rapport avec la phraséologie en général et le figement en particulier. Certains opposent par exemple les collocations aux colligations. Ces dernières

relèveraient d' « un phénomène de comportement grammatical préférentiel, de la même manière que la collocation est un phénomène de comportement lexical préférentiel » (D. Legallois & P. Gréa, 2006 : 11). Ce genre d'extension a introduit dans le débat la notion d'*entrenchment* qui « désigne un processus de stockage (storage) de l'information [...] si l'on peut stocker en mémoire des syntagmes figés (par exemple *pomme de terre*), on peut stocker des règles (flexionnelles, dérivationnelles, phonologiques, syntaxiques). De ce fait, l'*entrenchment*, contrairement à la notion de figement, est susceptible de s'appliquer à toute structure mémorisable par l'esprit et dont le stockage sera facilité par une fréquence importante » (*Ibidem*, 20). Cet élargissement implique toutes sortes de phénomènes comme « les règles grammaticales régulières », « les cadres sémantiques » (frames), « les métaphores conceptuelles », etc. Ce qui conduit à l'idée qu' « une construction est une généralisation des caractéristiques de la phraséologie » (*Ibidem*, 27).

Mais derrière tous ces phénomènes, rattachables d'une manière ou d'une autre au figement, se profile l'idée de *moule* ou de *matrice*, que L. Zhu (2022) a appliquée à un corpus lexicographique, ce qui lui a permis de récupérer automatiquement des phraséologismes. Dans le même cadre, la discussion a porté sur la dénomination du phénomène : les deux termes retenus sont *moule* et *matrice* ; le moule rend compte de l'aspect invariant, fixe, rigide, alors que *matrice* suggère la souplesse, la variation et surtout la production et la génération. « Le *moule* serait à la *matrice* ce que la *fixité* est à la *variation*<sup>6</sup> » (S. Mejri, 2023). La jonction des deux, avec ce qu'ils représentent chacun de son côté comme avantage, donnerait, comme indiqué en note, une nouvelle unité : l'*emmème* qui, « tout en étant partiellement fixe, est en même temps très souple, représentant ainsi un type sémiotique particulier qui se distingue [...] par :

- Sa nature syntagmatique, en tant qu'espace où s'exercent les deux forces de la dynamique [langagière] ;
- Son caractère hybride puisque la fixité et la variation peuvent, selon le cas, toucher n'importe quel aspect linguistique ;
- Sa discontinuité structurante sans laquelle il n'y aurait pas de distribution entre segments de l'espace saturés et segments à saturer » (*Ibidem*).

Cette entité, certes difficile à isoler, découle de l'observation de l'extrême imbrication entre combinatoire libre, analysée par G.G. au moyen des classes d'objets, et de l'ensemble des contraintes transformationnelles spécifiques, et la combinatoire

---

<sup>6</sup> Moule-matrice ou matrice-moule donnerait, par siglaison, M.M. (prononcé *em em*). D'où la proposition du terme *emmème*, qui présente l'avantage de s'intégrer dans le paradigme des unités telles que *phonème*, *morphème*, *lexème*, *grammème*, *phrasème*, etc. (S. Mejri 2023).



figée dont la gradation est de plus en plus fluctuante. Comme elle est au croisement de l'ensemble des dimensions du système linguistique, elle relève nécessairement du système symbolique qu'est la langue. C'est pourquoi il serait raisonnable de lui donner un statut sémiotique qui, vraisemblablement, se situe à l'interface de la langue et de la production langagière, cette zone où toute la dynamique des interactions entre les deux pans du langage prend forme. Une telle entité expliquerait l'automatisme de la production langagière, notamment chez les natifs, et l'extrême rapidité avec laquelle les enfants acquièrent leur langue naturelle. Une base neuronale est attribuée à l'*entrenchment* : « dans un réseau de neurones, les liens entre neurones sont pondérés et ces pondérations s'ajustent en fonction d'un algorithme d'apprentissage. Or cet algorithme est justement sensible aux régularités internes des patrons d'activation. Du coup, l'*entrenchment* peut être identifié à l'émergence d'un attracteur. En outre, le caractère gradué de l'*entrenchment* trouve lui aussi une reformulation évidente dans la « profondeur » de cet attracteur. L'*entrenchment* est donc un concept qui se veut neurobiologiquement plausible [...] » (D. Legallois & P. Gréa, 2006 : 20).

À un niveau cognitif beaucoup plus général, les automatismes peuvent être rattachés à la notion d'*heuristique* que Daniel Kahneman (2012 : 113) définit comme « une procédure [de pensée] simple qui permet de trouver des réponses adéquates, bien que souvent imparfaites, à des questions difficiles », qui serait à la base de ce que cet auteur appelle « le système 1 qui gouverne notre mode de fonctionnement cognitif à tous les niveaux et dans tous les domaines. Ce système est « rapide, intuitif et émotionnel ». Il s'oppose à un autre système, système 2, qui, lui, « est plus lent, plus réfléchi, plus contrôlé, plus logique ». Vu sous l'angle de ces deux modes de pensée, notre parole spontanée serait commandée par le système 1, évidemment sous le contrôle du système 2, dont l'intervention devient nécessaire lors de l'échec du premier. Ces deux systèmes correspondraient à l'opposition entre ces deux types de pensée : pensée automatique / pensée non automatique ; « pensée rapide / pensée lente » ; « pensée intuitive / pensée délibérée » (*Ibidem*).

Comme on vient de le voir dans cette section, la réflexion de G.G. sur la notion de moule, même si elle a un caractère encore intuitif dans son ouvrage, trace la voie devant une réflexion qui dépasse le cadre du figement pour embrasser des phénomènes beaucoup plus généraux qui ne cessent d'occuper suffisamment d'espace dans les débats relatifs au fonctionnement du langage. Un autre concept, qui joue un rôle central dans l'ouvrage de G.G., est celui de *polylexicalité*. Lui aussi témoigne d'une grande richesse heuristique conduisant, tout comme le processus du figement, à des problématiques générales toujours en discussion par les linguistes.

## 5. La polylexicalité

G.G. (1996 : 9) fait de la polylexicalité une caractéristique essentielle du figement : « La première condition nécessaire pour qu'on puisse parler de figement est que l'on soit en présence d'une séquence de plusieurs mots et que ces mots aient, par ailleurs, une existence autonome ». Découlent de cette caractéristique l'ensemble des spécificités des séquences polylexicales : elles ont un signifiant multiple dont les constituants sont susceptibles d'avoir par ailleurs leur propre combinatoire ; c'est pourquoi G.G. attire l'attention sur un

fait rarement mentionné. Ils [=les noms composés] relèvent donc d'une analyse à deux niveaux : du point de vue externe, ils ont dans la phrase une fonction de groupe nominal et leur structure interne n'est pas pertinente. En revanche, si on les analyse au regard de leur constitution interne, on observe que ce sont des suites qui n'ont pas la liberté de fonctionnement des groupes nominaux ordinaires [...].

(G.G., 1996 : 27)

Cette double combinatoire renvoie tout naturellement à l'ensemble des unités à signifiant pluriel que la tradition regroupe sous la dénomination de *locution*, notamment quand elles sont de nature grammaticale (locutions prépositives, conjonctives, etc.). G.G., contrairement à la tendance générale qui consiste à rejeter le discrédit sur de telles notions, lui donne une définition fondée sur ses critères propres d'analyse. Pour lui, « on pourrait appeler *locution* tout groupe dont les éléments ne sont pas actualisés individuellement » (*Ibidem*, 14). Une « locution prépositive avec le *désir de* [est] une instance intermédiaire entre une catégorie simple *pour* et un prédicat conjugué (*il avait le désir, il désirait, il était désireux de*) » (*Ibidem*). En d'autres termes, le figement conduit à annihiler toute relation prédicative interne à la séquence figée : « dans une suite donnée, il ne peut y avoir de relation prédicative entre les différents éléments qui sont dans la portée du figement » (*Ibidem*, 15). Sur le plan sémantique se pose la relation entre les significations des constituants et la signification globale de l'unité polylexicale (cf. section suivante).

Comme on le constate, G.G. met le doigt sur le point nodal du figement : les SF ont « la particularité qu'[elles] allient l'unité à la pluralité<sup>7</sup> » (*Ibidem*, 27).

<sup>7</sup> La problématique de l'un et du multiple est une problématique philosophique générale qui transcende la totalité de l'existant, comme le souligne Alfred North Whitehead (2004 : 71-72) dans ce passage : « je suggère de partir de l'idée de deux aspects de l'Univers :

– L'univers inclut un *facteur d'unité*, impliquant dans son essence la liaison des choses : unité de fin et unité de jouissance. Toute notion d'importance se réfère à cette unité ultime.

En d'autres termes, elles « présentent cette contradiction qu'[elles] fonctionnent comme une unité alors qu'[elles] sont constituées de plusieurs éléments lexicaux, que leur sens soit transparent ou opaque » (*Ibidem*, 28). Cette tension entre l'un et le multiple trouve sa solution chez G.G. dans « la constatation qu'une structure est d'autant plus libre que le nombre de relations linguistiques entre les différents composants est élevé. Quand il n'y a aucune relation syntaxique entre les différents éléments, nous disons que la structure est totalement figée » (*Ibidem*, 98).

Toujours est-il que la problématique de l'unité lexicale dans l'analyse linguistique demeure entière. Tout en reprenant tous les éléments de la question, il ne répond pas explicitement à la problématique du mot par rapport à ces unités plurielles : faut-il les considérer comme des unités simples et leur attribuer le statut de mot, avec tout ce que cela implique comme conséquences à la fois de description et d'analyse ? Si une telle solution est peu satisfaisante, y aurait-il une entité lexicale supérieure qui engloberait à la fois les mots (simples, construits par dérivation affixale ou tout autre procédé) et les unités polylexicales ? Si la réponse est positive, où la situer par rapport au système linguistique ?

Nous avons un début de réponse chez G.G. aussi bien dans son ouvrage sur le figement (1996) que dans son *Manuel d'analyse linguistique* (2012) où il a recours à l'équivalence fonctionnelle comme critère pour mettre sur le même pied d'égalité unité polylexicale et unité simple, c'est-à-dire le mot. Il n'hésite pas non plus à attribuer la dénomination de mots à ces unités de signifiant pluriel en employant des termes hérités de la tradition comme *nom composé* et *mot composé*. Il étend cet emploi même aux adjectifs et déterminants en parlant d'*adjectifs* et de *déterminants composés*. Ailleurs il a recours au terme *locution* en intitulant par exemple le chapitre VI de son ouvrage sur le figement « locutions adverbiales », même s'il précise au paragraphe 3 de ce même chapitre : « nous appelons *adverbes complexes* ceux qui sont constitués de plusieurs mots, donc ceux qui sont polylexicaux » (G.G., 1996 : 113). Se profilent derrière cet usage terminologique les idées suivantes :

- L'idée d'unité de fonctionnement de ces unités polylexicales ;
- L'idée qu'elles s'intègrent, tout comme les unités simples, dans les paradigmes des parties du discours concernées et en héritent l'ensemble des contraintes combinatoires ;

- 
- L'univers inclut aussi, un *facteur de multiplicité*, tout aussi fondamental : il y a de multiples actualités, chacune ayant son expérience propre, jouissant individuellement, et cependant se requérant l'une à l'autre.

Toute description de l'unité exigera les multiples actualités, et toute description de la multiplicité exigera la notion d'unité, dont dérivent celle d'importance et celle de dessein. En raison de l'indivisibilité essentielle des multiples réalités, il se produit des conflits entre les réalisations finies ».

- Le fait qu'il « reste le problème des séparateurs entre les différents éléments lexicaux, qu'on ne doit pas réduire à un simple problème de graphie » (G.G., 2012 : 198).

Avec cette dernière remarque, G.G. pose le problème fondamental de la continuité ou de la discontinuité du signifiant de l'unité lexicale. Comme il l'a souligné, cela pose des problèmes orthographiques et des difficultés pour le traitement automatique, mais l'idée fondamentale demeure que cette discontinuité dans le signifiant, héritée de la polylexicalité, conduit directement à la problématique du *mot* (Mejri, 2009), laquelle ne peut être résolue que si l'on s'interroge sur le statut du mot dans le système de la langue.

Sans faire le détour par l'abondante littérature sur la question (cf. en particulier A. Martinet, 1965, 1967 et 1968), nous retiendrons ces quelques citations qui nous serviront de jalons dans notre réflexion :

- a) « Il faudrait chercher sur quoi se fonde la division en mots – car le mot, malgré la difficulté qu'on a à le définir, est une unité qui s'impose à l'esprit, quelque chose de central dans le mécanisme de la langue »

(F. de Saussure, 1973 :154)

- b) « Forme et sens doivent se définir l'un par l'autre et ils doivent ensemble s'articuler dans toute l'étendue de la langue. Leurs rapports nous paraissent impliqués dans la structure même des niveaux et dans celle des fonctions qui y répondent, que nous désignons comme « *constituant* » et « *intégrant* ».

Quand nous ramenons une unité à ses constituants, nous la ramenons à ses éléments formels [...] l'analyse d'une unité ne livre pas automatiquement d'autres unités [...] Ainsi en opérant une analyse d'unités linguistiques, nous y isolons des constituants seulement formels.

Que faut-il pour que dans ces constituants formels nous reconnaissons, s'il y a lieu, des unités d'un niveau défini ? Il faut pratiquer l'opération en sens inverse et voir si ces constituants ont une fonction intégrante au niveau supérieur. *Tout est là : la dissociation nous livre la constitution formelle ; l'intégration nous livre des unités signifiantes* (c'est nous qui soulignons). Le phonème, discriminatoire, est l'intégrant, avec d'autres phonèmes, d'unités signifiantes qui le contiennent. Ces signes à leur tour vont s'inclure comme intégrants dans des unités plus hautes qui sont informées de signification. Les démarches de l'analyse vont, en direction opposée, à la rencontre ou de la forme ou du sens dans les mêmes unités linguistiques.

Nous pouvons donc formuler les définitions suivantes :

- La *forme* d'une unité linguistique se définit comme sa capacité de se dissocier en constituants de niveau inférieur.

- Le *sens* d'une unité linguistique se définit comme sa capacité d'intégrer une unité de niveau supérieur.

[...] Leurs rapports mutuels se dévoilent dans la structure des niveaux linguistiques, parcourus par les opérations descendantes et ascendantes de l'analyse, et grâce à la nature articulée du langage »

(E. Benveniste, 1966 : 127–129)

- c) « [...] Guillaume définit la genèse du mot comme un « acte essentiel de l'esprit » (L S2 99). La raison en est que le mot joue un rôle clef dans la construction du système de représentation qu'est la langue, car il est l'expression d'un équilibre entre la phrase et la pensée profonde. « Le mot est un être qui regarde de deux côtés à la fois : 1 du côté de la phrase vers laquelle il est appelé et qui le sollicite en quelque sorte d'entrer en elle ; 2 du côté de la pensée profonde, et c'est à ce regard que le mot doit l'universalisation qui en fait une catégorie finale d'entendement, une partie du discours [...] » « Le mot du français est un mot biphasé, formé par la consécution de deux périodes : une période de particularisation croissante menant à l'idée particulière visée, et une période de particularisation décroissante menant *in fine* à l'universalisation déterminée et *intégrante* que constitue la partie du discours » (LL1 049).

« Une remarque de portée générale, non dénuée d'intérêt en linguistique structurale, c'est qu'un morphème à double effet ne constitue jamais un mot indépendant en langue [...] ».

Rappelons la place privilégiée que Guillaume accorde au mot dans sa théorie du langage, la structure du mot conditionnant la structure de la phrase : « un mot, par constitution, apporte avec lui ses possibilités associatives en phrase. De sorte que la structure de la phrase apparaît conditionnée, et jouée par la structure du mot » (LL2 30)

(A. Boone & A. Joly, 2004 : 281–282)

De ces longues citations, rendues nécessaires par l'acuité du sujet, l'on peut retenir trois moments dans la réflexion à propos des différentes articulations du langage :

- Celui de Saussure qui pose l'existence du mot en tant qu'unité centrale dans un mécanisme de la langue, certes non encore définie, mais dont il faut chercher le fondement ;
- Celui de Benveniste qui fournit la méthode d'analyse qui préside à la détermination des différentes unités du langage, descendante et ascendante, croisant forme et sens, constituance et intégration ; *la forme* étant leur capacité de se dissocier en constituants de niveau inférieur » (constituance), *le sens* leur « capacité d'intégrer une unité de niveau supérieur ». Il est à préciser que

Benveniste n'évoque pas le mot, mais il prévoit clairement dans son analyse une place à une unité dont le morphème est intégrant : « Le phonème, discriminatoire, est l'intégrant, avec d'autres phonèmes, d'unités signifiantes qui le contiennent [=des morphèmes]. Ces signes à leur tour vont s'inclure comme unités intégrantes dans des unités plus hautes qui sont informées de signification ». On peut supposer que les « unités plus hautes » soient des mots, puisqu'il considère que « du phonème on passe au *signe*, celui-ci s'identifiant selon le cas à une forme libre ou à une forme conjointe (morphème). Pour la commodité de notre analyse, nous pouvons négliger cette différence, et classer les signes comme une seule espèce, qui coïncidera pratiquement avec le *mot*. Qu'on nous permette, toujours pour la commodité, de conserver ce terme décrié et irremplaçable » (*Ibidem*, 123). Mais le mot, tel qu'il est présenté ici est une « forme libre », la « forme conjointe » étant considérée comme un morphème. Nous pensons que l'auteur voit dans le mot une forme à part, à distinguer du morphème, puisqu'il le définit ainsi un peu plus loin (p. 124) : « Le mot peut donc se définir comme la plus petite unité signifiante libre susceptible d'effectuer une phrase, et d'être elle-même effectuée par des phonèmes ». Parmi les mots, il opère la distinction entre « mots *autonomes*, fonctionnant comme constituants de phrases (c'est la grande majorité), et mots *synnômes* qui ne peuvent entrer dans des phrases que joints à d'autres mots : ainsi en français *le* (*la...*), *ce* (*cette...*) [...] » (*Ibidem*, 124).

- Cette autonomie est explicitée par Guillaume qui y voit l'ingrédient essentiel qui, à notre avis, distingue le mot du morphème : la partie du discours qui donne forme au contenu lexical pour en faire un nom, verbe, adjectif, adverbe, etc. ; ce qui permet le passage du morphème, unité minimale dotée d'un sens et la phrase : un morphème, à lui seul, ne peut constituer une phrase que s'il possède l'autonomie que lui confère sa partie du discours. C'est grâce à cette caractéristique formelle que le mot peut être considéré comme l'unité prototypique d'une troisième articulation du langage, pouvant dans certains cas, si l'on se limite à l'analyse en niveaux, faire l'économie de la phrase : un mot peut couvrir tout le spectre des niveaux, du phonème à la phrase. Maurice Pergnier l'illustre parfaitement par l'exemple de *où* ? [u], qui est à la fois un phonème, un morphème, un mot, une phrase (1986 : 28–29). Il précise à juste titre que « le fait linguistique est en effet un donné *stratifié* comportant toujours plusieurs strates, sans lesquelles il n'existe pas. La totalité de ces strates est présente dans tout énoncé linguistique si inanalysable qu'il soit (ou qu'il paraisse) du point de vue de l'axe syntagmatique » (p. 28).

Ainsi une opposition fondamentale dans les hiérarchies s'impose : « hiérarchie de niveaux » et « hiérarchie des constituants » (*Ibidem*). La confusion entre ces

deux types de hiérarchie crée un angle mort qui sacrifie les unités de la troisième articulation du langage, qui reprend tout naturellement sa place grâce à l'intégration des unités linguistiques les unes aux autres pour faire émerger à chaque fois des unités dont la pertinence (le sens) participe à l'élaboration d'unités appartenant au niveau supérieur :

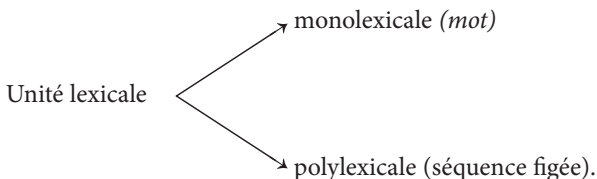
phonème → morphème → unité de la 3<sup>ème</sup> articulation → phrase.

L'exemple de [u] dans l'échange fourni par M. Pergnier : « – On se retrouvera à cinq heures, d'accord ? – Où ? » a le mérite de neutraliser les relations de constitution, c'est-à-dire les relations syntagmatiques.

Grâce aux travaux sur le figement, et conséquemment sur la polylexicalité, la question du *mot* se pose en d'autres termes : faut-il considérer les unités polylexicales comme des mots ou non ? Puisque l'autonomie de départ des constituants de ces séquences est annihilée à l'intérieur de l'espace couvert par le figement, l'emploi du *mot* se trouve justifié. Quand G.G. a recours aux deux termes de la tradition (*mot* ou *nom*) *composé* et *locution*, il pose de fait l'unité lexicale comme repère d'analyse, c'est-à-dire comme un niveau linguistique distinct du morphème, qu'il définit grâce à son appartenance catégorielle<sup>8</sup>. Si l'on pousse jusqu'à son terme cette analyse, qui rejoint celles de Saussure, Benveniste, Guillaume et Pergnier, l'on peut reprendre telles quelles les définitions du mot avancées par Benveniste et Guillaume, tout en substituant au terme mot celui d'unité lexicale, qu'elle soit monolexicale (*mot*) ou polylexicale (séquence figée) :

- « [L'unité lexicale] peut donc se définir comme la plus petite unité signifiante libre susceptible d'effectuer une phrase [...] » (Benveniste, *Ibidem*, 124)
- « [L'unité lexicale] est un être qui regarde des deux côtés [...] » (Guillaume, *Ibidem*).

Le mot serait une unité monolexicale selon cette répartition :



<sup>8</sup> Son ouvrage *Les expressions figées en français. Les noms composés et autres locutions* (1996) est structuré en chapitres renvoyant aux différentes parties du discours, qui constituent la base de toute son analyse.

Cette richesse heuristique s'exprime également avec la même acuité au niveau d'un autre critère des séquences figées retenu par G.G. : l'opacité sémantique.

## 6. L'opacité sémantique

Dans le premier chapitre de son ouvrage sur le figement, qu'il intitule « la notion de figement », G.G. (1996 : 9–23) examine « les propriétés communes qui caractérisent ce phénomène, que l'on doit considérer comme un des plus importants dans les langues » (p. 9). L'opacité sémantique figure en deuxième position après la polylexicalité. C'est dire l'importance que G.G. lui accorde. Comme elle découle tout naturellement de la polylexicalité, il la définit par la négative, c'est-à-dire par rapport à la compositionnalité : « cela veut dire que le sens d'une séquence est le produit de celui des éléments composants » (*Ibidem*, 11). Si elle ne l'est pas, elle « peut avoir deux lectures possibles : l'une transparente et l'autre opaque ». Il illustre cette situation pour l'exemple *Les carottes sont cuites*, « qui signifie que les légumes en question sont prêts à être mangés (sens compositionnel) ou que la situation est désespérée (sens opaque) [...] Il y a donc des suites dont le sens n'est pas fonction de celui des éléments constitutifs. Dans ce cas, nous dirons que nous sommes en présence d'une suite opaque ou sémantiquement figée et contrainte lexicalement » (*Ibidem*). Cette analyse, appliquée à des suites de niveau inférieur à la phrase, permet à G.G. de préciser que « l'opacité est un phénomène scalaire : elle peut être totale (*la clé des champs*), partielle (*clé anglaise*) ou inexistante (*clé neuve*) » (*Ibidem*).

L'opacité sémantique a pour corollaire les restrictions sémantiques : « ce qu'il faut souligner avec insistance, c'est que l'opacité sémantique et les restrictions syntaxiques vont de pair » (*Ibidem*, 12).

Avant de voir les implications de l'opacité sur l'aspect syntaxique des suites figées, il faut d'abord dégager l'ensemble des termes employés pour couvrir le champ de l'opacité. Cela nous permettrait de suivre les débats dont les principaux termes ont fait l'objet, et les avancées à la fois théoriques et appliquées auxquelles ils ont conduit.

Deux couples antonymiques s'opposent, à savoir d'un côté *sens compositionnel* (compositionnalité) / *sens non compositionnel* (non compositionnalité), de l'autre *sens transparent* (transparence) / *sens opaque* (opacité). Bien que les choses paraissent évidentes pour l'emploi de ces termes pour les habitués du figement, il n'en est rien en réalité dans l'usage qu'en font les linguistes.



Pour y voir un peu plus clair, nous consultons quelques ouvrages de base qui servent de références consensuelles, comme les dictionnaires de linguistique (Dubois, 2012[1994] ; Neveu, 2004) et des manuels comme ceux de Riegel *et al.* (2009[1994]), de Wilmet (2003), de Martin (2014) et de G.G. (2012). Comme chacun des deux couples de termes n'ont pas la même extension d'emploi, il serait pertinent pour la clarté de notre exposé d'écartier certains emplois qui n'intéressent pas le figement. Tel est le cas de l'emploi de l'opposition *transparence / opacité*, employée par les philosophes et les logiciens, notamment Gottlob Frege, et illustré chez les linguistes (Martin, 1992[1983]) ; Wilmet, 2003 ; Neveu, 2004 ; etc.) par les fameux exemples : 1. *Œdipe voulait épouser Jocaste* ; 2. *Œdipe voulait épouser sa mère*, dont l'interprétation dépend de l'univers de croyance de l'interprétant, celui de l'énonciateur qui sait que Jocaste est la mère d'Œdipe et celui d'Œdipe qui ne le sait pas. Ainsi « un contexte opaque (*vs* transparent) est un contexte linguistique dans lequel il est impossible de substituer l'une à l'autre des deux expressions coréférentielles sans modifier la valeur de vérité du contenu propositionnel » (Neveu, 2004 : 210). Un autre emploi, à écartier également, celui qui concerne l'analyse du discours, développé par Dubois *et al.*, qui :

note la présence ou l'effacement de l'émetteur par rapport à son discours et du point de vue du récepteur [...] On peut prendre comme référence d'opacité minimale le cas du livre scolaire [...] À l'opposé, l'opacité sera maximale dans la poésie [...].

(J. Dubois *et al.*, 2009 : 334)

D'autres contextes opaques sont étudiés comme les emplois autonymiques (Martin, 2014 : 11).

Quand il s'agit du figement, le couple des deux oppositions revêt d'autres contenus. Il s'agit de montrer dans quelle perspective ils s'inscrivent, de discuter certains de leurs emplois et de proposer une distribution que nous jugeons plus nuancée parce qu'elle correspond beaucoup plus aux faits linguistiques décrits :

### **a) La dichotomie *Compositionnalité / Non compositionnalité* :**

Elle s'inscrit clairement dans le cadre des relations entre le tout et ses parties, et vice-versa. Comme la langue est régie par les deux principes que sont la liberté combinatoire et le figement, la compositionnalité est le pendant naturel de la liberté combinatoire ; elle se définit par Robert Martin (2002 : 64) en ces termes : « Le sens d'une phrase est la résultante combinatoire du sens de ses parties : l'hypothèse est alors celle de la « compositionna-

lité<sup>9</sup> » ; les éléments se composent dans des ensembles de plus en plus vastes ». Tout se joue dans la relation partie-tout : comme la phrase (le tout) est le cadre dans lequel se détermine l'emploi des unités lexicales (les parties), il s'agit de savoir comment se construit le sens du tout à partir de celui des parties et comment les parties contribuent à l'émergence du sens véhiculé par le tout. Martin (2002 : 65) y voit une « difficulté [...] plutôt technique : comment en effet définir le sens des parties si ce n'est à travers le sens des phrases ? ». Pour lui, la « démarche [...] qui semble beaucoup plus opératoire : on part de la phrase pour aller aux parties et non l'inverse (*Ibidem*). Comme

le sens en soi n'est pas directement observable, et accessible seulement par des signes qui le manifestent, une véritable méthode d'objectivation sémantique [se dégage] : le sens est objectivable, pour peu que l'on parte de la phrase, par des conditions de vérité que les locuteurs compétents sont en état, sinon d'énumérer (c'est le rôle du linguiste), du moins de valider.

(R. Martin, 2002 : 65)

Tout autre est le point de vue de G.G. à propos de la compositionnalité :

La notion de grammaire est fondée sur la liberté combinatoire des mots, le sens d'une séquence étant le produit de celui de ses éléments constitutifs. Cette réalité est connue sous le nom de compositionnalité [...] La liberté combinatoire s'entend comme la possibilité d'utiliser librement les éléments des classes d'objets autorisées par le schéma d'arguments d'un prédicat donné. Le schéma d'arguments constitue l'identité du prédicat, ce que nous avons appelé son emploi. Dès lors qu'une position argumentale correspond à une classe d'objets, on est en présence d'une construction libre.

(G.G., 2012 : 197)

Robert Martin comme G.G. sont d'accord que c'est dans le cadre du tout (la phrase) que se détermine le sens global, mais chacun le décrit dans les termes de son approche : Robert Martin a recours à la vériconditionnalité, G.G. à l'analyse prédicative. Mais l'un comme l'autre pose la compositionnalité comme principe d'interprétation des suites régies par la liberté combinatoire. Il s'ensuit que les deux éléments auxquels on peut ramener la compositionnalité sont d'un côté la relation partie-tout et la combinatoire libre des constituants.

---

<sup>9</sup> Ce terme a deux orthographes, avec ou sans le dédoublement du *n* : Robert Martin l'écrit avec un seul *n*, alors que G.G. recourt au dédoublement de cette consonne, qu'on rencontre par ailleurs dans *compositionnel*.

On l'aura compris : ce sont les mêmes éléments qui entrent dans la définition de la *non-compositionalité* : « Toute langue [...] porte en elle-même une part importante de non-compositionalité (R. Martin, 2014 : 138). G.G. va plus loin :

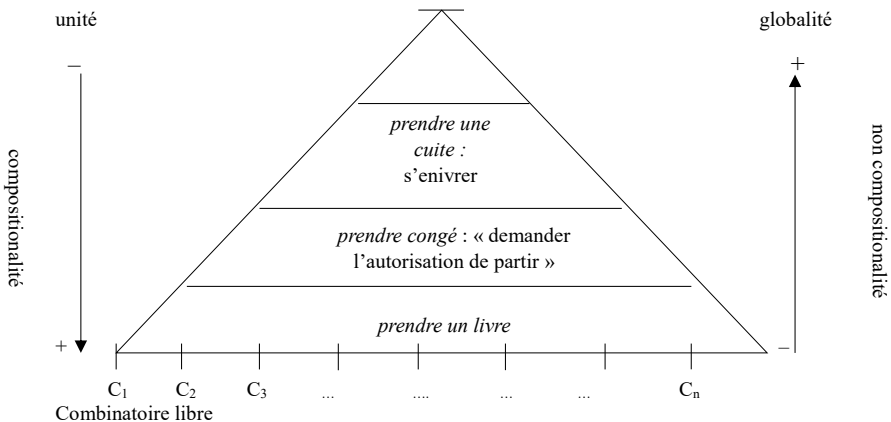
une propriété des langues naturelles est la non-compositionalité d'un très grand nombre de séquences de mots : cette caractéristique affecte toutes les catégories grammaticales et toutes les fonctions syntaxiques. Une description systématique montre que le figement est un phénomène massif qui couvre une part importante de la surface d'un texte.

(G.G., 2012 : 8-9)

Si la compositionalité est le pendant naturel de la liberté combinatoire, son corollaire, la non-compositionalité, s'impose avec la limitation de cette liberté. La conséquence des blocages de la combinatoire entraîne un décrochage sémantique qui, selon le degré de figement, éloigne plus ou moins le sens global du tout de celui des parties constitutives : « le sens d'une expression figée est donc un sens global qui dépasse ce que l'on peut comprendre sur la base des mots qui la composent. On dit qu'il est *idiomatique, non compositionnel* » (Abeillé *et al.*, 2021 : 249). Les exemples les plus prototypiques sont ceux où il y a une rupture totale entre le sens compositionnel et le sens global, non-compositionnel. Parler de *quatrième mur* ne fournit aucun élément sur le contenu global, propre au théâtre : « Le mur virtuel séparant la scène des spectateurs » (GR). C'est pourquoi l'on dit qu'*un acteur brise le quatrième mur*, quand il « s'adresse directement aux spectateurs » (*Ibidem*).

En résumé :

### Schéma 1



C = constituants

## **b) Transparence / Opacité :**

Découle de ce qui précède l'autre dichotomie : transparence / opacité. Nous avons déjà évoqué les emplois du concept d'opacité qui ne nous concernent pas ici, illustrés par l'exemple d'Œdipe ou de celui des usages autonymiques, en mention, des mots de la langue. Ce qui nous intéresse, c'est ce que G.G. (1996 : 11) dit à propos d'une suite comme *Les carottes sont cuites* qui « peut avoir deux lectures possibles : l'une est transparente et l'autre opaque ». Dans le premier cas, elle signifie que les légumes en question sont prêts à être mangés (sens compositionnel) » (*Ibidem*) ; dans le deuxième cas « la situation est désespérée » (*Ibidem*). Et G.G. d'ajouter : « comme on le voit dans la seconde *interprétation*, il y a donc des suites dont le sens n'est pas fonction de celui des éléments constitutifs » (*Ibidem*), c'est nous qui soulignons).

Partant de ces citations, la situation semble se compliquer davantage, puisque l'auteur emploie « sens compositionnel » pour dire qu'il est transparent et qu'il définit l'opacité par la non-compositionalité. Pour démêler cet écheveau, il faut préciser le cadre dans lequel se conçoit cette dichotomie et quelle relation elle entretient avec la première dichotomie.

Si nous avons souligné dans la citation de G.G. le mot *interprétation*, c'est parce que nous cherchons à mettre en saillance le cadre dans lequel s'inscrivent la transparence et l'opacité des suites figées : celui de leur interprétation, c'est-à-dire celui du décodage des séquences. Une telle opération englobe forcément la première dichotomie : une suite est interprétée soit en fonction des règles de la compositionnalité, le corollaire de la liberté combinatoire, soit en fonction de celle de la non-compositionalité, c'est-à-dire selon une fixité sémantique en rupture avec la compositionnalité. C'est la raison pour laquelle G.G. (2012 : 198–199) parle de l' « *interprétation compositionnelle* » et de la « *lecture littérale* ».

Mais quand on se situe au niveau de l'interprétation uniquement, tout en tenant compte de la compositionnalité ou non des séquences figées, on introduit, en plus de la relation partie-tout, le contexte de l'emploi des suites concernées. Ainsi aurions-nous deux sources d'opacification : les relations partie-tout et le contexte d'emploi.

### **I. Les facteurs opacifiants dans le cadre parties-tout :**

Comme « l'opacité est un phénomène scalaire : elle peut être totale (*la clé des champs*), partielle (*clé anglaise*), ou inexistante (*clé neuve*) » (G.G., 1996 : 11), il faut préciser ce que la suite à interpréter comporte comme élément opacifiant,

c'est-à-dire bloquant l'interprétation compositionnelle, pas nécessairement transparente<sup>10</sup>.

Pour en faire l'inventaire, nous en fournirons deux exemples prototypiques :

- La présence de noms propres : *tomber de Charybde en Scylla* ; *le baiser de Judas* ; *baiser Lamourette* ; *franchir, passer le Rubicon*. Dans toutes ces suites, la présence des noms propres bloque l'interprétation compositionnelle parce que le nom propre n'a pas un contenu intensionnel inscrit en langue. « L'opacité qui en résulte n'est interprétable que par des voies fort éloignées de l'explication synchronique » (R. Martin, 2014 :134). C'est dans l'histoire de la langue que l'on peut trouver le sens compositionnel et avoir une interprétation plausible. Ainsi le *GR* nous dit à propos des suites citées :
  - *Charybde* renvoie à « un gouffre situé dans le détroit de Sicile, en face d'un écueil appelé Scylla » ; d'où le sens global : « néchapper à un mal que pour tomber dans un autre pire encore » ;
  - *Judas* dans *baiser de Judas* (ou baiser de la mort), « par allusion au baiser donné par Judas à Jésus, pour le désigner aux soldats romains » ; d'où le sens global : « témoignage d'affection trompeur et perfide » ;
  - *Baiser Lamourette* : « par allusion au rapprochement ébauché par Lamourette à l'Assemblée législative (1792) où des adversaires s'embrassèrent » ; d'où le sens : « réconciliation éphémère » ;
  - *Passer, franchir le Rubicon* : « par allusion à César qui, en franchissant cette petite rivière formant la limite entre la Gaule cisalpine et l'Italie, déclarait au Sénat la guerre civile » ; d'où le sens attribué à la séquence : « prendre une décision irrévocable, passer hardiment à l'action ».
- Les transferts tropiques comme la métaphore et la métonymie :
  - Dans une expression comme *un ange passe*, le lien entre le milieu des couvents<sup>11</sup> et le silence qui s'installe dans une conversation n'est pas direct. Le sens global devient plausible quand on l'établit métonymiquement en faisant le pont entre le passage de l'ange et le silence que cela induit ;
  - On trouve dans les dénominations des plantes, pierres précieuses et autres objets, toute une moisson d'unités polylexicales conçues sur la base du mécanisme métaphorique : *œil-de-bœuf* (lucarne), *cœur-de-pigeon*<sup>12</sup> (variété de

<sup>10</sup> Nous verrons par la suite qu'une séquence peut être de sens compositionnel sans être pour autant transparente et vice-versa.

<sup>11</sup> Le *GR* fait la mention suivante à propos de l'origine de cette suite : « selon G. Sand, cette expression est née dans les couvents ».

<sup>12</sup> En réalité, il s'agit dans ces expressions d'une métaphore doublée d'une métonymie : la dénomination repose sur un rapprochement de similitude, à partir duquel l'on dénomme l'objet (une caractéristique → dénomination du tout).

- cerise). Moins la proximité entre le domaine source et le domaine cible est grande, plus l'opacité est importante.
- Les anciennes croyances: Robert Martin (2002:139) donne *reprenre du poil de la bête* comme exemple non compositionnel, donc opaque, pour dire que: « seule l'histoire peut éclairer ce contenu. Il se justifie sans doute par la croyance que, pour guérir une morsure, il faut y mettre quelques poils de la bête qui l'a produite. On se ressaisit donc en reprenant du poil de la bête ».
  - Le fonds des textes religieux et littéraires: deux exemples suffisent pour illustrer cette origine:
    - *Le chemin de Damas (de quelqu'un)*: « Le lieu et le moment où la vérité lui apparaît ». Allusion à la conversion de Saint Paul, due à une vision alors qu'il se rendait à Damas » (A. Rey & S. Chantreau, 1989: 233).
    - *Montrer patte blanche*: « par allusion à la chèvre de la fable »: montrer un signe de reconnaissance convenu, dire le mot de passe nécessaire pour entrer quelque part » (GR).

## **II. Les facteurs contextuels :**

L'interprétation des séquences opaques ne dépend pas seulement du degré de décrochage entre le sens des constituants et le sens global; elle est également tributaire du contexte d'emploi. Certains contextes agissent sur l'opacité interne à des degrés divers. D'autres le sont moins, d'où une incongruité dans l'enchaînement séquentiel entre la suite opaque et le reste du contexte. Le contexte est soit cotextuel (cotexte), soit situationnel. Ces exemples nous serviront d'illustrations :

- Cotextes :
  - *Béret Savoyard*, employé dans l'exemple suivant :  
*Après le plat de résistance et avant le dessert, on nous a servi un Béret Savoyard parfumé et fruité avec des arômes de noisette*  
 ne présente pas beaucoup de difficultés pour ceux qui en ignorent la signification. L'opacité est alors battue en brèche par les cotextes de gauche et de droite: l'opacification entraînée par les deux facteurs internes que sont la métaphore-métonymie de *béret* et le gentilé *Savoyard*<sup>13</sup>se trouve ainsi éclairée (rendue plus transparente, puisque le contexte, même s'il n'indique pas explicitement qu'il s'agit d'un fromage); l'on comprend qu'il s'agit d'aliments servis à un moment précis du repas;

<sup>13</sup> Il faut ajouter un 3<sup>ème</sup> facteur opacifiant, le caractère exocentrique de la dénomination.

- Au contraire, la même suite figée dans une phrase comme *le béret Savoyard se vend cher* conserve un degré d'opacité plus important, puisque le cotexte ne suggère que l'idée de marchandise : cela pourrait être un béret, comme cela pourrait renvoyer au fromage.
- Contextes situationnels :
  - *Aller à Tataouine*, « aller très loin, au bout du monde » (expressio.fr), employé par un Tunisien, perd toute sa signification opaque, alors qu'elle conserve toute son opacité pour celui qui ne connaît pas la ville de Tataouine, dans le Sud tunisien.

Deux constats sont à retenir :

1. L'opacité et la transparence, tout en impliquant respectivement la (non) compositionnalité comme facteurs, se situent clairement du côté de l'interprétation ; ce qui implique les univers de croyance des locuteurs : toute séquence ne faisant pas partie du stock lexical du locuteur résiste inéluctablement à une interprétation correcte. Cela n'est pas propre aux suites figées : toute unité lexicale se comporte ainsi (cf. R. Martin, 1992, à propos des énoncés scientifiques tels que *Les rutacées sont des phanérogames angiospermes*) ;
2. Les deux dichotomies ne se recouvrent pas parfaitement, puisque l'interprétation dépend des compétences des locuteurs. Ce qui est opaque pour l'un pourrait être transparent pour l'autre et vice-versa. Si l'on y ajoute le caractère scalaire de ce phénomène, l'on aboutit à des configurations où les deux éléments des deux dichotomies peuvent s'opposer. Ainsi aurait-on :
  - des suites figées compositionnelles mais dont l'interprétation demeure opaques pour certains locuteurs, faute d'une maîtrise parfaite des emplois. L'exemple type de ces séquences peut être illustré par les formules telles que *Et avec ceci, (Ce sera tout ?)*, « communément évoqué par les petits commerçants pour diriger la vente vers le paiement » (expressio.fr), qui risque d'échapper à un étranger, surtout pour ce qui est du cadre énonciatif de son emploi ;
  - des suites non-compositionnelles, notamment celles qui sont très courantes, ne posent aucun problème d'interprétation aux natifs : elles sont perçues comme transparentes. *Avoir un chat dans la gorge*, malgré la bizarrerie de l'image, est toute transparente pour un natif.

Une dernière remarque à propos de l'opacité sémantique, très peu évoquée dans la littérature, mais qui est pertinente dans le fonctionnement du système, à savoir son économie générale : le jeu sur les quatre termes des deux dichotomies fournit à la langue une économie extraordinaire qui s'ajoute à celle de la polylexicalité. Si avec la polylexicalité on réemploie des syntagmes ayant par ailleurs une signification littérale pour faire émerger une nouvelle signification

globale, avec l'opacité s'ajoute une nouvelle économie, celle des facteurs opacifiants, qu'ils soient tropiques ou non.

## 7. Les contraintes pragmatiques

Depuis l'ouvrage de G.G. sur le figement, les typologies des séquences figées se sont affinées et ont apporté des éclairages nouveaux sur ce phénomène. La typologie la plus complète, on la doit à Igor Mel'čuk (cf. en particulier 2013: 145). En croisant les deux critères que sont la nature des contraintes (par rapport au sens/ par rapport au message conceptuel), et la (non) compositionnalité, l'auteur dégage trois grandes catégories de phrasèmes (= suites figées) : les locutions, les collocations et les clichés. Dans cette dernière catégorie figurent les clichés ordinaires et les pragmatèmes : le premier est « un phrasème sémantico-lexical compositionnel<sup>14</sup> », (*Ibidem*, 142), le second est « un cliché contraint par la situation de son emploi<sup>15</sup> » (*Ibidem*, 143). L'intérêt théorique d'un tel concept réside dans l'attribution d'un « contenu conceptuel précis – pour l'exemple à *paraître* : « Le texte en question est en train d'être publié » (*Ibidem*) – à une forme linguistique contrainte, dont l'emploi est ancré dans une situation d'énonciation précise. Cette triple contrainte participe à isoler un type particulier de séquence dont la forme est en quelque sorte la contrepartie de sa dimension pragmatique, c'est-à-dire de ses conditions d'emploi<sup>16</sup>. C'est la pragmatique inscrite en langue.

Le pragmatème est un nouveau concept en cours de stabilisation, grâce notamment aux travaux de Xavier Blanco (2013) et de Wiesław Banyś (2018). Nous nous limitons dans ce qui suit à en retenir les principales caractéristiques telles qu'elles sont présentées dans Xavier Blanco et Salah Mejri (2018). Pour éviter de tomber dans le piège de la définition du pragmatème en termes de conditions nécessaires et suffisantes, trop rigide pour rendre compte de l'extrême variation et souplesse du phénomène, nous avons opté pour le choix de l'approche prototypique qui présente l'avantage de conserver l'air de famille partagé par tous les membres sans conduire à l'exclusion ou à la rupture avec la catégorie. Ainsi les pragmatèmes monolexicaux intègrent-ils les conditions suivantes : un signifié restreint, un signifiant polylexical, un sémantisme compositionnel et des emplois ritualisés.

<sup>14</sup> Illustré par *Quel âge avez-vous (as-tu) ?*

<sup>15</sup> Exemple : à *paraître*.

<sup>16</sup> D'autres travaux cherchent à cerner cette dimension pragmatique par d'autres moyens, comme les actes de langage (cf. *les actes du langage stéréotypés* chez M. Kauffer, 2013).

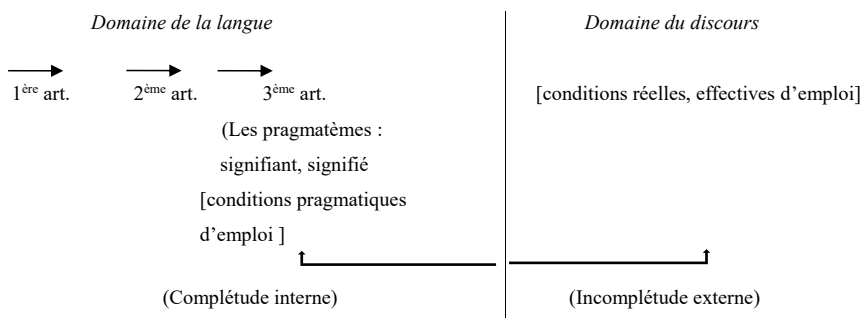


### I. Première caractéristique :

« un pragmatème ne correspond jamais à une partie du discours, il s'agit toujours d'un énoncé complet même quand il est formé d'une seule unité lexicale » (Blanco & Mejri, 2018). Même si son origine se conçoit dans les parties du discours de la langue, il n'en hérite pas le fonctionnement linguistique : contrairement aux autres séquences figées, les pragmatèmes, de par leur complétude, ne sont pas des unités intégrant dans les unités prévues en langue. On ne s'en sert pas pour former des unités supérieures. D'où leur intérêt épistémologique pour la langue fondée sur la triple articulation. Avec les pragmatèmes, l'on a l'exemple prototypique de l'unité de la troisième articulation, c'est-à-dire celle qui atteint la complétude sémiotique exigée pour un signe, fruit des deux articulations, celles des phonèmes et des morphèmes, pour rendre compte d'un signifié ou d'un contenu conceptuel précis, assurant ainsi une communication bien ancrée dans la situation d'énonciation prévue à cet effet. La triple articulation atteint avec les pragmatèmes l'extrême limite des énoncés préconstruits en langue<sup>17</sup>. Leur insertion dans le discours se fait sans la combinatoire<sup>18</sup>. La complétude de ce type d'énoncé concerne l'association entre un signifiant contraint et un contenu conceptuel précis. Pour que le pragmatème assure sa fonction dans la communication courante, il a besoin de l'adéquation entre les conditions énonciatives d'emploi prévues dans sa structure conceptuelle (l'ensemble des contraintes pragmatiques) et la situation effective de son emploi. Sans une telle adéquation, le pragmatème sera amputé de la fonction sémiotique à laquelle il est destiné.

En schéma :

Schéma 2



Il n'en est pas de même des autres unités lexicales intégrables dans des unités supérieures, celles qui appartiennent à des parties du discours. Contrairement

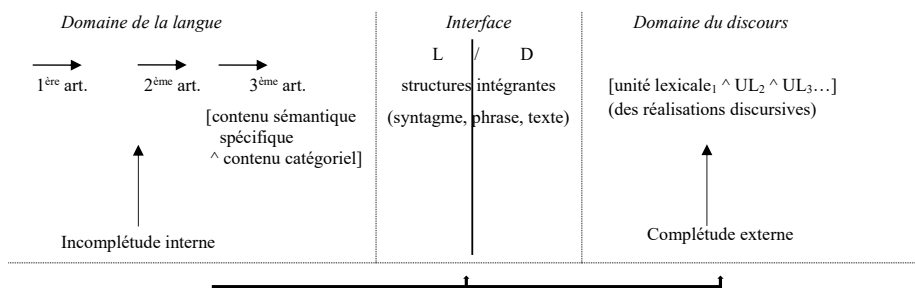
<sup>17</sup> Point qu'ils partagent avec les interjections et les énoncés sentencieux.

<sup>18</sup> Même si des passerelles peuvent exister (cf. X. Blanco & S. Mejri, 2018 : 25–57).

aux pragmatèmes, elles ne se caractérisent pas par une complétude interne, exigeant des règles d'emploi de nature pragmatique (une sorte de grammaire, de syntaxe); elles appartiennent à des paradigmes d'unités (nom, verbe, adjectif, adverbe, préposition, etc.), appelés parties du discours, permettant justement leur intégration dans la phrase. En d'autres termes, elles portent en elles-mêmes une incomplétude à combler dans les réalisations discursives comme la phrase ou le texte.

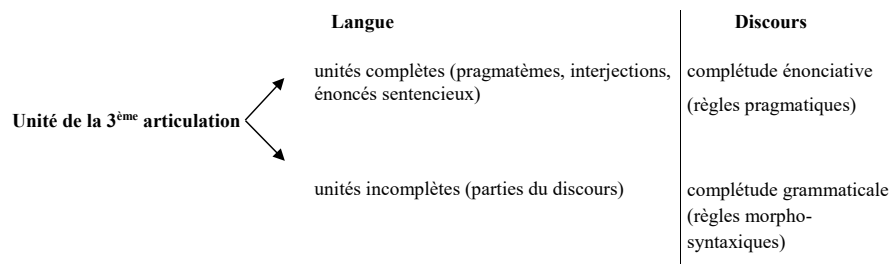
En schéma :

### Schema 3



Avec les unités intégrables, grâce à leurs contenus catégoriels, l'incomplétude dans le discours est comblée au moyen des règles grammaticales déjà prévues dans les unités de troisième articulation. En termes de complétude et d'incomplétude, nous aurons cette distribution :

### Schéma 4



## II. Deuxième caractéristique :

Les pragmatèmes prototypiques sont polylexicaux, sans que cette caractéristique soit exclusive. Les pragmatèmes monolexicaux sont de l'ordre de 5 à 7 % (X. Blanco & S. Mejri, 2018 : 31). La polylexicalité, étant déjà suffisamment développée, nous

nous contentons de retenir dans ce qui suit des éléments ayant une retombée théorique :

- La complétude n'est pas restreinte à la polylexicalité. Des mots simples suffisent à être des pragmatèmes : *Allo !, Halte !, Banco !*;
- L'intérêt de la polylexicalité dans les pragmatèmes réside dans la complétude en langue qui condense les configurations possibles des unités dépassant la taille de l'unité monolexicale, à savoir le syntagme, la phrase et même des constructions dont certains constituants relèvent de la combinatoire (semi-) libre<sup>19</sup>. Les exemples suivants sont empruntés à l'index des pragmatèmes (*Ibidem*, 193–204) :
  - Des exemples de phrases : *Attachez votre ceinture ; Cédez le passage ; Roulez au pas !, Surtout ne dis pas merci, etc.*
  - Des exemples de syntagmes : *Tenue correcte exigée ; Respecter la dose prescrite ; À tes souhaits ! ; Hors service ; etc.*
  - Des exemples à configurations variables : *À la santé de [N] ; [NOM PROPRE] à l'appareil ; Bouchon à [NOMBRE] kms ; Fait à [LIEU], le [DATE] ; etc.*

### **III. Troisième caractéristique, la compositionnalité :**

Le pragmatème prototypique se laisse déchiffrer selon les règles de la compositionnalité, mais cela n'empêche pas d'avoir des pragmatèmes plus ou moins réfractaires à ces règles comme les pragmatèmes antithétiques (*Cause toujours ! ; Tu m'intéresses !*) ou des formules comme *À tes souhaits !* Il ne faut pas confondre évidemment le signifié rattaché aux éléments constitutifs du pragmatème polylexical et le concept pour lequel il a été créé, qui renferme, en plus du signifié, les conditions d'emploi.

### **IV. Quatrième caractéristique, le rituel langagier :**

« Le pragmatème est une forme ritualisée » (*Ibidem*, 30), le rite renvoyant à « une entité très formalisée (*Ibidem*) et le rituel à un type de comportement. Il s'agit donc d'un comportement langagier régi par des conventions qui en restreignent la réalisation. Le comportement a deux faces : l'une est verbale (l'énoncé du pragmatème), l'autre concerne les règles qui codifient l'emploi ritualisé des éléments verbaux : un ancrage énonciatif précis (lieu, temps, prédication...), le mode de réalisation du pragmatème (à l'oral, par écrit, affichage, accompagnement par des

<sup>19</sup> Syntagmes et phrases sont à la fois des possibles de langue, de par la combinatoire des unités lexicales qu'ils comportent, et des possibles du discours, de par la créativité discursive.

codes parallèles : gestes, grimaces...) [...] dimension comportementale (pour les individus) et sociale (régulation des comportements), les sanctions liées au manquement de l'emploi du pragmatème exigé dans la situation d'annonce adéquate, ou à l'inadéquation de la forme employée, etc.» (*Ibidem*, 35). Avec cette dernière caractéristique où le verbal n'est qu'une composante du rituel, l'on touche à la dimension culturelle où l'idiomaticité langagière est au service de l'identité individuelle ou collective. Si les pragmatèmes sont denses culturellement, c'est parce qu'ils dépassent le verbal, réduit à l'une des composantes, pour englober une attitude, une interaction, une posture, etc. qui engagent le locuteur dans ses interactions avec les autres conformément à des règles préétablies<sup>20</sup>. Ainsi pourrions-nous dire qu'avec le pragmatème, les séquences figées atteignent une très grande densité sémiotique qui peut s'analyser selon les différentes strates qu'elles renferment.

## 8. La saillance sémiotique des suites figées

Une caractéristique fondamentale des suites figées, rarement mentionnée, pourtant très prégnante, est leur épaisseur sémiotique. Nous entendons par là la densité des strates qu'elles comportent : tout en partageant avec les autres unités lexicales les trois articulations qui participent de leur genèse, à savoir les niveaux phonémique, morphémique et lexical, elles s'en distinguent par au moins deux autres strates spécifiques, celles qui impliquent la littéralité de la séquence et le sens global correspondant au tout qu'elles représentent.

Un exemple nous servira d'illustration : il montre que la fixité de chaque strate se trouve solidaire de celle des autres strates :

1<sup>ère</sup> strate : /k/œu/r/d/ə/b/œf/ ; /c/œu/r/d/e/b/œu/f/ = unités de la 1<sup>ère</sup> articulation

2<sup>ème</sup> strate : /kœr/də/bœf ; /cœur/de/bœuf/ = unités de la 2<sup>ème</sup> articulation

3<sup>ème</sup> strate : /kœr də bœf ; /cœur de bœuf/ = unité de la 3<sup>ème</sup> articulation.

Il suffit qu'on touche à l'une des unités de chaque articulation pour que tout le reste s'en ressente. Le résultat de chaque opération (substitution, effacement, mutation, etc.) est naturellement le défigement de l'unité poylexicale :

---

<sup>20</sup> C'est par ce biais que d'autres séquences autonomes comme les proverbes ont des emplois pragmatiques.

. /k/œ/r/d/ə/b/œ/f/ ; /c/œu/r/d/e/b/œu/f/ → /s/œ/r/d/ə/b/œ/f/ ; /s/œu/r/d/e/b/œu/f/ = *sœur de boeuf*<sup>21</sup>

. /k/œ/r/d/ə/b/œ/f/ ; /c/œu/r/d/e/b/œu/f/ → /k/œ/r/d/ə/v/a/f/ ; /c/œu/r/d/e/v/a/ch/e/ = *cœur de vache*.

C'est par le biais de la 3<sup>ème</sup> articulation qu'interviennent les interprétations littérales et globales. Cela peut se produire sans qu'on touche à l'intégrité des deux premières articulations. Dans l'énoncé suivant :

Le diable est dans les détails, surtout ceux des contrats d'assurance,

l'expression *le diable est dans les détails* est employée telle quelle sans la moindre manipulation ; une simple anaphore pronominale réactive le sens premier de *détail* en l'appliquant à ceux des *contrats d'assurance* ; ainsi obtient-on une double interprétation : la première, globale, consiste à se prémunir contre les conséquences fâcheuses du manque d'attention ; la seconde applique le *détail* aux contrats d'assurances, le plus souvent trop détaillés et présentés souvent sous un format très condensé rendant leur lecture très difficile. Ce dédoublement est plus flagrant quand la séquence comporte une strate métaphorique. Tel est le cas de l'expression *cœur de bœuf* qui dénomme une variété de tomate sur la base d'un rapprochement avec la forme du cœur. Employée dans cet énoncé, elle fait exploser les interprétations au gré des jeux impliqués par le locuteur :

Il n'aime pas les cœurs de bœuf, il a une préférence pour les meufs de beurs.

Dans cet énoncé, surchargé de préjugés et de sous-entendus peu glorieux pour ceux qui en assument le contenu, *cœur de bœuf* est employé dans un sens global ; ce n'est qu'avec la deuxième phrase qu'une nouvelle interprétation s'impose, mettant sur le même plan des préférences culinaires et des attirances pour des personnes. Avec le jeu chiasmatisé, une sorte de parallélisme s'effectue entre *cœur* et *meuf* d'un côté et *bœuf* et *beur* de l'autre. Il faut y ajouter l'ensemble des connotations associées à l'ensemble des unités monolexicales impliquées dans ces énoncés.

Conclusion : plus la stratification est dense de ces suites, plus leur saillance sémiotique est grande. Plus une séquence est saillante, plus elle donne matière à des défigements.

<sup>21</sup> Les unités de la 1<sup>ère</sup> articulation ont un impact direct sur celles de la 2<sup>ème</sup> articulation, soit en donnant lieu à un morphème nouveau, soit à un segment non attesté (comme *dœur*, *noeur*).

G.G. (1996: 20) souligne à juste titre que « le figement peut être mis en évidence grâce à l'effet provoqué par le jeu du défigement, qui consiste à briser le carcan qui caractérise les suites figées ». Ce passage fait du défigement beaucoup plus qu'un critère de figement ; il attire l'attention sur le carcan (=la fixité) à briser, c'est-à-dire l'ensemble des possibles de discours que la fixité, préconstruite, favorise. Ainsi le défigement répond-il à deux caractéristiques, à la fois distinctes et solidaires : la densité de la stratification de ces signes linguistiques et le degré de fixité. La productivité langagière qui implique l'usage de ces signes se fait de deux manières : selon leur emploi normal, c'est-à-dire dans leur globalité, mais également selon le grand potentiel que représentent les deux caractéristiques déjà mentionnées. Tout ce qui est fixe dans la langue appelle tout naturellement les tentatives de « déstabilisation ». Selon la formule de Gertrud Gréciano (1983), seul ce qui est figé se défige.

Mais peu importe la strate concernée par les manipulations. Tout se joue au niveau lexical, qu'il s'agisse des constituants des suites figées qui reprennent vie individuellement grâce au défigement, ou de la séquence globale, parce que le lexique est l'ultime aboutissement de l'articulation du langage. Avec les unités de la 3<sup>ème</sup> articulation entrent en jeu la combinatoire et tout le potentiel des règles conduisant à l'émergence des idées que véhiculent les productions langagières. Que les règles soient de nature pragmatique ou de nature grammaticale, leur application donne aux lecteurs cette capacité extraordinaire de produire à l'infini des énoncés adaptés à toutes les personnes, situations, événements, états, et prédictions d'une manière générale. Même si le principe de récursivité est le même, les charpentes peuvent prendre une infinité de configurations qui n'ont de limites que des normes sociales qui varient selon les sociétés et les époques.

## 9. Conclusion

L'ensemble des réflexions exposées dans cette contribution reprennent peu ou prou des échanges que j'ai eus avec G.G. lors de nos rencontres régulières et des voyages effectués ensemble<sup>22</sup>. Cela m'a permis d'apprécier la profondeur de la réflexion du linguiste et la pertinence de ses argumentations. Partant de Lexique-grammaire et de la théorie de Harris, il a réussi à faire de la sémantique, grâce à l'analyse prédicative, le cœur de la description linguistique. Formalisée sous la forme de

---

<sup>22</sup> Entre autres en Pologne : à Varsovie, à Cracovie et à Katowice.

classes d'objets, il a pu montrer que l'emploi d'une unité lexicale correspond au schéma d'arguments s'il s'agit d'un prédicat, et à l'ensemble des séries de prédicats auxquels l'unité s'associe s'il est question d'un argument. Le tout consiste à décrire l'ensemble des configurations syntaxiques associées à l'emploi concerné. Brisant la séparation érigée par certaines approches entre lexique, syntaxe et sens, il a su montrer qu'en décrivant les différents emplois, l'on ne peut pas faire l'économie du figement lexical, qui est rarement absolu, mais qui se présente sous la forme d'une gradation plus ou moins importante, que le linguiste doit prendre soin de décrire par le menu s'il cherche à avoir des applications informatiques.

Son ouvrage sur le figement (1996) et son manuel d'analyse linguistique (2012) donnent un condensé sur ce phénomène massif dans les langues ; ils fournissent également à la communauté des linguistes des outils méthodologiques à la fois efficaces pour discriminer les séquences figées et les décrire, et surtout des pistes de réflexion systémiques favorisant l'intégration du figement à part entière dans l'analyse et la description des langues. Cet apport méthodologique est, comme nous l'avons vu, d'une richesse heuristique considérable. L'ensemble des éléments de réflexion de G.G. conduisent vers des interrogations fondamentales, théoriques et appliquées, sur le fonctionnement des systèmes linguistiques en général<sup>23</sup> et sur celui de la langue française en particulier : le figement est le pan symétrique de la langue qui vient s'opposer et compléter celui de la combinatoire libre. L'extrême imbrication des deux est fondée sur le principe universel suivant : tout dans la langue est fait pour signifier en ayant les unités lexicales (unité de la 3<sup>ème</sup> articulation) comme point de convergence de tout le potentiel du système (phonologie, morphologie, grammaire) et comme point de diffraction donnant lieu à une infinité d'énoncés. La dialectique entre les deux régit la dynamique générale du système langagier.

## Références citées

- Abeillé, A., Delaveau, A. & Gautier, A. (2021). *La grande grammaire du français*. Actes Sud.
- Anscombe, J.-C. (2000). Parole proverbiale et structures métriques. *Langages* 139, 6–26.

---

<sup>23</sup> J'ai traduit en arabe avec Béchir Ouerhani l'ouvrage sur le figement (2008) et nous avons pu vérifier la pertinence des éléments retenus dans l'analyse du figement, que nous avons illustrés aisément par des exemples en arabe.

- Anscombe, J.-C. (2011). Figement, idiomatité et matrices lexicales. Dans J.-C. Anscombe & S. Mejri, *Le figement linguistique : la parole entravée* (17–40). Honoré Champion.
- Anscombe, J.-C. & Mejri, S. (éds) (2011). *Le figement linguistique : la parole entravée*. Honoré Champion.
- Banyś, W. (2018). Nouveaux anciens paradigmes : approche orientée Objets, Classes d'objets, Psychologie écologique et Linguistique. *Neophilologica* 30, 25–41.
- Banyś, W. (2020). Pragmatèmes au pays de la prosodie. *Neophilologica* 32, 89–116.
- Benveniste, E. (1966–1974). *Problèmes de linguistique générale, I et II*. Gallimard.
- Bertin-Delbey, A. (1991). Compte rendu de l'ouvrage *Les constructions converses du français, L'information grammaticale* 49, 59.
- Blanco, X. (2013). Microstructure évolutive pour un dictionnaire de pragmatèmes. Dans S. Mejri, I. Sfar & M. Van Campenhandt (éds), *L'unité en Sciences du langage* (139–150). Éditions des archives contemporaines.
- Blanco, X. & Mejri, S. (2018). *Les pragmatèmes*. Classiques Garnier.
- Boone, A. & Joly, A. (2004). *Dictionnaire terminologique de la systématique du langage*. L'Harmattan.
- De Rouilhan, P. (1984). Sujet, prédicat, objet, concept chez Frege. *Histoire, Épistémologie, Langage* 6(1), 91–99.
- Dubois, J. et al. (2012[1994]). *Dictionnaire de linguistique et des sciences du langage*. Larousse.
- Eranckel, J.-J. (éd.) (1989). *La notion de prédicat*. Laboratoire de linguistique formelle de l'Université Paris 7.
- Forsgren, M., Jonasson, K. & Kronning, H. (éds) (1998). *Prédication, assertion, information*. Uppsala University Library.
- François, J. & Legallois, D. (2006). Autour des grammaires de constructions et de patterns. *Cahiers du Crisco* 21, 1–73.
- Grand Robert de la langue française*, version numérique, 2017.
- Gréciano, G. (1983). *Signification et dénotation en allemand. La sémantique des expressions idiomatiques*. Klincksieck.
- Gross, G. (1989). *Les constructions converses du français*. Droz.
- Gross, G. (1996). *Les expressions figées en français*. Ophrys.
- Gross, G. (1996). Prédicats nominaux et compatibilité aspectuelle. *Langages* 121, 54–72.
- Gross, G. (2010). *Sémantique de la cause*. Peeters SLP.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Septentrion.
- Gross, G. (2021). Des perspectives rigoureuses pour la linguistique. *Neophilologica* 33, 1–32.
- Gross, G. & Prandi, M. (2004). *La finalité. Fondements conceptuels et genèse linguistique*. De Boeck supérieur.



- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 7–52.
- Gross, M. (1993). Les phrases figées en français. *l'Information grammaticale* 59, 36–41.
- Guillet, A. & Leclère, C. (éds) (1981). Formes syntaxiques et prédicats sémantiques. *Langages* 63.
- Harris, S. Z. (1973). Les deux systèmes de la grammaire : prédicat et paraphrase. *Langages* 29, 55–81.
- Jespersen, O. (1971[1924]). *La philosophie de la grammaire*. Les Éditions de Minuit. Traduction de M. M. Léonard.
- Kahneman, D. (2012). *Système 1, système 2. Les deux vitesses de la pensée*. Flammarion.
- Kauffer, M. (2013). Le figement des « actes de langage stéréotypés » en français et en allemand. *Pratiques* 159–160, 42–54.
- Le Pesant, D. & Mathieu-Colas, M. (éds) (1998). Les classes d'objets. *Langages* 32(131).
- Legallois, D. & Gréa, P. (2006). La grammaire de construction : un bâtiment ouvert aux quatre vents. *Cahiers du CRISCO* 21, 1–19.
- Legallois, D. & François, J. (éds) (2006). Autour des grammaires de construction et des patterns. *Cahiers du CRISCO* 21.
- Lemaréchal, A. (2022). Des prédicats à perte de vue.... (Ryle 1933) : pourquoi faire ? Ethnocentrisme et tabous. *Bulletin de la Société de Linguistique de Paris* 117(1), 51–109.
- Martin, R. (1992 [1983]). *Pour une logique du sens*. PUF.
- Martin, R. (2014). *Comprendre la linguistique*. PUF.
- Martin, R. (2021). *Linguistique de l'universel*. Académie des Inscriptions et Belles Lettres.
- Martinet, A. (1965). Le mot. *Diogène* 51, 39–53.
- Martinet, A. (1967). Syntagme et syntème. *La linguistique* 2, 1–14.
- Martinet, A. (1968). Mot et syntème. *Lingua* 21, 294–302.
- Mathieu-Colas, M. (1996). Typologie de la composition nominale. *Cahiers de lexicologie* 69, 65–118.
- Mejri, S. (2002). Le figement lexical : nouvelles tendances. *Cahiers de lexicologie* 80(1), 213–225.
- Mejri, S. (éd.) (2009). *Le français moderne : La problématique du mot* 77(1).
- Mejri, S. (2009). Le mot : problème théorique. *Le Français Moderne* 77(1), 68–82.
- Mejri, S. (2018). La phraséologie française : synthèse, acquis théoriques et descriptifs. *Français moderne* 1(86), 5–32.
- Mejri, S. (2018). Les pragmatèmes et la troisième articulation du langage. *Verbum* 40(1), 7–19.
- Mejri, S. (2023). La néologie polylexicale. *Synergies Tunisie* 6.
- Mejri, S. & Ouerhani, B. (2008). Traduction en arabe de l'ouvrage de Gaston Gross *Les expressions figées en français* (1996). Publications du CERES.

- Mejri, S. & Zhu, L. (2020). Données dictionnaires informatisées. Réseaux inférentiels et phraséologiques. *Le français moderne* 78(1), 102–136.
- Mel'čuk, I. (2013). Tout ce que nous voulions dire sur les phrasèmes.... *Cahiers de lexicologie* 102(1), 129–149.
- Muller, C. (2013). Le prédicat, entre (méta)catégorie et fonction. *Cahiers de lexicologie* 102, 51–65.
- Neveu, F. (2004). *Dictionnaire des Sciences du langage*. Armand Colin.
- Pergnier, M. (1986), *Le mot*. PUF.
- Rey, A. & Chantreau, S. (1989), *Dictionnaire des locutions et expressions figurées*. Les usuels du Robert.
- Riegel, M. et al. (2009 [1994]). *Grammaire méthodique du français*. PUF.
- Saussure, F. (de) (1973). *Cours de linguistique générale*. Payot.
- Sułkowska, M. (2019). Quelques remarques sur la phraséologie appliquée. *Neophilologica* 31, 445–455.
- Whitehead, A. N. (2004). *Modes de pensée*. Vrin.
- Wilmet, M. (2003). *Grammaire critique du français* (3e éd.). Duculot.
- Zhu, L. (2020). Moule locutionnel lexicographique et traitement des phraséologismes. *Les Cahiers du dictionnaire* 11, 147–163.
- Zhu, L. (2016). Pour une notion de moule dans la phraséologie. *Cahiers du dictionnaire* 8, 97–109.
- Zhu, L. (2022). Discours dictionnaire, moule phraséologique et corpus textuel. *Langages* 225, 127–146.



**Claude Muller**

Université Bordeaux Montaigne  
France

 <https://orcid.org/0000-0002-7355-1765>

# Les pronoms démonstratifs neutres du français devant les subordonnées de type relatif

**French neutral demonstrative pronouns before relative-type subordinate clauses**

## **Abstract**

This work deals with the construction of relatives and cleft clauses that follow a demonstrative pronoun in French. It examines their mode of acquisition of meaning and reference, either by referential increase obtained by predicative interactions and lexical meaning, or by the process of focusing on a referent in the situation or context. We show that the original pronoun *ce* is no longer demonstrative in all these structures, and how the various demonstratives are used in actual French in those contexts.

## **Keywords**

Demonstrative, relative, cleft, focusing

## **1. Introduction**

J'ai rencontré Gaston Gross, il y a environ 47 ans, dans le laboratoire de Paris-7, le LADL, chez Maurice Gross, notre commun directeur de thèse. Ce fut une longue amitié, ponctuée de rendez-vous de travail, jurys divers à Paris comme à Bordeaux, chez lui ou chez moi, et dans divers colloques un peu partout. Gaston avait l'amitié franche et joviale, on pouvait compter sur lui. Je le croyais, à tort hélas, lui, ce solide lorrain, d'une longévité à toute épreuve. C'est un plaisir et un honneur pour moi

de lui rendre hommage en dédiant ce texte à sa mémoire. Il n'y est pas cité : nos travaux parcouraient des domaines différents, Gaston a laissé sa marque en lexicologie, et tout particulièrement par l'introduction de la sémantique dans l'étude des classes lexicales, alors que mes recherches m'ont plutôt orienté vers la syntaxe, et la sémantique de la syntaxe. Nous nous entendions fort bien, parce qu'au-delà de nos centres d'intérêt, notre vision de ce que doit être la recherche en linguistique était proche : Gaston a toujours cherché à répondre à une question théorique par un examen humble, minutieux et sans esbroufe des données linguistiques.

## 2. Les pronoms neutres démonstratifs : *ce* et les autres

Ce travail fait suite à une série d'articles<sup>1</sup> examinant la syntaxe et la sémantique des pronoms démonstratifs neutres du français : *ce*, *ça*, *ceci*, *cela*, dans leurs emplois devant des subordonnées. Les démonstratifs neutres sont quatre pronoms *ce*, *ceci*, *cela*, *ça*, formes dénués de sens lexical propre (sauf le trait « non humain »<sup>2</sup>), de genre ou de nombre (masculin singulier par défaut). Leur interprétation exige des informations soit de situation, soit de contexte. Ces informations sont de deux types : d'une part, de contenu sémantique nominal, d'autre part, relatives à leur détermination. Le contenu sémantique nominal de ces pronoms est celui d'un « non nommé » (Maillard, 1974), ou d'un objet perçu non répertorié par le lexique, « non classifié » (Kleiber, 1984), le trait non humain étant le seul élément de sens à peu près établi<sup>3</sup> ; ce contenu sémantique peut être explicité par le contexte ou la situation extralinguistique. La détermination, elle, fait problème, puisque le pronom neutre (nom vague plutôt que pronom) se présente sans aucun déterminant morphologique, et on notera aussi qu'il ne s'agit pas dans cette série de ce que Corblin (1990) a nommé des « déterminants sans nom » : morphèmes à nom amovible, comme *beaucoup* / *beaucoup de gens*, *certains* / *certains enfants*. Si déterminant il y a, ce devrait être *ce* qu'on trouve aussi dans les noms, et qui entre en composition avec *lui* / *elle* dans *celui*, *celle*. Comme il n'y a pas de noyau nominal morphologique dans nos pronoms neutres, ce qui

<sup>1</sup> Muller 2018, 2020, 2023.

<sup>2</sup> Le trait « humain » s'inscrit déjà dans une catégorie, les personnes ; en l'absence de toute classification, un « objet » perçu l'est comme « chose » avant d'être catégorisé, cf. Kleiber (1998a : 105), à propos de *ça*.

<sup>3</sup> Même s'il y a une exception de taille, la reprise d'un nom par *ce* sujet du verbe *être*, applicable aux humains : *Jean-Paul, c'était quelqu'un de bien*.

distingue par exemple *cela* de *celui-là*, il faut admettre que le *ce* dans les pronoms neutres est bien à distinguer du *ce* déterminant<sup>4</sup>.

On a naturellement tendance à interpréter ces pronoms comme des démonstratifs sémantiques. Les groupes nominaux démonstratifs, introduits par le déterminant *ce*, sont bien connus, et bien que faisant partie au sens large des définis, ils s'en distinguent par diverses propriétés, dont la principale est l'invite par un énonciateur à chercher dans le contexte ou la situation, à portée de perception, un référent adéquat, ce que Corblin (1987b : 242) décrit comme l'identification du référent « par proximité » ; suivant Kleiber (1984), les têtes de groupes nominaux démonstratif appartiennent aux « symboles indexicaux [...] (ayant) « un sens fonctionnel de désignation ». L'identification de leur référent ne se fait pas par le contenu sémantique de leur groupe nominal, qui intervient secondairement comme élément attributif aidant à la reconnaissance, mais selon un processus « instructionnel » dans le contexte énonciatif (Kleiber, 1998a : 105). Dans le cas des pronoms neutres, le contenu sémantique du groupe nominal étant à peu près vide, c'est donc surtout ce processus instructionnel qui permet la compréhension du sens en contexte. On admettra l'hypothèse de Kleiber (1998b : 221) à propos de *ça* : les démonstratifs neutres (exception faite de certains emplois de *ce* comme on le verra) sont *token-réflexifs*, associant au sens référentiel initial vide, comme signalé ci-dessus, mais enrichi par les interactions actanciennes, un sens instructionnel attaché à leur emploi. Celui-ci, selon Diessel (2014 : 122) consiste à attirer l'attention du locuteur sur un référent accessible en partant de la perspective fixée par le locuteur : référent de situation (exophorique), ou bien textuel, en amont (anaphorique) ou en aval (cataphorique) ; c'est cette opération énonciative (pilotée par un énonciateur, réel ou fictif selon le genre de l'énoncé), qu'on nommera pour faire court « monstration » dans la suite de ce travail. Il y a un trait quasiment universel des démonstratifs, qui est lié à ce sens instructionnel piloté par le locuteur, et lié à l'espace d'accessibilité du référent, c'est l'opposition, qu'on retrouve en français avec *cela* / *ceci*, porteurs de suffixes de localisation, entre un démonstratif proximal (*ceci*) et un démonstratif distal (*cela*) comme le montre entre autres Cornish 2017. En outre *cela*, plus fréquent que *ceci*, est la forme non marquée de cette opposition, et *ça* est la forme usuelle dans le registre familier, des formes suffixées *ceci* ou *cela* mais il a acquis des contextes propres<sup>5</sup>.

<sup>4</sup> L'opposition est soulignée dans Corblin (1987a : 88) qui oppose *Je prendrai celle-ci, de chemise* à *\*Je prendrai ceci de chemise* ; les pronoms neutres sont des noms « indistincts », des formes « saturées », inaptes à la reprise pronominale de noms coréférents.

<sup>5</sup> On n'emploie pas *cela* devant les impersonnels familiers *ça caille, ça pue* dans lesquels le démonstratif n'a pas de référent précis ; inversement, *cela* reste possible devant verbe *être* : *Cela est intéressant* mais *\*Ça est intéressant*.

Dans cet ensemble, le pronom *ce*, sans suffixe de localisation, fait figure d'élément hétérogène, morphologiquement basique, syntaxiquement contraint, mais historiquement premier<sup>6</sup> : *ce* n'est guère utilisable comme élément isolé, au contraire des trois autres formes, mais il bénéficie paradoxalement d'une distribution bien plus vaste<sup>7</sup>. De fait, comme on le verra sur les contextes de nominal précédant une subordonnée, il est souvent impossible de le remplacer par un autre démonstratif. Est-il démonstratif au sens qu'on vient de définir ? On se rappellera que *ce* a bien dans certains contextes une interprétation de démonstratif, par exemple dans les reprises de contenu propositionnel avec *et ce*, *sur ce*, *ce faisant*, mais il ne s'agit pas d'un introducteur de subordonnée ; par exemple avec l'expression figée *ce faisant* :

- (1) *Alors on le voit tendre la main, dénouer un ruban entortillé et, ce faisant, converser avec l'enfant d'un air très naturel.* (A. M. Garat, *Chambre noire*, 1990)<sup>8</sup>

Le *ce* en (1), proclitique sur le participe présent, équivaut au démonstratif à valeur anaphorique *cela* postposé ou à un groupe nominal démonstratif équivalent, *en faisant cette chose*. L'intérêt de (1) est aussi de montrer l'une des fonctions principales de ces pronoms : on peut les utiliser comme support d'un contenu propositionnel, par nature sans lexicalisation propre. Ici, c'est la phrase qui précède *ce*, référent par anaphore.

Dans les relatives, et dans une construction proche, les clivées, l'utilisation d'un de ces pronoms neutres comme antécédent du pronom relatif, donc comme support référentiel de la position vide dans la subordonnée, permet de donner, par relation cataphorique tenant au matériel lexical, un peu de substance sémantique au pronom neutre, de façon souvent marginale :

- (2) *Je vois ce que tu vois*

<sup>6</sup> Les quatre formes dans leur aspect actuel sont d'apparition presque contemporaine, dans le courant du 12<sup>e</sup> siècle, mais *ce* provient d'une réduction phonique d'une forme plus ancienne de type *ço*, alors que *ceci/ cela* sont suffixées sur *ce* et que *ça* semble être la réduction de *cela* peut être contaminée par le locatif ancien *çà* qui subsiste dans la locution *çà et là*. La forme d'origine est donc bien *ce*.

<sup>7</sup> Introducteur de complétive, cf. Muller (2023), d'interrogative indirecte, d'exclamative. On se limitera ici aux relatives et aux clivées.

<sup>8</sup> Les exemples ponctués et datés proviennent pour la plupart du corpus « contemporain » de la base Frantext (textes postérieurs à 1979).

permet ainsi de supposer que *ce* représente une chose vue aussi bien par le locuteur que son interlocuteur ; mais on n'en sait guère plus ; s'y ajoutent selon les contextes, des renvois au véritable contenu, par exemple :

- (3) *Je photographie tout ce que je n'avais pas pu la fois précédente et repars. Du moins, c'est ce que je crois. L'antivol de mon vélo refuse de s'ouvrir.* (P. Artières, *Vie et mort de Paul Geny*, 2013)

Dans cette scène, *ce que je crois* renvoie par anaphore à (*je*) *repars*. Le groupe nominal (antécédent plus relative) est suffisant pour permettre ce genre de mise en relation textuelle. Or, il apparaît que les formes diverses des pronoms ne sont pas toutes également utilisables ; que se passe-t-il si *ça*, *cela*, *ceci* est utilisé à la place de *ce* ? En l'occurrence, *c'est ça que je crois* n'est guère différent en acceptabilité, peut-être légèrement bizarre. Pour l'exemple (2) :

- (2') ? *Je vois cela que tu vois*

est plus étrange, sans être inacceptable ; mais pourquoi plus étrange ? Si on prend :

- (4) *Nous sommes tous des poissons morts, voilà ce que je crois, moi.* (E. Krecké, *Syncope, roman*, 2012)

Il est cette fois exclu de dire :

- (4') \**Voilà ça que je crois, moi*

C'est ce genre de bizarreries, ou de franches inacceptabilités, qu'on se propose d'examiner, dans relatives et clivées<sup>9</sup>, caractérisées également par un élément vide dans la position argumentale de la subordonnée. On peut déjà pressentir que les différences les plus nettes sont entre *ce* et les autres formes, parce que ce qui distingue *ça* de *cela*, c'est sans doute surtout son caractère plus familier, et parce que l'opposition entre *ceci* et *cela*, entre une localisation proche, une forme de grossissement, et au contraire une vision plus détachée ou plus globale, reste assez secondaire. On peut se douter que le phénomène de la « monstration », typique des démonstratifs, doit jouer un rôle, pas seulement pour les relations phoriques,

---

<sup>9</sup> On trouve notamment *ce que* dans des interrogatives indirectes et des exclamatives, constructions qu'on n'examinera pas ici ; le *ce* n'y est pas concurrencé par les autres formes, ce qui va dans le sens d'un statut catégoriel différent de celui des démonstratifs.

mais aussi du fait de la concurrence entre deux fonctionnements sémantiques sur lesquels on devra se pencher pour comprendre l'utilisation des formes, l'enrichissement sémantique par les contenus croisés du pronom et des propositions où il figure directement ou par lien structural, et la démarche instructionnelle propre aux démonstratifs. C'est donc ce qui différencie *ce* des autres démonstratifs qui sera l'enjeu essentiel. On examinera successivement les relatives, et les phrases à pronom neutre introduites par *c'est*, relatives ou clivées.

### 3. Les pronoms neutres dans les relatives

#### 3.1. Les emplois définis distingués des emplois démonstratifs

Dans Muller (2020), j'avais remarqué que le *ce* antécédent des relatives ne se comporte pas comme un démonstratif sémantique<sup>10</sup>. Le pronom n'est pas séparable par une pause de la relative, qui est obligatoirement restrictive; l'interprétation du contenu est strictement le sens compositionnel, donc référentiel, de la relative associée – avec cette différence, par rapport à la complétive<sup>11</sup>, que l'enrichissement cataphorique dû à la relative est moindre puisque la relation entre *ce* et le contenu de la subordonnée n'est pas appositive, mais est le produit de l'information obtenue par identité de référence de l'antécédent à la position argumentale dans la relative :

(5) *Donne-moi ce que tu tiens*

En (5), le sens construit pour *ce* est ainsi celui de *x*, dans *tu tiens x*, donc « chose tenue par toi ». La relative restrictive donne à *ce* un sens défini non démonstratif, *la chose que tu tiens*. Il n'y a donc pas ce qu'on trouve normalement dans le démonstratif, le processus de « monstration » pour enrichir cette interprétation. Dans ces contextes, le remplacement de *ce* par un démonstratif est assez bizarre :

<sup>10</sup> Cette hypothèse est par ailleurs confortée par celle, analogue, proposée par Hirschbühler et Labelle (1990) à propos de *celui* devant les relatives, dont il est un antécédent non séparable, qui se comporte selon eux comme un pronom défini.

<sup>11</sup> Dans la complétive (cf. Muller, 2023), *ce* ne figure qu'après préposition, mais il est le support cataphorique exact du contenu propositionnel de la subordonnée: *Je ne suis pas mort en fait: je mange, je bois; mais je suis mort en ce que mon œuvre est terminée...* (Beauvoir, *La cérémonie des adieux*, 1981).



(5') ??*Donne-moi ceci que tu tiens*

Ce qu'on peut comprendre par un effet de saturation concurrentielle : le démonstratif *ceci* est suffisant pour la communication, de même que la relative restrictive le serait. Il faut tout de même préciser que l'apport sémantique du démonstratif suppose une forme de proximité entre l'objet désigné et le locuteur, alors que la relative décrit cet objet comme « tenu » par l'interlocuteur. Ce sont des informations qui diffèrent, tant par leur contenu que par la manière dont elles sont communiquées. Ainsi, avec un autre énoncé, mais respectant le même état physique des interactions entre objet et protagonistes, *ceci* serait plus justifié et plus adéquat que *ce* :

(6) *Rends-moi ceci qui m'appartient*

En effet, *ceci* est le terme le plus approprié<sup>12</sup> pour désigner un objet particulier, localisé, à la fois possédé par l'interlocuteur et proche du locuteur. Surtout, seul le démonstratif sémantique est pleinement interprétable : *ce* en (6) ne bornerait pas dans l'espace le ou les objets visés, et prendrait une valeur trop générale. L'opposition entre *ce* et *ceci* tient ici au caractère démonstratif, « cette chose-ci », du pronom, alors que *ce* n'est qu'un nominal rendu défini par sa relation à la relative restrictive qu'il introduit, dépourvu du pouvoir de « monstration ».

Ces exemples montrent l'opposition entre *ce* et un démonstratif sémantique à contenu exophorique. Qu'en est-il lorsque la phrase est construite sur des relations textuelles, ou endophoriques ? Deux exemples, l'un réel, l'autre construit, vont nous permettre de le préciser :

(7) *Je me souviens, en riant maintenant, de ceci que je me répétais tout le temps : Mon Dieu, si cette femme était jolie, j'aurais vraiment tout pour moi.* (M. Havet, *Journal*, 2005)

(8) *Il m'a dit ceci que je n'oublierai jamais : sois un homme !*

Dans les deux cas, la relative est restrictive, mais comme généralement, peu informative ; le véritable contenu, visé par la « monstration » du pronom, est la phrase qui vient ensuite, en relation de cataphore. Précisons le point suivant : si on remplace *ceci* par *ce*, le résultat obtenu n'est pas nécessairement inacceptable<sup>13</sup> ;

<sup>12</sup> Le *ça* est évidemment possible en registre courant.

<sup>13</sup> Elle gagnerait en (7) à être bornée dans le temps : ...*de ce que je me répétais tout le temps pendant cette période...* Et la relation cataphorique est moins impérieuse qu'avec *ceci* : le complément d'information sur le sens de *ce* n'est pas requis.

avec (7), la relative *ce que je me répétais tout le temps* peut, même si *ce* n'y est pas vu comme démonstratif – *la chose que* plutôt que *cette chose* –, être cataphorique, cette relation étant accessible aux groupes nominaux définis dans certaines conditions, notamment suivis de relatives restrictives (Kęsik, 1989 : 120).

En (8), la relative contribue au sens référentiel : « chose que je n'oublierai jamais », mais le démonstratif (non seulement morphologique, mais sémantique) *ceci* impose en plus la construction d'un sens déictique, qui oriente ici sous forme cataphorique vers la phrase impérative qui suit, sans laquelle il faudrait supposer un référent non contextuel, par exemple la désignation d'un texte écrit. Dans l'un ou l'autre cas, *ce* est impossible cette fois :

(8') \**Il m'a dit ce que je n'oublierai jamais : sois un homme !*

Sans la phrase impérative, (8') serait également inacceptable ; et la présence de cette adjonction ne suffit pas à la rendre possible. Comment l'expliquer ? Et d'où vient la différence entre (7) et (8) ? On peut s'en tenir à l'opposition entre *ce* nominal défini et *ceci* nominal démonstratif pour le comprendre. Si on suppose que *ce* n'est plus démonstratif dans cette position en français actuel, il n'est pas non plus indéfini (si c'était le cas, la phrase (8') serait acceptable avec le sens *il m'a dit une chose que je n'oublierai jamais*). Il est donc défini, et sa définitude tient à la relative restrictive. Le contenu lexical de la phrase ne permet pas d'obtenir ce sens, avec l'unicité qui s'y attache :

(8'') \**Il m'a dit la chose que je n'oublierai jamais*

Cela ne tient pas à l'impossibilité d'avoir une fonction cataphorique pour *ce que*+relative, comme on l'a signalé, mais à l'absence d'unicité définie. Rappelons que le problème ne se pose pas avec un groupe nominal démonstratif, *cette chose que je n'oublierai jamais* : l'unicité tient alors à la monstration, c'est le pointage démonstratif qui la constitue.

Un autre exemple, cette fois avec une variation possible entre *ce* et *ceci* :

(9) *Le Père Noël m'a apporté ceci que j'espérais depuis toujours : un train électrique.*

(9') *Le Père Noël m'a apporté ce que j'espérais depuis toujours : un train électrique.*

Les exemples vus plus haut présentent un pronom intégré dans sa proposition, avec le sens défini « la chose » pour *ce*, et le sens démonstratif « cette chose »

pour les autres formes. Pourquoi (9') est-il possible alors que (8') est exclu ? On a supposé que *ceci* mobilise deux méthodes d'appropriation du sens, alors que *ce*, défini non démonstratif, est limité à son apport lexical, le contenu de la relative. On notera que (9) n'est interprétable, comme attendu, que si la continuation avec le nom est présente, ou si la situation donne la perception du sens de *ceci* ; pour (9'), la première partie de l'énoncé suffit, même si le sens exact de *ce* reste non élucidable. La phrase est acceptable, avec ou sans sa deuxième partie. En effet, la définitude exige une condition, qui est ici réalisée, celle de l'unicité du référent, et le contexte de (9) permet cette interprétation :

(9'') *Le Père Noël m'a apporté la chose que j'espérais depuis toujours*

La différence entre *ce*, nominal défini malgré sa morphologie, et *ceci*, démonstratif est donc d'une part que *ce* est interprétable sans obligation d'une indication de contenu par anaphore ou cataphore, mais avec toute l'extension permise par le contenu lexical de la relative, si bien que la définitude doit s'associer à l'unicité ensembliste, alors que *ceci* exige une interprétation possible, contextuelle ou de situation, mais sans contrainte de l'unicité éventuelle qu'imposerait la relative, puisque l'unicité résulte de la monstration.

On peut encore montrer cette différence en examinant (10) :

(10) *Il m'a légué tout ce qu'il possédait*

Si *ce* était démonstratif, ou remplacé par *ceci*, le groupe nominal signifierait *toutes ces choses* avec une limite spatiale à l'interprétation, celle des objets visibles, montrés par le locuteur ; mais dans (10), le sens est bien « toutes les choses qu'il possédait », dans un sens juridique, au-delà du perceptible.

On peut aussi tirer parti d'une opposition remarquée depuis longtemps entre groupes nominaux définis et démonstratifs pour montrer que *ce qu-* n'est pas démonstratif : les groupes nominaux définis précédés de *tous* et suivis d'un syntagme verbal négatif voient le quantifieur dans le sujet affecté par la négation alors que ce n'est pas le cas avec un groupe nominal démonstratif (avec un contenu lexical analogue, cette propriété étant susceptible de varier selon ce contenu) :

(11) *Tous les enfants n'aiment pas les épinards*

(11') *Tous ces enfants n'aiment pas les épinards*

On interprétera (11') de préférence comme une négation totale sans effet particulier sur le sujet (*aucun de ces enfants n'aime les épinards*) et (11) comme une

négation portant sur *tous* (*pas tous les enfants*). Avec la relative *tout ce que*, l'interprétation est celle des définis, comme en témoigne l'interprétation univoque du proverbe :

(12) *Tout ce qui brille n'est pas or = Ce n'est pas tout ce qui brille qui est de l'or*

A comparer à (12') qui a le sens attendu avec les démonstratifs

(12') *Tout ceci qui brille n'est pas de l'or*

Pour les constructions à pronom relatif visible, du fait de la présence de prépositions, on utilise aussi *ce* comme support nominal permettant la complémentation de celles-ci, un peu comme dans les complétives, et obligatoirement<sup>14</sup> en cas de prépositions différentes. Le *ce* qui tire sa signification du croisement des déterminations des prédications principale et subordonnée est également un défini (ci-dessous, il équivaut à *la chose*) sans la nécessité ou la possibilité d'une démarche de monstration, et sans remplacement par un autre pronom neutre :

(13) *...personne à Saïda n'avait la moindre idée de ce à quoi pouvait bien ressembler la vraie griffe de Jean-Paul Gaultier...* (C. Bentz, *Oublier les fleurs sauvages*, 2021)

Il en irait différemment si la phrase était segmentée :

(13') *Personne (...) n'avait la moindre idée de ça, à quoi pouvait bien ressembler le griffe de J.P. Gaultier*

Cette possible réinterprétation comme démonstratif du terme qui articule la subordonnée quand il y a rupture et apposition détachée de celle-ci pourrait illustrer le processus d'évolution probable qu'a connu le démonstratif *ce* vers un pronom défini. Il est lié à sa cliticisation, cette modification syntaxique s'associant à un changement sémantique. On peut comprendre ainsi ce changement : la cliticisation de *ce* antécédent, jointe à l'impossibilité de rendre l'indéfini saillant par accentuation, impose comme seule voie à l'interprétation sémantique de *ce* l'association lexicale du contenu de *ce* avec une relative restrictive. Il n'y a plus dès lors de voie « instructionnelle », liée à la monstration, qui puisse sub-

<sup>14</sup> On a le choix avec la même préposition : *Il s'est assis sur quoi tu t'es assis* ou *il s'est assis sur ce sur quoi tu t'es assis* (la seconde version est plus fréquente).

sister. Cela n'exclut pas que l'ensemble *ce qu-* + relative entretienne des relations anaphoriques ou cataphoriques avec le contexte : ces relations sont alors du type de celles possibles pour les groupes nominaux définis à antécédent et relative, et pas celles limitées au seul pronom. Cette réduction du sens de *ce* est rendue possible par l'existence des autres pronoms démonstratifs, qui gardent leur suffixe localisant, *-ci* et *-là*, ou pour *ça* une autonomie forte de mot isolé, non cliticisable.

Pour la vaste gamme d'emplois de *ce* dans les relatives, on peut se reporter à Muller (2020). Outre les contenus propositionnels, le sens de *ce* se colore de diverses interprétations selon les contextes, quantité, partie, fragment d'objet ou nuance de propriété, collection d'objets ou d'individus, objets approximatifs et sans nom répertorié, comme en (14) :

- (14) *Je me demandais où elle dénichait des étrangetés pareilles, [...] des filles montées sur des échasses pas croyables qui riaient pour n'importe quoi en secouant ce qui leur tenait lieu de chevelure.* (A. Gavalda, *je voudrais que quelqu'un m'attende quelque part*, 1999)

### 3.2. Les pronoms démonstratifs autres que *ce* dans les relatives de type déterminatif

Il y a une particularité, observée depuis longtemps, souvent signalée (Gary-Prieur, 1998 : 45) : les relatives à antécédent démonstratif ne sont pas corrélées à celui-ci de la même façon que dans les constructions à antécédent défini, sans être pourtant facultatives ou appositives. La relative peut y être indispensable pour la compréhension exacte du groupe nominal, comme le sont les relatives déterminatives des définis, mais s'y ajoute le démonstratif et sa propre voie de repérage du référent, la démarche de « monstration », par renvoi soit à la situation, soit au contexte ; s'y ajoutent, aussi, pour les groupes nominaux à déterminant démonstratif (mais peut-être pas pour les pronoms, comme on le verra plus loin), les emplois à déterminant démonstratif et relative épinglant un objet mémoriel supposé commun aux interlocuteurs (Kleiber, 2004a, 2004b) :

- (15) *Tu te souviens de ce prof qui ne donnait que des bonnes notes ?* (Kleiber, 2004b : 232)

Nous reviendrons plus loin sur la possibilité pour un pronom démonstratif de figurer en antécédent dans un énoncé du type de (15).

Le cas le plus courant d'utilisation d'un pronom démonstratif est la combinaison de la double démarche signalée plus haut, l'accroissement du sens référentiel par la relative, et l'établissement, piloté par le locuteur, d'une relation soit à la situation, soit au contexte antérieur ou postérieur, distincte de la relative :

- (16) *Tout ceci que je te raconte, je ne l'ai raconté à personne.* (A. Jenni, *L'art français de la guerre*, 2011)
- (17) *Un homme tout vêtu de blanc [...] regarda son croquis [...]. « Vous vous interrompez trop souvent, jeune homme. Je comprends que vous vouliez vérifier si cela que vous avez tracé est vrai, mais pour que votre peinture vive [...] il faut que vous n'interrompiez pas votre souffle. Vous devez vous laisser guider par l'unique trait du pinceau ».* (A. Jenni, *ibid.*)

En (16), *ceci* renvoie aux pages qui précèdent ; la relative permet de mieux cerner le sens du groupe nominal, *ce que je te raconte*, à quoi s'ajoute avec *ceci* l'immédiateté de la présence, juste auparavant, de ces choses racontées. En (17), *cela* a une fonction identique : à la fois tête de la relative *ce que vous avez tracé*, et forme longue de pronom qui renvoie cette fois à la situation concrète décrite dans le récit, au dessin tracé. Dans les deux cas, le *ce* aurait suffi, mais le démonstratif sémantique permet de dissocier l'antécédence de la relative, la partie lexicale du sens, de l'indication déictique qui trouve son aboutissement dans le contexte.

Il arrive que la forme suffixée du démonstratif soit la seule possible, sans possibilité d'emploi de *ce*, et c'est toujours pour les raisons signalées en 3.1. *supra* : l'interprétation définie est incompatible avec le contexte :

- (18) *Mais je suis venu à l'émission de Pierre Sipriot parler de mon livre [...]. Aussi bien ai-je dû me convaincre que c'est moi qui ai écrit cela que je reconnais, où je me reconnais si peu.* (C. Mauriac, *Signes, rencontres et rendez-vous*, 1983)

Le démonstratif renvoie aux pages écrites par le romancier Claude Mauriac, qui est aussi le locuteur : la phrase *c'est moi qui ai écrit ce que je reconnais* n'aurait aucun sens dans ce contexte ; l'interprétation par monstration (figurée ici) est indispensable à la compréhension : *ces pages (de mon livre) que je reconnais*.

On a relevé cependant, en explorant les corpus d'exemples réels, de nombreux cas dans lesquels le démonstratif est employé sans utilité communicative, et pour lesquels on peut s'interroger :

- (19) *Olivier s'était senti un peu jaloux de ne pas partager cela que Samuel portait en lui et, par réciprocité, il avait évoqué la vie de son village aux confins de l'Auvergne.* (R. Sabatier, *Les fillettes chantantes*, 1980)
- (20) *L'effet qu'elle [la photographie] produit sur moi n'est pas de restituer ce qui est aboli (par le temps, la distance), mais d'attester que cela que je vois, a bien été.* (R. Barthes, *La chambre claire*, 1980)

L'emploi de *ce* en (19) et (20) aurait été parfaitement normal, en lieu et place de *cela*. La relative, *cela que Samuel portait en lui*, figure dans le discours intérieur d'Olivier, rapporté en (19), et *cela* évoque à la fois, pour Olivier, l'altérité malgré sa proximité du monde intérieur de Samuel : effet de style donc ; il en va de même en (20) : *ce que je vois* est neutre ; *cela que je vois* y ajoute un effet de proximité, de réalité saillante, pour l'auteur de l'essai. Mais contrairement aux exemples (16) à (18), on ne trouve pas de justification précise dans le texte à une démarche « instructionnelle » distincte : les deux procédés de construction du sens convergent vers la relative. On a trouvé de nombreux exemples de ce type dans les textes littéraires récents, peut-être un effet de mode. D'autre part, cette innovation est peut-être d'ordre historique : les relatives en *ce* pourraient suivre la voie évolutive des complétives en *ce* (Muller, 2023), qui consiste à renforcer un démonstratif démonétisé, devenu simplement défini, par une forme suffixée qui reste, elle, sémantiquement démonstrative. Bien évidemment, les deux explications sont compatibles l'une avec l'autre.

Des raisons structurelles peuvent nécessiter l'emploi des formes suffixées. Ainsi, la complémentation du pronom neutre par un adjectif, entre antécédent et relative, interdit ci-dessous l'emploi de *ce* : \*...*ce de très urgent que le malade attendait* :

- (21) *Il [le pharmacien de garde de nuit] lisait les ordonnances avec soin, il vérifiait plusieurs fois, [...], il jugeait l'allure de son client ; puis il partait dans l'arrière-boutique aux étagères et rapportait ceci de très urgent que le malade attendait en se balançant d'une jambe sur l'autre...* (A. Jenni, *L'art français de la guerre*, 2011)

Cependant ici aussi, il y a mise en saillance avec le démonstratif, le terme proximal, le plus marqué, une manière ici d'adopter le regard du malade et de signifier son impatience. Aucun démonstratif ne serait stylistiquement neutre dans ce contexte, et on pourrait hésiter entre « la chose très urgente » et un indéfini, « quelque chose de très urgent ». En (21) se lit la recherche par l'auteur d'une mise en scène énonciative de ses personnages.

De telles constructions ne sont pas généralisables avec toutes les formes : *ça*, peut-être parce que relevant d'un registre plus familier, semble échapper à ce réemploi de démonstratifs focalisables devant les relatives, du moins dans celles qui ne suivent pas le verbe *être* (on les examinera plus loin)<sup>15</sup>. Les exemples construits sont peu acceptables, et, ce qui intrigue, nettement moins que les groupes nominaux démonstratifs correspondants :

(22) \**Tu as appris ça qui est arrivé à Paul ?*

(22') *Tu as appris cette histoire qui est arrivée à Paul ?*

Dans ce cas c'est *ce* qui est le pronom naturellement usité :

(22'') *Tu as appris ce qui est arrivé à Paul ?*

Mais c'est avec un sens défini, et non démonstratif.

On peut reconnaître dans *cette histoire* en (22') ce que Kleiber (2004a) nomme un groupe nominal « démonstratif cataphorique spécifique », qui figure normalement devant une relative restrictive, et Kleiber décrit cette construction comme l'appariement à l'antécédent démonstratif d'une information (celle détaillée dans la relative) présentée comme probablement déjà connue de l'interlocuteur, simplement « rafraîchie », réactivée, ce qui autorise le démonstratif (cf. *supra*, exemple (15)). Il me semble qu'il existe une condition structurale à cet appariement : l'existence d'une base nominale permettant l'enrichissement référentiel du groupe nominal démonstratif. Le « *ça* », dénué de tout contenu sémantique nominal hors de sa signification vague de « chose », et surtout utilisé le plus souvent en référence à du réel immédiat, pourrait sinon bloquer, en tout cas rendre difficilement acceptable, un tel processus d'accrétion du sens lexical par adjonction d'une relative, tel qu'il apparaît dans les groupes nominaux à déterminant démonstratif. L'explication de la difficulté de telles constructions pourrait tenir, d'une part à l'absence de contenu lexical propre dans le démonstratif, rendant difficile le lien avec le contenu lexical de la relative, d'autre part à l'appétence particulière de *ça* pour les référents de situation, rendant sa saisie communicative largement indépendante d'une relative. C'est dans ces contextes de renvoi à de l'extralinguistique qu'on peut trouver des exemples nettement plus acceptables de cette construction, qui n'a plus rien d'un rappel mémoriel :

<sup>15</sup> Sur 200 occurrences de *ça que* examinées dans les corpus « contemporain » de Frantext, 185 relèvent de clivées.



(23) *Pierre a rapporté ça qui est tombé de ta poche*

(24) *Je vous rends ça qui vous revient de droit*

L'utilisation « mémorielle » du démonstratif en (25) ci-dessous, sur le modèle de (15) (Kleiber, 2004b), semble impossible avec *ça*, alors qu'elle le pose aucun problème avec *ce* (sans surprise : ce renvoi mémoriel est également possible avec un nominal défini suivi d'une relative) ; l'utilisation de *ça* est exclue :

(25) *Tu te souviens de cet incident qui nous a tellement effrayés, quand on a fait l'ascension du pic du Glacier blanc ?*

(25') *Tu te souviens de ce (\*ça) qui nous a tellement effrayés, quand on a fait l'ascension du pic du Glacier blanc ?*

Pourquoi le pronom *ça* ne fonctionne pas du tout dans ce contexte ? Il se pourrait que l'immédiateté de la référence de *ça* ne permette pas le grand écart temporel entre le présent et le souvenir passé évoqué dans la relative ; et le lien entre *ça* et la relative est différent, en termes de cohésion, de celui qui existe dans un groupe nominal à déterminant démonstratif.

### 3.3. Le cas des relatives apposées reprises de contenu propositionnel

Les reprises de contenu propositionnel utilisent naturellement les pronoms neutres de façon libre :

(26) *Prends le raccourci, ça/cela nous fera gagner du temps ; ...ce sera préférable*

(27) *...je ravitaillais les bonnes personnes. Et ce, depuis assez longtemps pour que mon appartenance au bon côté ne puisse être mise en doute. (A. Jenni, L'art français de la guerre, 2011)*

En (26), *ce* est exclu devant *fera* parce que ce pronom ne peut être sujet d'un verbe autre que *être* ; mais en (27), il est démonstratif sémantiquement, et pourrait alterner avec *ceci* ou *cela*, *ça*.

Pendant, il existe un type particulier de relative à reprise de contenu propositionnel, dans laquelle le démonstratif suivi de la relative restrictive forme une prédication apposée complète, sans autre lien de subordination. Dans ce cas, *ce* est le seul pronom neutre possible :

- (28) *Il travaille de mieux en mieux, ce (\*ceci/cela/ça) qui étonne tous ses professeurs*

Il est impossible en (28) de remplacer *ce* par un autre pronom neutre, ce qu'on peut expliquer (Muller, 2018) en supposant que *ce*, contrairement aux autres pronoms neutres, non seulement n'est pas démonstratif dans ce contexte, mais équivaut à un nom sans détermination, à un nominal nu, avec le sens approximatif de *chose*. Cette construction, très courante, permet d'associer à un contenu propositionnel généralement antérieur celui de la relative restrictive qui est ici une prédication nouvelle, cette relative ayant dans ses arguments un élément coréférent. La relation de coréférence entre *ce* et le contenu propositionnel antérieur est possible du fait de la vacuité sémantique du démonstratif, et n'a pas besoin du support d'un déterminant quel qu'il soit. Il se trouve que *ce* est la seule forme morphologiquement démonstrative qui ait cette faculté de jouer le rôle d'un nom nu.

La même explication peut être proposée pour une construction morphologiquement proche, celle des relatives indépendantes, utilisées comme titres de livre ou de film, ou de morceau de musique :

- (29) *Ce que je crois* (livre, Jacqueline de Romilly)  
 (30) *Ce dont rêvent les filles* (titre français du film de Dennie Gordon, *Reluctant debutante*, 2003)  
 (31) *Ce que raconte le vent d'Ouest* (Debussy, prélude)

Le *ce* renvoie à un contenu dont le lecteur, spectateur ou auditeur pourra prendre connaissance ; la seule indication qui permette de s'en faire une idée est la relative restrictive qui suit le pronom, et ici aussi, le référent n'est pas démonstratif, c'est le contenu seul qui importe : *choses que je crois* par exemple, et c'est la fonction de support référentiel du nom nu qui est mise en œuvre. Il ne serait pas impossible ici d'avoir, avec le nom *chose*, l'article défini, mais le mode énonciatif, le titre, neutralise l'opposition entre nom nu et nom défini. Il est aisé de vérifier que dans ces emplois, les autres pronoms neutres sont tout à fait inacceptables.

#### 4. Les constructions *c'est ce, ça, cela, ceci qu-*

Il faut examiner à part ces constructions, qui mêlent clivées et relatives de type présentatif ou explicatif. On y observe, pour les démonstratifs pleins, *cela, ceci,*

ça beaucoup plus d'occurrences que dans les relatives ; et *ce* y est également très présent.

#### 4.1. Distinguer relatives et clivées

Auparavant, nous devons distinguer les véritables relatives et les clivées. Les clivées utilisent le verbe *être* après démonstratif (au présent, *c'est*) introduisant un terme focalisé, suivi d'une structure subordonnée défective dont le terme vide, thématique, coréférent au focus ; l'articulation subordonnante varie selon les cas, une conjonction<sup>16</sup> lorsque le focus est un équivalent d'adverbe :

(32) *C'est en forgeant qu'on devient forgeron*

ou le type relatif, notamment avec un terme vide correspondant au sujet :

(33) *C'est ce livre qui raconte l'histoire*

La subordonnée peut présenter des variations de construction<sup>17</sup> lorsqu'il y a une préposition, soit associée au focus, soit restée dans la subordonnée :

(34a) *C'était d'un cancer de la prostate que souffrait Mitterrand*

(34b) *C'était un cancer de la prostate dont souffrait Mitterrand*

ou encore figurant dans les deux composantes :

(34c) *C'était d'un cancer de la prostate dont souffrait Mitterrand*

Enfin, la clivée peut être vue comme un cas d'extraposition, le sujet *c'* étant coréférent au contenu thématique de la subordonnée, ce qui autorise parfois une équivalence étroite avec la pseudo-clivée :

---

<sup>16</sup> Si on admet, comme je le fais, les explications de type génératif pour cet ensemble, la différence est estompée en construction non propositionnelle par des conjonctions à fonction de support pronominal relatif, notamment *qui* sujet et *que* non-sujet des clivées ; dans les deux types, la position argumentale du correspondant d'un pronom relatif est vide et co-indiciée avec un antécédent.

<sup>17</sup> Toutes ne sont pas également utilisables ; la forme la plus courante place la préposition dans la principale ; toutes restent bien attestées en français actuel.

(35) *Ce dont souffrait Mitterrand, c'était (d')un cancer de la prostate*

Il y a eu de nombreuses discussions sur les clivées, notamment sur leur rôle fonctionnel, leur syntaxe, leur histoire et leur structure sémantique<sup>18</sup>. Il est certain que d'autres constructions existent, avec la même forme *c'est X que P*, qui ne sont pas des clivées, mais des relatives. Sémantiquement, la clivée est une construction phrastique complexe visant à focaliser un argument quelconque d'une structure prédicative simple, en l'isolant comme complément de *c'est*, et en faisant suivre ce segment de la subordonnée constituée de ce qui reste de la structure prédicative. La jonction sous forme de relative à pronom exprimé ne se réalise que lorsque cet argument s'y prête. Surtout, la relative, si elle existe, ne peut servir de justification à l'emploi du défini devant l'antécédent, puisque celui-ci, tel quel, est par ailleurs argument de la phrase simple correspondante. Par exemple :

(36) *C'est la personne qui est venue tout à l'heure*

est très probablement<sup>19</sup> une relative « présentative », parce que la base prédicative de la clivée devrait être *La personne est venue tout à l'heure*, avec le défini devant « personne », non justifié par la relative, ce qui n'est pas plausible avec ce nom. Si, comme dans de nombreuses clivées, il y avait un présupposé sur « une personne est venue tout à l'heure », le groupe nominal sujet de la clivée serait démonstratif :

(36') *C'est cette personne qui est venue tout à l'heure*  
= *C'est cette personne, la personne qui est venue tout à l'heure*

Il est donc beaucoup plus facile de trouver des clivées à focus défini avec des groupes nominaux comme les noms propres ou les démonstratifs, qui n'ont pas besoin d'une expansion relative pour leur assurer la définitude.

A côté des relatives « présentatives », structure binaire dans laquelle le sujet dans *c'est* est anaphorique ou exophorique, il existe des constructions focalisées, structures ternaires n'ayant pas toutes les propriétés des clivées, comme ce jeu de scène, après un brouhaha de voix, dans *Cyrano* :

(37) *Jodelet, qui a regardé :*  
– *C'est Montfleury qu'on hue !* (Rostand, *Cyrano de Bergerac*, I, 4)

<sup>18</sup> Cf. par exemple Muller (2003).

<sup>19</sup> L'acceptabilité pour une clivée serait bonne avec *C'est le chien qui fait du bruit*, parce que le chien sera défini par familiarité, ce qui ne convient pas avec un nom à sens général comme *la personne*.

La phrase ne veut certainement pas dire *celui qu'on hue, c'est Montfleury*; c'est une variante ternaire (à prédication seconde) de « présentative », dans laquelle l'énoncé réfère directement au bruit entendu en coulisse<sup>20</sup>. Cette construction focalise le nom attribut de *c'est*, et contrairement à la présentative « binaire »<sup>21</sup>, permet les démonstratifs :

(38) *C'est ça qu'on entend ! (Montfleury qui se fait huer)*

mais ne permet pas la diversité de construction caractéristique des clivées :

(37') [le bruit] *C'est Montfleury dont on se moque ! / \*C'est de Montfleury qu'on se moque !*

#### 4.2. Distinguer *c'est ce qu-* et *c'est ça qu-*

Le *ce* exige une relative déterminative et, comme on l'a vu, fonctionne comme un défini non démonstratif; on distinguera donc (39) de (40) comme respectivement étant une clivée et une relative :

(39) *C'est ça qui est arrivé à Paul*

(40) *C'est ce qui est arrivé à Paul*

Il est en effet impossible de focaliser *ce*, ou de disloquer la phrase :

(40') *\*C'est ce, qui est arrivé à Paul*

Un des critères de reconnaissance des clivées est lié à la possibilité de variation de la position ou de la redondance de la préposition associée au focus, comme on l'a signalé *supra*. Ainsi, (41) est une clivée avec redondance de la préposition *de*, présente aussi dans *dont* :

(41) *C'est de ça dont vous voulez me parler ?* (Télévision : Arte, 19.05.2022, *En thérapie*, Claire, 22h45)

---

<sup>20</sup> Ces trois constructions avec *c'est*, présentative, clivée, construction à prédication seconde, sont bien distinguées dans Riegel *et al.* (1994 : 820). Il reste probablement d'autres constructions à singulariser dans cet ensemble.

<sup>21</sup> Contrairement à *voilà*, qui est typiquement entrée de présentative (cf. 4 *supra*).

L'utilisation de *ce* au lieu de *ça* n'autorise, comme prévu, que la syntaxe des relatives<sup>22</sup> :

- (42a) *C'est ce dont vous voulez me parler ?*  
 (42b) *??C'est de ce dont vous voulez me parler ?*  
 (42c) *\*C'est de ce que vous voulez me parler ?*

Il s'agit bien ici d'une relative déterminative, donc d'une construction de type présentatif. Le type redondant, (42b), bien que parfois attesté, comme le montre (43), est alors erroné :

- (43) *...le film [Kaamelott] a surpris ses fans, avec notamment une scène : celle du baiser d'Arthur et Guenièvre, à l'initiative du roi de Logres. Était-ce un baiser d'amour ? Ou un signe équivoque du fait qu'Arthur accepte enfin son destin ? C'est de ce dont on va parler aujourd'hui (hitek.fr>cinéma, 20.08.2021)*

Il faudrait<sup>23</sup> supprimer le *de* devant *ce* (relative) ou utiliser *ça* en place de *ce* (clivée).

On ne trouvera donc pas de clivées avec *ce* constituant seul le focus. Avec *c'est ce+Relative*, la structure est celle d'une relative attachée à un défini : *la chose que/dont...* et le sens n'est défini que par cette relation étroite avec la relative. Ceci n'exclut pas une interprétation indexicale, par la situation, ou anaphorique / cataphorique, cette relation étant accessible aux groupes nominaux définis suivis de relatives, comme on l'a déjà dit.

<sup>22</sup> L'inacceptabilité de (42b) est moins nette parce que *de* peut se justifier par un contexte extérieur à l'énoncé, « à propos de » par exemple.

<sup>23</sup> On peut trouver des constructions *c'est de ce dont* si le *de* est justifié par ailleurs, comme dans: *Ce dont une chose peut avoir besoin, c'est de ce dont elle provient puisque cette chose est supérieure à elle et meilleure qu'elle.* (S. Roux, *Transcendance et relation, Plotin et l'antinomie du principe*, 2012, cairn-info). Ici, on a affaire à une pseudo-clivée, ce qui explique le *de* devant *ce*, repris de la complémentation d'*avoir besoin*, le *dont* étant le pronom introduisant la relative déterminative.

### 4.3. Les emplois de *c'est ce qu-*

Le cas le plus fréquemment observé<sup>24</sup> est un cas particulier de « présentatives », celui de constructions « explicatives »<sup>25</sup>, dans lesquelles *c'est ce qu-* est le développement d'une séquence antérieure par anaphore, par exemple :

- (44) *Je vais avoir six ans. Je suis « un homme ». C'est ce que me dit mon père sur le chemin de la « grande école ».* (M. Genevoix, *Trente mille jours*, 1980)
- (45) *N'est pas putain ou tueur qui veut. C'est ce que j'apprendrai avec la pénurie.* (F. D'Eaubonne, *L'indicateur du réseau*, 1980)

La relative est une extension cataphorique de *ce*, mais pour un contenu sémantique marginal (chose, dite, expliquée, apprise) : l'essentiel du contenu, par l'intermédiaire du *ce* sujet, est explicité anaphoriquement. L'explicitation cataphorique est nettement moins fréquente :

- (46) *Je pensais, à sa suite [Max Jacob] – et c'est ce que je voulais illustrer dans le texte qu'il déchira – , qu'on pouvait, presque à l'infini, faire naître de nouvelles images en désorganisant la phrase...* (E. Jabès, *Du désert au livre*, 1980)
- (47) *Pourquoi quelqu'un qui ne fume pas s'amuserait-il à trimballer son cendrier d'un appartement à l'autre ? – Ouais, fait l'Asperge, c'est ce que je dis toujours aux clients : quand on a trop de trucs, on finit par ne plus savoir ce qu'on possède.* (J. L. Benoziglio, *Cabinet portrait*, 1980)

### 4.4. Les pronoms neutres *ceci, cela*, dans les constructions en *c'est*

Sur corpus, les constructions *cela que* avec le type relatif comportent une majorité de clivées<sup>26</sup>. Avec *c'est cela qui / que*, la question de pose de savoir si certaines des constructions observées sont des clivées ; on verra que la question est surtout liée à l'interprétation sémantique de ce qu'on appelle une clivée.

---

<sup>24</sup> Frantext, corpus « contemporain » (textes postérieurs à 1979, 300 occurrences *ce que* examinées, parmi lesquelles 13 *c'est ce que* : 11 constructions explicatives anaphoriques, pour 2 cataphoriques.

<sup>25</sup> Proches de certaines constructions appelées « identificatives » par Apothéloz (2018 : 33).

<sup>26</sup> Dans le corpus « contemporain » de Frantext, 758 résultats *cela que*. Sur les 200 premières occurrences, on compte 11 relatives et 151 clivées.

De nombreuses clivées comportent un focus avec préposition, *c'est pour / comme / après / en / cela que*. Dans toutes ces constructions, *cela* renvoie à du contenu langagier qui précède. Ainsi, en (48), le démonstratif est lié à la subordonnée en tant qu'élément d'un circonstanciel de cause non explicité, cette explicitation étant antérieure :

- (48) *On ne se connaissait pas de tels ennemis. Mais eux nous connaissent. C'est souvent dans ce sens que fonctionne la destruction, à cause de cela que les pauvres sont condamnés à avoir les moins vides...* (G. Bienne, *Le silence de la ferme*, 1986)

La relation est parfois cataphorique :

- (49) *...la diaspora moderne des œuvres, et c'est à cela que je voulais en venir, n'est selon moi que la conséquence de cette insuffisance foncière.* (Y. Bonnefoy, *Un rêve fait à Mantoue*, 1980)

Lorsque le démonstratif est seul en position focalisée, la signification contrastive des clivées est parfois évidente :

- (50) *Dernier carton, j'ouvre. Des lettres de mon oncle, de ma sœur, d'amis, de collègues, envoyées à mes diverses adresses (...). Ce n'est pas cela que je cherche. Je fouille dessous...* (S. Doubrovsky, *Un homme de passage*, 2011)

Le *cela* utilisé se distingue nettement de *ce*, également possible ici, en ce qu'il manifeste l'intention déictique, orientée généralement vers le contexte antérieur, et donc dissociée de la subordonnée : *Ce n'est pas cela, ce que je cherche*. Comme *cela* réfère à un contenu antérieur, et malgré l'absence de toute pause entre le pronom et la subordonnée, la détermination de cet unique terme est double : à la relative déterminative qui suit, comme au contenu anaphorique.

L'addition des deux déterminations n'implique peut-être pas, dans certains cas, une équivalence sémantique avec une pseudo-clivée, qui serait en (51) ci-dessous *ce que je ne comprends pas, c'est cela*, avec thématization de la subordonnée. L'information dans la subordonnée semble aussi nouvelle, comme elle le serait si l'auteur avait écrit *c'est quelque chose que je ne comprends pas*, et la clivée semble se ramener à une façon de focaliser *cela* : *et cela, cette*



*chose, je ne la comprends pas, sans la hiérarchie informative des clivées les plus caractéristiques*<sup>27</sup> :

- (51) *Le mystère qui différencie les hommes et fait de l'un d'entre eux un poète, c'est cela que je ne comprends pas, dit-elle.* (M. Havet, *Journal*, 2005)

Certaines clivées pourraient ainsi résulter de la seule mise en relief par focalisation.

Même chose dans (52) : la relative ne semble pas thématique, c'est plutôt le résultat d'un bilan dont le détail figure dans l'énumération reprise anaphoriquement par *cela* :

- (52) *... cet effort vers un confort de masses, rude et moderne (possibilité d'envoyer des télégrammes, casse-croûte au Foyer du Soldat, tasse de café gratuite à la cantine), c'est tout cela qui donne à l'ensemble un caractère de fête fasciste.* (J. P. Sartre, *Carnets de la drôle de guerre*, 1983)

Il peut arriver, mais ça n'est pas le cas le plus fréquent, que la référence essentielle pour le pronom soit cataphorique, et située au-delà de la subordonnée, comme en (53) :

- (53) *On voyait les uns boire, poser leur verre, avancer leur main pour attraper un gâteau, prendre une bouteille. C'est cela que Vérida aimait : ce moment où les invités s'excitaient les uns les autres...* (J. Duvignaud, *L'ère de la République*, 1957)

L'emploi de *ceci* est beaucoup moins répandu dans ces constructions<sup>28</sup> avec *être*, et surtout dans des clivées avec préposition ; la « monstration » se répartit entre référence déictique cataphorique, sur le modèle des complétives, et référence à la situation. L'interprétation ne diffère guère des constructions avec *cela* :

- (54) *Elle me dérangeait, certes, mais elle était éblouissante. [...] A l'aise dans ses vêtements d'une qualité discrète, mais ancestrale. C'est ceci qui était*

---

<sup>27</sup> Fait déjà observé par Sandfeld (1977 : 123) à propos de *C'est Sidonie qui va être étonnée*. « Cette phrase nous apprend deux choses : 1. Il y a quelqu'un qui va être étonné et 2. Ce quelqu'un, c'est Sidonie ». Il en donne un exemple avec le démonstratif *ça* : *C'est ça qui va être flatteur pour la maison Desvarenes !* (Sandfeld, 1977, p. 124).

<sup>28</sup> Seulement 72 *ceci que* dans le corpus « contemporain » de Frantext. Parmi lesquels seulement 37 constructions de type relatif, dont 6 clivées, et aucune avec seulement *ceci* en focus.

*le plus frappant : la tradition qui se manifestait derrière tant d'aisance...*  
(J. Semprun, *L'écriture ou la vie*, 1994)

ou référence de situation :

(55) *Il prend un crayon, un morceau de papier, une enveloppe, n'importe quoi et il dessine. « C'est ceci que vous voulez ? C'est entendu ». Et il file.* (E. Peisson, *Parti de Liverpool*, 1932)

La référence de *ceci* est plus rarement anaphorique, en particulier quand le contexte permet une sélection vers le dernier contenu support de référence possible, comme en (56) au segment débutant par *Mais* :

(56) *Vieillir, cela n'est rien. Mais cette seconde gâtée, ce calme différé, un peu plus loin encore, c'est ceci qui est fatigant.* (A. de Saint-Exupéry, *Courrier Sud*, 1928)

#### 4.4. *c'est ça qu-*

Les constructions de ce type sont essentiellement<sup>29</sup> des clivées, dans lesquelles *ça* n'est pas nécessairement l'antécédent direct de l'élément vide dans la subordonnée. Il existe beaucoup de constructions sur le modèle de *c'est à cause de ça que*, *c'est pour ça que*, *c'est comme ça que*, pour ne citer que les combinaisons les plus fréquentes :

(57) *Ah, c'est un homme comme ça que j'aurais voulu !* (J. Giono, *Olympe*, 1982)

Le pronom est anaphorique, et n'est pas antécédent de la subordonnée : *c'est un homme comme ça, ce que j'aurais voulu*.

En antécédent unique, le sens du pronom est équivalent à *cette chose*, donc à un nom vague à déterminant démonstratif :

(58) *Il n'y avait pas de hiérarchie possible. Et c'est ça que j'ai petit à petit mis au clair pour moi.* (J. P. Sartre, *Entretiens avec J.P. Sartre*, 1981)

<sup>29</sup> Sur le corpus Frantext « contemporain » (à partir de 1980), 1114 résultats ; sur les 200 premières occurrences de « ça que », 185 sont des clivées, aucune relative n'y figure.

- (59) *Tu veux que la police me tombe sur le dos ? Hein ? C'est ça que tu veux ?*  
(Y. Beauchemin, *le matou*, 1981)

Dans la plupart des constructions, la subordonnée donne au démonstratif une interprétation qui reste périphérique : le contenu sémantique est essentiellement anaphorique en (58), (59). On a évoqué plus haut la diversité sémantique des clivées, qui reste assez peu explorée ; il arrive assez souvent que les clivées soient simplement des moyens de focaliser un argument, sans engager une structure informative à focus nouveau distingué de la subordonnée présupposée. C'est l'interprétation qui se dégage de (58) : la subordonnée est informative : *et ça, je l'ai mis au clair...*

De même :

- (60) *On ne peut pas faire confiance à ces gens. C'est ça que tu ne veux pas comprendre.* (N. Mathieu, *Leurs enfants après eux*, 2018)

La construction n'est pas motivée par un arrière-plan plus ou moins présupposé, et une focalisation qui distingue le bon argument : *c'est cette chose et pas une autre*. Le sens est simplement : *ça, tu ne veux pas le comprendre*. D'ailleurs, le sens de *ça* en (60) ne diffère guère de la construction identique où *ce* serait utilisé, typiquement dans une relative « explicative » :

- (60') *On ne peut pas faire confiance à ces gens, c'est ce que tu ne veux pas comprendre*

L'interprétation anaphorique subsiste. Le sens de *ce* en (60') est cependant tellement vague qu'il semble bien proche d'un indéfini, *c'est quelque chose que, c'est une chose que*, avec la même relation anaphorique.

Sur le plan de la syntaxe, (60) est-il encore une clivée, avec la variété de constructions qui les caractérise ? Il semble possible d'utiliser la mobilité de position d'une préposition avec *ça* (mais pas avec *ce*) :

- (61) *On ne peut pas faire confiance à ces gens, c'est à ça qu'il fallait s'attendre / c'est (à) ça à quoi il fallait s'attendre*  
(62) *On ne peut pas faire confiance à ces gens, c'est (de) ça dont j'avais peur / c'est de ça que j'avais peur*  
(63) *On ne peut pas faire confiance à ces gens, c'est ce à quoi il fallait s'attendre / \*c'est à ce qu'il fallait s'attendre*  
(64) *On ne peut pas faire confiance à ces gens, c'est ce dont j'avais peur / \* c'est de ce que j'avais peur*

On y verra donc bien des clivées, même si elles semblent assez éloignées de la structure informative en focus et thème subordonné des clivées prototypiques, et fort proches sémantiquement des « présentatives » utilisant *ce*.

## Conclusions

On conclura de cet examen que le pronom neutre *ce* devant les relatives est toujours à considérer comme un pronom défini non démonstratif ou un nom nu, selon les contextes ; surtout, un outil à tout faire dans la syntaxe de la subordination : les relatives ayant *ce* pour antécédent sont restrictives, sans interférence déictique, et les constructions avec *c'est ce qu-* sont des relatives et non des clivées. Inversement, les autres pronoms neutres sont bien des démonstratifs, dont on a tenté de détailler le fonctionnement de recteurs de subordonnées qui restent restrictives, en ce qu'elles affectent significativement l'extension du groupe nominal qu'elles construisent, mais qui ne déterminent pas la définitude et l'unicité de cet ensemble, ces propriétés étant affectées par la « monstration » indépendamment du contenu lexical. Le démonstratif antécédent est donc à la fois enrichi sémantiquement par l'expansion que constitue celle-ci, et support d'une relation de désignation par proximité visant le contexte ou la situation, sans concurrence entre les deux modes de construction du sens.

La distribution des constructions nous a permis de percevoir ce qui est peut-être une évolution, l'utilisation surtout littéraire, à des fins expressives, des démonstratifs *ceci* et *cela* en-dehors des relations endophoriques ou de situation dans les relatives, en lieu et place de *ce*. C'est un phénomène analogue à celui que nous avons pu observer sur les complétives à préposition (Muller, 2023). Dans les clivées, il semble qu'on utilise assez souvent un type simplement focalisé de ces constructions, sans hiérarchie informative nette. La répartition des formes est assez conforme à ce qu'on en sait : *ceci* est environ dix fois moins fréquent que *cela* ; *ça* figure surtout dans les usages familiers et ne semble pas s'intégrer aussi facilement que les formes suffixées dans un groupe nominal démonstratif comme antécédent de relative, sauf usages exophoriques. L'opposition entre *ceci* et *cela* est ici assez secondaire<sup>30</sup>, même si *ceci* est plus fréquent comme support cataphorique. Lorsqu'elle est mise en œuvre, elle privi-

<sup>30</sup> Elle est plus marquée dans d'autres contextes, cf. Corblin (1987a : 90), qui attribue le trait « délimité » à *ceci* pour le distinguer de *cela*.

légie *ceci* pour signifier une grande proximité, et comme support de cataphore, semblant confirmer ce que dit Kleiber (2004a : 228–229), sur la nécessité pour la cataphore d'être très proche de son référent, ce qui la distingue nettement de l'anaphore.

## Références citées


- Apothéloz, D. (2018). Examen d'une famille de constructions : les constructions identificatives. *Scolia* 32, 13–41.
- Corblin, F. (1987a). *Ceci* et *cela* comme formes à contenu indistinct. *Langue française* 75, 75–93.
- Corblin, F. (1987b). *Indéfini, défini et démonstratif*. Droz.
- Corblin, F. (1990). Les groupes nominaux sans nom du français. Dans G. Kleiber & J. E. Tyvaert (éds), *L'anaphore et ses domaines. Recherches Linguistiques* 14 (63–80). Centre syntaxique de l'Université de Metz.
- Cornish, F. (2017). SN démonstratifs et anadeixis : sens « spatial » ou valeurs tributaires d'une stratégie pragmatique potentielle ?. *Journal of French Language Studies* 27, 215–239.
- Diessel, H. (2014). Demonstratives, Frames of Reference, and Semantic Universals of Space. *Language and Linguistics Compass* 8(3), 116–132.
- Gary-Prieur, M.-N. (1998). La dimension cataphorique du démonstratif. Étude de constructions à relative. *Langue française* 120, 44–50.
- Hirschbühler, P. & Labelle, M. (1990). *Celui* comme noyau de syntagme nominal. *Travaux de linguistique* 20, 109–122.
- Kęsik, M. (1989). *La cataphore*. PUF.
- Kleiber, G. (1984). Sur la sémantique des descriptions démonstratives. *Linguisticae Investigationes* 8(1), 63–85.
- Kleiber, G. (1998a). *Des cerisiers, ça fleurit au printemps* : une construction bien énigmatique. Dans E. Werner, R. Liver, Y. Stork & M. Nicklaus (éds), *Et multum et multa* (95–112). Gunter Narr.
- Kleiber, G. (1998b). Au générique : tout ça pour ça. Dans J. Pauchard & J. E. Tyvaert (éds), *La variation (domaine anglais). La généralité* (195–231). Presses Universitaires de Reims.
- Kleiber, G. (2004a). Anticipation, mémoire et démonstratifs cataphoriques. Dans R. Sock & B. Vaxelaire (éds), *L'anticipation à l'horizon du présent* (221–236). Pierre Mardaga.

- Kleiber, G. (2004b) Sémantique, référence et discours : le cas des démonstratifs cataphoriques spécifiques. Dans A. Auchlin *et al.* (éds), *Structures et discours. Mélanges offerts à Eddy Roulet* (231–245). Nota Bene.
- Maillard, M. (1974). Essai de typologie des substituts diaphoriques. *Langue française* 21, 55–71.
- Muller, C. (2003). Naissance et évolution des constructions clivées en « c'est...que... » : de la focalisation sur l'objet concret à la focalisation fonctionnelle. Dans P. Blumenthal & J. E. Tyvaert (éds), *La cognition dans le temps* (100–120). Niemeyer.
- Muller, C. (2018). L'emploi de *ce* dans les reprises de contenu propositionnel. *Scolia* 32, 117–138.
- Muller, C. (2020). Le *ce* antécédent des constructions de type relatif : un pronom démonstratif?. *Langue française* 205, 101–119.
- Muller, C. (2023). Les démonstratifs neutres du français en cataphore d'une subordonnée complétive. *Linguisticae Investigationes* 46(1), 18–40.
- Riegel, M., Pellat J. C. & Rioul, R. (2009). Grammaire méthodique du français. PUF.
- Sandfeld, K. (1977(1936)). *Syntaxe du français contemporain. Les propositions subordonnées*. Droz.



**Adriana Orlandi**

Université de Modène et Reggio d'Émilie  
Italie

 <https://orcid.org/0000-0002-6638-265X>

## Regards croisés sur l'adjectif

### Adjectives at the intersection of perspectives

#### Abstract

In this study, we offer an overview of the main issues addressed by Gaston Gross in his analysis of the adjectival category. Gross's aim was to create an electronic dictionary of adjectives intended for automatic language processing. His research work focused on the search for criteria for the identification and description of the various subclasses of adjectives, as well as on the accurate description of adjectival meaning using the notion of "emploi". We will discuss some key concepts of the *Théorie des classes d'objets* that have been applied to the study of the adjective. These include the notion of adjectival predicate and adjectival use ("emploi"). We will attempt to draw a summary of Gaston Gross's achievements, while also discussing some critical issues.

#### Keywords

Adjectives, adjectival predicates, relational meaning, polysemy, word classes

## 1. Introduction

Dans cette contribution, nous souhaitons rendre hommage à Gaston Gross (le linguiste et l'ami), à travers l'examen d'un des domaines de recherche sur lesquels il est revenu à plusieurs reprises au cours de sa vie professionnelle : les adjectifs. Notre objectif est de dresser un bilan de son activité de recherche à ce sujet, en analysant et en discutant les points saillants de sa méthode ainsi que les principaux résultats atteints. Nous aimerions montrer à quel point l'étude de l'adjectif a pu tirer profit de la Théorie des classes d'objets, sans négliger néanmoins de prendre en compte les limites de cette démarche.

Notre amitié avec Gaston Gross est née en 2002 lors du Colloque sur les Adjectifs non prédicatifs qu'il avait organisé à la Maison des Sciences de l'Homme à Paris. À cette époque, Gaston envisageait la constitution d'un dictionnaire électronique des adjectifs, et ses recherches se heurtaient à la nécessité d'appréhender le comportement d'un certain nombre d'adjectifs qu'il était difficile pour lui d'intégrer dans son dictionnaire en raison de leur non prédicativité. D'où l'organisation de ce colloque qui fut sans doute le premier à être entièrement consacré à la question du blocage prédicatif de l'adjectif, et qui contribua à accélérer le processus de description et de recensement des adjectifs non prédicatifs. Gaston Gross appela le colloque « Les adjectifs non prédicatifs », ce qui, entre autres, constitua d'un point de vue terminologique un signe de rupture par rapport au passé, dans la mesure où la tradition exigeait que l'on utilise l'opposition *épithète vs attribut*.

Le dictionnaire était destiné « à la traduction automatique, à l'introduction d'une synonymie automatique dans les logiciels de traitement de texte, à la recherche d'informations dans des textes ou sur le Web, à l'aide à la rédaction, etc. » (Gross *et al.*, 2001 : 1). Le projet était ambitieux et nécessitait d'un grand investissement de ressources (temps et financements). Malheureusement, il n'a pas pu être achevé.

Nous passerons ici en revue un certain nombre de points concernant l'étude de l'adjectif selon la perspective de la Théorie des classes d'objet. Nous nous bornerons à l'analyse de quelques questionnements majeurs. Le premier est celui des critères de délimitation de la catégorie adjectivale, que nous développerons dans la section §2. Ensuite, nous nous concentrerons sur la notion de prédicat adjectival (§3), qui est assez problématique dans l'étude de l'adjectif, justement à cause des adjectifs non prédicatifs. Pour finir, nous nous pencherons sur la notion d'« emploi » (§4), qui représente sans aucun doute la contribution principale de Gaston Gross à la réflexion sur l'adjectif.

## 2. Problèmes de délimitation de la catégorie adjectivale

Les études sur l'adjectif étant liées au projet de constitution d'un dictionnaire électronique, le premier souci de Gaston Gross était d'identifier des critères permettant la reconnaissance automatique des parties du discours. En ce qui concerne l'adjectif, Gaston Gross a souligné à plusieurs reprises que cette catégorie est plus difficile à reconnaître automatiquement que les noms et les verbes, car si pour les noms et les verbes il est possible d'identifier des propriétés morphologiques



distinctives, pour les adjectifs « il n'existe pas de propriétés morphologiques qui permettraient, indépendamment de la syntaxe, de délimiter à coup sûr un adjectif » (Gross, 2012 : 117). En effet, les noms ont un genre inhérent, masculin ou féminin, et peuvent être déclinés suivant le nombre, singulier ou pluriel. Les verbes, quant à eux, ont une flexion permettant d'exprimer le mode, le temps, la personne et le nombre. Tous les mots qui sont classés en tant que noms ou en tant que verbes possèdent les mêmes propriétés morphologiques. En revanche, pour les adjectifs, le critère morphologique n'est pas suffisant, car s'il est vrai que l'adjectif s'accorde en genre et en nombre avec le substantif, on compte de nombreuses exceptions à la règle : les adjectifs épiciens (*juste, calme*), et les adjectifs invariables (*snob, orange, anticellulite*).

Précisons tout de suite que pour Gaston Gross la possibilité de reconnaître un nom ou un verbe, tout en étant nécessaire, n'est pas suffisante. En effet, « [i]l est [...] tout à fait insuffisant pour l'analyse automatique d'une phrase [...] d'être en mesure de localiser et de reconnaître une forme verbale [...], car celle-ci peut correspondre à des fonctions syntaxiques très diverses » (Gross, 2012 : 21) : les locutions verbales, par exemple, ont un comportement syntaxique complètement différent par rapport aux constructions à verbe support. De même, pour le nom, « [l]'indexation morphologique automatique n'est utile que si elle est accompagnée de l'établissement de la fonction syntaxique de l'élément dans la phrase ou le texte » (*ibid.* : 22) : il existe en effet des substantifs qui fonctionnent comme des « arguments fondamentaux ou élémentaires » (*table*) et des substantifs qui, en revanche, génèrent eux-mêmes des arguments (*rêve*).

En ce qui concerne l'adjectif, Gross (2012) utilise le critère syntaxique d'un côté pour résoudre certaines ambiguïtés catégorielles, de l'autre côté pour regrouper les adjectifs en sous-classes homogènes. Ainsi montre-t-il que l'environnement syntaxique aide à « identifier le statut catégoriel du terme » (Gross, 2012 : 118) dans un certain nombre de cas limites, comme par exemple celui des adjectifs pouvant être interprétés comme des participes passés (*fatigué*), ou jouant le rôle de prépositions (*sauf, excepté*) ou d'adverbes (*fort* dans *il pleut fort*). Dans toutes ces situations, il est possible de réduire l'ambiguïté grâce à des tests syntaxiques.

En outre, la syntaxe, et plus particulièrement le critère distributionnel, est ce qui permet le repérage de trois sous-classes de l'adjectif : les adjectifs prédicatifs, les « classificateurs » (non prédicatifs) et les « arguments » (non prédicatifs).

Or, cette focalisation sur le critère syntaxique ne va pas chez Gaston Gross jusqu'à la reconnaissance du rôle que la syntaxe peut jouer dans la détermination des limites externes de la catégorie. Dans le volume des *Cahiers de Lexicologie* consacré à l'adjectif non prédicatif, Gross (2005a : 5) nie qu'il soit possible

de séparer de manière claire et univoque la classe des adjectifs des autres parties du discours :

Il n'existe aucune définition unique qui permette de les séparer des autres catégories grammaticales. Tous les adjectifs ne sont pas prédicatifs, tous n'ont pas de substantifs associés, tous ne sont pas scalaires, tous ne désignent pas des qualités, tous ne sont pas coordonnables à l'aide de la conjonction *et*. Dès lors qu'il n'existe pas de propriété définitionnelle de la catégorie, la seule démarche consiste à mettre au point une batterie de critères, de tester chaque adjectif au regard de l'ensemble de ces propriétés et de constituer des ensembles homogènes en regroupant ceux des adjectifs qui ont les mêmes propriétés.

(Gross, 2005a : 5)

Force est de constater qu'un critère distributionnel nous permettant de distinguer les adjectifs des autres parties du discours existe : c'est la possibilité pour tout adjectif d'apparaître dans un SN – généralement à la droite du nom – en tant qu'expansion du nom tête (cf. Orlandi 2020, §4.7). Comme l'observe Goes (1999 : 128), « [r]efuser la fonction épithète constitue [...] un refus du caractère adnominal tout court : les éléments en question ne sont plus des adjectifs ». À ce critère s'ajoute le critère morphologique de la concordance, qui n'est pas cependant toujours efficace. Le caractère adnominal est partagé par les déterminants ; néanmoins les déterminants ne sont pas des expansions mais des constituants essentiels du noyau du SN et ne risquent pas de se confondre avec les adjectifs (cf. Prandi, sous presse, pour un approfondissement sur la distinction entre noms, verbes, adjectifs et déterminants).

Du point de vue du traitement automatique des textes, le critère distributionnel aide à résoudre certaines ambiguïtés dues à la difficulté de distinguer le nom de l'adjectif dans le cas où la tête du syntagme et son expansion sont des lexèmes au double statut de nom et d'adjectif (par exemple, *militaire* et *idéal*, qui peuvent être à la fois des noms et des adjectifs). Dans ce cas, puisque la position non marquée pour l'adjectif (au moins dans les langues romanes) est après le nom, on assignera de préférence le statut d'adjectif à l'élément le plus à droite : cf. *un militaire idéal* vs *un idéal militaire* (exemple de Goes, 1999 : 102).

Bref, il est possible de nuancer l'affirmation selon laquelle « il n'existe pas de propriété définitionnelle de la catégorie [adjectivale] » (Gross, 2005a : 5) : le critère existe et c'est une combinaison entre une propriété morphologique et une propriété distributionnelle. Cela dit, comme c'est le cas pour les noms et pour les verbes, il est nécessaire de procéder à une description fine de chaque adjectif afin d'en permettre le traitement automatique dans les phrases et dans les textes.

### 3. Notion de prédicat adjectival

L'adjectif est essentiellement considéré par Gaston Gross comme un prédicat. Un prédicat est défini comme « un mot qui opère une sélection déterminée parmi les mots du lexique pour établir avec eux un schéma formant la base d'une assertion. Les arguments sont les éléments lexicaux ainsi sélectionnés par les prédicats dans le cadre de la phrase simple » (Gross, 2012 : 13). À cet égard, il convient de souligner que ce que Gross appelle *prédicat* devrait être appelé *prédicateur*, si l'on s'en tient à la distinction introduite par Lyons (1977 : 434) selon laquelle le prédicat est la relation grammaticale identifiée par le syntagme verbal, et le prédicateur est l'élément qui garantit la fonction prédicative, c'est-à-dire le verbe, le nom ou l'adjectif. Dans la suite de cet article, le terme *prédicat* dans la terminologie de Gaston Gross sera donc compris comme pivot prédicatif, et donc comme prédicateur au sens de Lyons (1977).

En ce qui concerne la classe des adjectifs, il s'agit en grande partie de prédicats. Ceci est confirmé par le fait que sur les 30.000 lignes prévues pour le dictionnaire électronique des adjectifs, environ 24.000 concernent des adjectifs prédicatifs (Gross, 2005a), ce qui représente la majorité des emplois adjectivaux (nous reviendrons plus loin sur la notion d'emploi). Être un prédicat signifie avoir des arguments. Dans le cas de l'adjectif, les arguments sont le sujet auquel l'adjectif s'applique dans le cadre d'une phrase simple (*[Cette tarte]<sub>0</sub> est bonne*), et un ou deux compléments éventuellement régis par l'adjectif (*[Luc]<sub>0</sub> est bon [en mathématiques]<sub>1</sub>*)<sup>1</sup>.

Gaston Gross n'a jamais ignoré le problème de la non prédicativité adjectivale, au contraire : non seulement il y a consacré le Colloque de 2002 mentionné dans l'Introduction, mais il a toujours considéré la prédicativité comme étant le critère de différenciation principal au sein de la catégorie de l'adjectif, voire comme « le paramètre d'analyse le plus important » (Gross, 2005a : 6). Néanmoins, dans ses écrits sur l'adjectif, il arrive qu'il attribue parfois aux seuls adjectifs prédicatifs des propriétés valables pour tous les adjectifs, et qu'en revanche il attribue à tous les adjectifs des propriétés qui ne sont valables que pour les adjectifs prédicatifs. Nous explorerons le premier cas de figure dans la section §3.1, et le deuxième dans la section §3.2.

---

<sup>1</sup> Pour une typologie des arguments des adjectifs cf. Gross (2012 : 121).

### 3.1. Adjectifs, prédicats et valeur relationnelle

Gross (2012 : 120) observe que « le sens d'un adjectif (et de façon générale d'un prédicat) change et cela de façon compacte, en fonction des classes sémantiques (*classes d'objets*) décrivant leurs arguments ». De toute évidence, Gaston Gross se réfère ici aux adjectifs prédicatifs puisqu'il fait référence aux arguments. Cependant ces adjectifs ne sont pas les seuls à changer de sens en fonction des classes d'objets auxquelles ils s'appliquent. La propriété en effet ne concerne pas de façon exclusive les adjectifs prédicatifs, mais tous les adjectifs en tant que termes insaturés. L'ambiguïté réside, à notre avis, dans le fait de considérer l'adjectif comme un prédicat, car s'il y a bien un dénominateur commun entre tous les adjectifs, ce n'est pas le fait d'être des prédicats (des prédicateurs), mais des termes non saturés, c'est-à-dire des expressions qui ont besoin d'être complétées par d'autres expressions afin de remplir leur fonction au sein de la phrase simple<sup>2</sup>. Les adjectifs étant des modificateurs, ils ont nécessairement besoin d'un substantif, que ce soit au sein du SN (adjectif épithète) ou d'une phrase copulative adjectivale (adjectif attribut)<sup>3</sup>.

Le caractère non saturé des adjectifs découle du fait que les adjectifs ont pour contenu des concepts relationnels, c'est-à-dire des concepts dont l'occurrence requiert la présence ou l'existence d'une autre entité (Croft, 1991 : 62–63). Cette caractéristique, que les adjectifs ont en commun avec les verbes, a été désignée de différentes manières au cours des siècles par les philosophes et les linguistes. Dans la tradition grammaticale occidentale, elle est connue sous le nom de *dépendance référentielle*. Comme l'affirme Riegel (1993 : 8), par exemple,

Une propriété comme 'rouge' est par définition la propriété de quelque chose [...]. C'est un concept descriptif dont les occurrences particulières (les cas de 'rouge(eur)') ne se réalisent que si elles sont 'supportées' par quelque chose qu'elles caractérisent et dont elles présupposent l'existence. Autrement dit, les propriétés n'ont pas l'autonomie référentielle caractéristique des concepts typants [les noms].

(Riegel, 1993 : 8)

<sup>2</sup> La notion de saturation a été introduite par Frege (1891) et par la suite transférée au domaine de l'analyse linguistique par Tesnière (1959). Prandi (2004 : 125) la définit comme suit : « An expression is saturated when it is able to perform its function without being completed by another expression. A noun phrase is thus a saturated expression, because it can be used as a referential expression as it is. At a higher level, a nuclear sentence is also a saturated expression, because it is capable of framing a process. An unsaturated expression, on the other hand, has to be completed in order to perform its function. A verb, for instance, is an unsaturated expression, because it can frame a process on condition that its free positions are saturated by appropriate arguments ».

<sup>3</sup> Pour des raisons d'économie, nous ne traiterons pas ici le cas de figure du complément prédicatif du sujet (*Paul se dit intelligent*) et de l'objet (*Je trouve Paul intelligent*).

Cette différence dans le degré d'autonomie référentielle entre les noms et les adjectifs, déjà soulignée par les auteurs de la *Logique* de Port Royal (1662[1992 : 40]), est également présente dans l'opposition entre *universaux typants* et *universaux caractérisants* formulée par Strawson (1959[1973 : 189]) :

Un universel typant fournit un principe pour distinguer et compter les particuliers individuels qu'il rassemble. Il ne présuppose aucune méthode ou principe antécédents pour individuer [...] les particuliers qu'il rassemble. Par contre, les universaux caractérisants, s'ils fournissent des principes pour grouper, et même pour compter, les particuliers, les fournissent uniquement pour des particuliers que l'on distingue déjà ou que l'on peut distinguer par l'emploi d'une méthode ou d'un principe antécédent. En gros, [...] certains noms communs qui désignent des particuliers introduisent des universaux typants, tandis que les verbes et les adjectifs qui s'appliquent à des particuliers introduisent des universaux caractérisants.

(Strawson, 1959[1973 : 189])

De même, pour Strawson (1985 : 60), « la référence aux attributs présuppose la référence aux objets mais non pas réciproquement ». Guillaume (1973 : 203), quant à lui, introduit l'opposition entre *incidence interne* et *incidence externe* : « le propre de l'adjectif, c'est de ne pas être incident à lui-même, mais à un support dont il n'emporte pas la prévision concrète ». Croft (1991 : 62–63) définit la dépendance référentielle sous le terme de *valence* (ou *relationalité inhérente*), emprunté à Tesnière (1959). Cette référence à une dimension « relationnelle » est également présente chez Prandi (2004 : 122–123) qui distingue les concepts *ponctuels* (ou *classifiants*) des concepts *relationnels*<sup>4</sup>. Alors que les concepts ponctuels « are concepts that subsume instances of beings under categories » (*ibid.* : 122), les concepts relationnels « impose relations on instances of beings » (*ibid.*).

Or, pour en revenir à la relation entre saturation et caractère ponctuel ou relationnel d'un concept, Prandi (2004 : 124) observe que : « [p]unctual concepts, which group kinds of beings, are meanings of saturated expressions – typically nouns – while relational concepts, which draw processes and properties, are meanings of unsaturated expressions – typically verbs and adjectives »<sup>5</sup>. Les adjectifs ont la caractéristique d'avoir un signifié relationnel. Cela veut dire que leur contenu ne se définit que sur la base de leur connexion avec un nom. Ainsi,

<sup>4</sup> La distinction entre concepts ponctuels et relationnels remonte à Aristote (*Les Catégories*, 5, 2a).

<sup>5</sup> Cela ne veut pas dire que tous les noms ont pour contenu un concept ponctuel. Le nom *description* est un nom de procès, non saturé, qui fonctionne exactement comme le verbe *décrire* : *Jean a fait la description du projet* = *Jean a décrit le projet*.

l'adjectif *grand* aura un sens différent selon qu'il s'accompagne à *maison*, à *joie* ou à *travailleur*<sup>6</sup>. Dans le premier cas, l'adjectif modifie un nom de <bâtiment> (*grande maison*), et il désigne la dimension ; dans le deuxième, il modifie un nom d'<émotion> (*grande joie*), et il désigne l'intensité<sup>7</sup> ; dans le troisième, il modifie l'<agent habituel> d'une action (*grand travailleur*), et il indique le haut degré dans l'accomplissement de l'action. Lorsque des synonymes sont disponibles, les différentes significations de l'adjectif (« emplois » au sens de Gross) sont également confirmées par la présence de synonymes différents, une *grande maison* étant une maison *vaste, spacieuse*, une *grande joie* une joie *intense, vive*, un *grand travailleur* un travailleur *appliqué, laborieux*.

Jusqu'à-là l'analyse recoupe bien la méthode développée dans la Théorie des classes d'objets. La seule différence est que, si l'on considère la fonction de *grand* dans les exemples donnés, seulement dans les deux premiers l'adjectif est un prédicat au sens de Gross, car ici l'adjectif peut devenir le pivot prédicatif d'une phrase copulative adjectivale :

- *Cette maison est grande.*
- *Ma joie est grande.*

Dans le troisième exemple (*grand travailleur*), en revanche, l'adjectif ne peut pas être considéré comme un pivot prédicatif à proprement parler, car le modificateur n'a pas la fonction de qualifier le référent du nom tête du syntagme, et de ce fait il est inapte à la position prédicative<sup>8</sup>. L'énoncé *Ce travailleur est grand* a clairement un sens différent, l'adjectif *grand* se référant ici à la taille du sujet.

Pour utiliser la terminologie de la Théorie des classes d'objets, on pourrait donc affirmer que l'adjectif *grand* dans l'expression *grande maison* et *grande joie* est un prédicat alors que dans *grand travailleur* ne l'est pas. Néanmoins, dans toutes les expressions, y compris la dernière, le sens de l'adjectif est le résultat de l'interaction avec le nom.

<sup>6</sup> Précisons que *grand* est un adjectif prototypique dont la position non marquée est avant le nom et non après.

<sup>7</sup> Dans *grande joie*, l'adjectif modifie un nom qui, à son tour, est un terme non saturé. L'adjectif est donc ici une expression non saturée au deuxième degré.

<sup>8</sup> Dans Orlandi (2020), nous avons montré que le blocage prédicatif de certains adjectifs s'explique non pas en termes sémantiques mais syntaxiques et fonctionnels. La phrase copulative adjectivale fonctionne comme un « moule rigide » (Blinkenberg, 1960) capable d'imposer la relation de qualification entre sujet et prédicat. Pour cette raison, les adjectifs qui n'ont pas pour fonction de qualifier sont exclus de la position prédicative, et lorsque l'adjectif peut avoir plusieurs interprétations, l'occurrence prédicative sélectionne la lecture qualificative.

Aussi, d'après la définition de prédicat que nous avons citée au début de cette section, l'adjectif est un prédicat lorsqu'il sélectionne un ou plusieurs arguments au sein d'une phrase simple. Les adjectifs non prédicatifs sont considérés comme des adjectifs « qui n'ont pas d'arguments, c'est-à-dire qui ne font pas partie de la classe des prédicats » (Gross, 2005b : 31). Parler des adjectifs en termes de « prédicats » et d'« arguments » nous amène à considérer *maison* et *joie* comme des arguments sélectionnés par *grand*, et *travailleur* comme un « non-argument », alors que de toute évidence ce substantif n'a simplement pas la possibilité d'occuper la place de sujet d'une phrase copulative adjectivale dans laquelle *grand* garderait le même sens qu'il a en position épithète (« qui travaille beaucoup »). Si *travailleur* n'est pas un « argument » au sens de Gross, il représente néanmoins l'expression permettant à l'adjectif d'être saturé. Il n'y a donc pas de différence, à nos yeux, entre les trois expressions du point de vue de la manière dont le signifié de l'adjectif se définit.

Gross (2005a : 7) ajoute à ceci que les outils d'analyse de la Théorie des classes d'objets « ne s'appliquent pas [aux] [...] adjectifs non prédicatifs ». Or, si une bonne partie des tests syntaxiques mis en place pour l'étude des prédicats adjectivaux ne sont sans doute pas applicables aux adjectifs non prédicatifs, il n'en reste pas moins vrai qu'une description de ces adjectifs au moyen des classes d'objets est possible. Cela dit, l'analyse des adjectifs non prédicatifs demande la mise au point d'outils d'analyse spécifiques (qui restent en bonne partie à étudier).

En définitive, nous estimons qu'analyser les adjectifs en termes de « prédicats » et « arguments » finit par gommer le dénominateur commun de la catégorie : le caractère relationnel des adjectifs. Autrement dit, il s'agit de prendre en compte le fait que les adjectifs sont des termes non saturés qui ont besoin non pas d'un « argument » mais d'un « terme saturé » pour accomplir leur tâche de modificateurs au sein d'une phrase simple. Selon le type de structure, le terme saturé peut être un substantif (1), un syntagme nominal (2), voire une proposition (3) :

- (1) *Il a une grande maison*
- (2) *Cette maison est grande*
- (3) *S'être mis dans cette situation est fou*

Le caractère relationnel de l'adjectif est une caractéristique qui est partagée par tous les membres de la catégorie, y compris les adjectifs non prédicatifs. Dès lors, il serait sans doute plus efficace de décrire l'adjectif au moyen du couple « saturé » / « non saturé » plutôt qu'au moyen du couple « prédicat » / « argument ».

Ajoutons également avec Prandi (2011 : 39), que « la vocation relationnelle n'implique pas l'absence ou la pauvreté de contenu lexical ; tout au contraire,

[...] la capacité de créer des relations est inséparable de la présence d'un contenu ». Avoir un signifié relationnel ne veut donc pas dire être dépourvu de signifié.

Enfin, un cas de figure à part est constitué par les expressions figées, qui se distinguent par leur sens figé. Le contenu de l'adjectif ne fait pas ici l'objet d'une négociation avec le contenu du nom. C'est pourquoi, comme Gaston Gross l'a souligné à plusieurs reprises, les adjectifs qui entrent dans des expressions figées doivent constituer des entrées distinctes au sein du dictionnaire électronique. Prenons l'exemple de *grande maison* et *grande mosquée*. Les substantifs *maison* et *mosquée* appartiennent à la classe d'objets des <bâtiments>. Cependant, alors qu'une *grande maison* est une maison de grandes dimensions, une *grande mosquée* ne désigne pas nécessairement une mosquée de grandes dimensions, mais une mosquée destinée à la prière du vendredi. Ici l'adjectif n'a pas la fonction de qualifier : *grand* a une valeur classifiante et typologisante, une grande mosquée étant un type particulier de mosquée. *Grande mosquée* n'a donc pas le même schéma d'arguments que *grande maison*, et doit être traité à part.

### 3.2. Épithète et fonction prédicative

Comme chacun sait, l'adjectif a deux rôles syntaxiques fondamentaux : la position épithète (*un enfant heureux*) et la position attribut (*cet enfant est heureux*). À ces deux rôles s'ajoute l'apposition (*l'enfant ouvre les cadeaux, heureux*), une forme de modification « en position détachée » (Riegel, Pellat & Rioul, 1994[2006 : 190]), où le modificateur adjectival se présente comme une expansion séparée par une pause bien marquée, que ce soit au moyen d'une virgule (à l'écrit) ou d'une rupture intonative (à l'oral).

Gross (2012 : 120) considère « les épithètes et les appositions comme des fonctions prédicatives, tout comme les attributs, à la différence près qu'ils ont perdu leur actualisation ». Cette affirmation repose sur l'idée selon laquelle dans la phrase *Donne-moi le cahier bleu*, l'épithète « est la réduction d'un emploi comprenant le verbe être : *Donne-moi le cahier qui est bleu* » (*ibid.*). La même analyse vaut pour l'apposition : dans la phrase *Les joueurs, fatigués, ont renoncé à l'entraînement du soir*, « le prédicat adjectival peut recevoir une actualisation qui lui est propre : *Les joueurs, qui étaient fatigués, ont renoncé à l'entraînement du soir* » (*ibid.*).

Or, si l'idée selon laquelle l'apposition représente une fonction prédicative paraît désormais rencontrer un consensus général de la part des linguistes, l'idée que l'épithète soit elle aussi une fonction prédicative, tout comme les attributs, est une généralisation qui ne concerne en réalité qu'une partie des adjectifs épithètes, à savoir les adjectifs prédicatifs.



En ce qui concerne l'apposition, nous nous bornons à rappeler que bien qu'elle soit souvent connue sous le nom d'« épithète détachée » (par exemple chez Le Goffic, 1993), elle peut être considérée comme « un prédicat externe secondaire de la phrase, déterminant sémantiquement le concept du SN fonctionnant comme base, sans restreindre l'extension de celui-ci » (Forsgren, 1988 : 138), comme cela arrive pour l'adjectif attribut. Selon Gaston Gross, l'apposition peut s'analyser comme la réduction d'un emploi comprenant le verbe support *être*. Le caractère secondaire de la prédication est dû au fait que l'apposition s'insère à l'intérieur d'une prédication principale, par définition primaire, dont elle dépend.

Quant à l'adjectif épithète, l'idée que ce dernier soit une fonction prédicative consistant dans la réduction d'un emploi comprenant le verbe *être*, fait écho aux Générativistes. La position de la Grammaire générative est bien résumée par Ruwet (1967 : 406) qui observe : « Quant aux adjectifs en position d'épithète, ils sont construits, par des transformations généralisées, à partir des adjectifs en position attribut ». La même perspective est présente chez Riegel (1985 : 13) dans le célèbre passage affirmant que : « l'adjectif qualifiant semble essentiellement voué à la fonction attribut [...]. C'est même sa seule fonction si l'on admet que ses deux autres constructions, épithète et apposée, sont dérivées d'une relative attributive enchâssée ».

Or, l'hypothèse d'une dérivation prédicative de l'adjectif épithète, qui, chez Gaston Gross, est sans doute redevable de sa vision de l'adjectif comme « prédicat », a été contestée depuis longtemps. On a pu constater en effet que cette hypothèse n'est vérifiée que pour un sous-ensemble des adjectifs. Chez les théoriciens de la Grammaire générative, il est désormais reconnu que les adjectifs épithètes ne peuvent pas tous dériver d'une relative enchâssée, les adjectifs se divisant ainsi en deux groupes : ceux pour lesquels cette transformation est possible et ceux qu'on appelle « direct modifiers of the NP » (Cinque, 2010)<sup>9</sup>, que l'on ne peut en aucun cas dériver d'une relative enchâssée. Comme Cinque (2014 : 2) le souligne, l'idée que certains adjectifs peuvent dériver d'une relative enchâssée alors que d'autres ne le peuvent pas a une longue tradition<sup>10</sup>.

À ces deux classes d'adjectifs sont associées un certain nombre de propriétés (cf. Cinque 2010 et 2014 pour une description détaillée), comme par exemple le

---

<sup>9</sup> Les adjectifs appelés « direct modification adjectives » sont définis comme des « functional elements [...], with phrasal status, merged as specifiers of distinct heads of the extended projection of the NP [...], not derivable from (reduced) relative clauses » (Cinque, 2010 : 43). Nous signalons au passage que pour Prandi (2004) la « modification directe » désigne le phénomène opposé, à savoir la modification par attribution d'une qualité.

<sup>10</sup> Il cite à ce propos les études de Winter (1965), Bolinger (1967), Motsch (1967), Berman (1974), Levi (1975), Emonds (1976) et Siegel (1976).

caractère non restrictif et non intersectif de la « direct modification », qui s'oppose au caractère restrictif et intersectif des adjectifs dérivés d'une relative enchâssée. D'après Cinque (2010, 2014), dans les langues romanes les adjectifs se caractérisant par une « direct modification » correspondent aux adjectifs antéposés, alors que les adjectifs postposés gardent toujours une certaine ambiguïté entre les deux sources d'interprétation ; c'est pourquoi *les très hauts bâtiments de New York* et *les bâtiments très hauts de New York* peuvent ne pas avoir la même extension, les deuxièmes ne pouvant désigner que les bâtiments « qui sont hauts ».

Le résultat le plus intéressant des recherches menées dans le cadre de la Grammaire Générative est le constat que « the principal divide is between predicative and non-predicative adjectives, which correspond in the adnominal case to the relative clause and direct modification sources, respectively » (Cinque, 2014 : 1). Les adjectifs prédicatifs seraient donc les seuls à pouvoir s'analyser comme le résultat de la dérivation d'une relative enchâssée, alors que les adjectifs non prédicatifs seraient exclus de cette transformation. Ce résultat ne doit pas surprendre. Les adjectifs non prédicatifs ne pouvant pas apparaître dans une phrase copulative adjectivale, il est normal qu'ils soient exclus de la possibilité de dérivation d'une phrase relative. Si on considère la phrase *Il se promène avec son éternel chapeau noir*, l'adjectif *éternel* s'analyse comme étant un adjectif non prédicatif, et plus précisément, suivant la dénomination de Bolinger (1967), comme un modificateur « de la référence » (*reference modifier*). Ces adjectifs, d'habitude antéposés, sont soumis au blocage prédicatif et ne peuvent pas apparaître dans des constructions comme *\*Son chapeau noir est éternel*. Il faut donc s'attendre à ce qu'ils soient également exclus d'une relative enchâssée : *\*Il se promène avec son chapeau noir qui est éternel*.

L'affirmation de Gaston Gross selon laquelle l'épithète serait une fonction prédicative ayant perdu son actualisation ne peut donc pas être généralisée : elle ne s'applique qu'à un sous-ensemble des adjectifs qui est l'ensemble des adjectifs prédicatifs.

#### 4. Notion d'emploi

L'emploi, défini comme « un schéma prédicatif donné et l'ensemble des propriétés qui lui sont rattachées » (Gross, 2012 : 33), est l'une des notions-clé de la Théorie des classes d'objets. Cette notion est appliquée à la catégorie de l'adjectif comme suit :

Si l'on parle d'adjectifs ici, il convient de préciser qu'il ne s'agit pas d'adjectifs morphologiques, c'est-à-dire les habituelles entrées de dictionnaire, mais d'emplois, c'est-à-dire des différents « sens » de chaque adjectif. Par exemple, il ne suffit pas de dire que l'adjectif *juste* est un appréciatif de valeur (*Son comportement a été juste*), ce qui ne constitue qu'un de ses sens, il peut aussi être un adjectif aléthique (*Cette information est juste*), un adjectif d'adéquation à une norme (*Ce calcul est juste*), un adjectif de taille (*Ce pantalon est juste*), de bon fonctionnement (*Cette pendule est juste, ce violon est juste*), etc. La mise en évidence de chacun de ces sens n'est possible que si pour chaque adjectif morphologique on a, dans un premier temps décrit l'ensemble de ses schémas d'arguments. Nous appelons *emploi* chacune de ces significations caractéristiques d'un schéma d'arguments spécifique.

(Gross, 2005a : 5–6)

Nous avons déjà en partie discuté cette notion d'emploi dans Orlandi (2020). Nous reprenons ici l'essentiel de ces réflexions car elles nous permettent de faire le point sur l'approche de l'adjectif développée par Gross.

Tout d'abord, l'idée que l'emploi coïncide avec la signification pertinente d'un mot n'est pas nouvelle, mais les visions divergent quant à la nature de l'emploi. Pour Wittgenstein (1953 : §43), « [l]a signification d'un mot est son emploi dans le langage ». Dans un effort de reconduire les mots « de leur usage métaphysique à leur usage quotidien » (*ibid.* : rem. 116), Wittgenstein considère l'emploi non pas comme une propriété systématique du langage, mais comme un fait contingent, à savoir l'utilisation contextuelle d'un mot. En linguistique, ce point de vue est partagé entre autres par les pragmaticiens ainsi que par les sémanticiens s'appuyant sur la théorie de la Sémantique interprétative (Rastier, 1987, 1994, 2001), qui vont jusqu'à nier l'existence même du signifié.

Gaston Gross a une vision de l'emploi très éloignée de cette conception éphémère de la signification tout imprégnée de connotations pragmatiques. Il a le mérite de ramener l'emploi à une dimension non contingente, le définissant comme une signification liée à un schéma d'arguments. À titre d'exemple, voyons quelques emplois des adjectifs *juste* et *âpre* (tirés de Gross, 2005b : 15–16) :

<i>juste</i> /N0 : hum	<i>âpre</i> /N0 : <objet concret>
<i>juste</i> /N0 : hum/N1 : avec hum	<i>âpre</i> /N0 : <lieu géologique>
<i>juste</i> /N0 : action<récompense, punition>	<i>âpre</i> /N0 : <fruit>
<i>juste</i> /N0 : action<calcul, déduction>	<i>âpre</i> /N0 : <dispute>
<i>juste</i> /N0 : <vêtement>	<i>âpre</i> /N0 : <propos>
<i>juste</i> /N0 : <instr. de mesure>	<i>âpre</i> /N0 : <combat>
<i>juste</i> /N0 : <instr. de musique>	<i>âpre</i> /N0 : <phénomène météo>

On voit bien que les emplois des deux adjectifs sont complètement soustraits à la dimension contextuelle de la communication. Comme l'écrivent Fasciolo et Zheng dans ce volume, « [u]ne telle notion d'emploi [...] n'est pas discursive, énonciative ou pragmatique, mais sémantique et stable ». Aussi, le fait que chaque schéma d'arguments donne lieu à un emploi différent (à une différente signification de l'adjectif) reflète le caractère relationnel de la signification adjectivale, le substantif qui sature l'adjectif étant ici représenté par sa classe d'objets d'appartenance indiquée en chevrons.

Gross remplace donc l'adjectif morphologique par ses différents emplois, ce qui constitue sans aucun doute un pas en avant dans l'étude de cette catégorie. La littérature sur l'adjectif a souvent mentionné la nécessité d'abandonner la notion d'adjectif au profit d'un autre concept appelé *emploi*, *comportement*, *valeur* (Goes, 1999), ou encore *lecture* (Nowakowska, 2004), mais ces réflexions n'ont jamais fait l'objet d'une théorisation de la notion d'emploi et n'ont pas réussi à prendre véritablement en compte la dimension relationnelle de l'adjectif. Goes par exemple observe que le terme *adjectif de relation*<sup>11</sup> ne délimite pas tant un sous-groupe spécial d'adjectifs, mais « un comportement particulier aux adjectifs dénominaux » (*ibid.* : 256). Ainsi l'adjectif *volcanique* a-t-il une valeur relationnelle dans le syntagme *Pierre volcanique* et qualificative dans *caractère volcanique*. Selon Goes (2011 : 185), « nous avons [...] affaire à une polysémie contextuelle ». De même, Nowakowska (2004 : 40) observe à propos de la double lecture (relationnelle et qualificative) de l'adjectif *enfantin* (*l'école enfantine* vs. *la naïveté enfantine de ma mère*) que :

[s]'agissant à l'évidence du « même » adjectif, on devrait [...] généraliser le terme de « lecture » ou celui d'« emploi » et décrire *enfantin* comme un adjectif polysémique pouvant actualiser une lecture relationnelle ou une lecture qualificative. Cela reviendrait à renoncer à parler d'« adjectif » relationnel.

(Nowakowska, 2004 : 40)

Dans ce passage, on retrouve la même référence à la polysémie, vue ici comme un ensemble de signifiés potentiels à actualiser : les emplois sont essentiellement considérés comme des signifiés inhérents, la saturation du substantif n'intervenant que pour actualiser l'un d'eux.

---

<sup>11</sup> Il convient de noter qu'en linguistique, le terme *relationnel* est utilisé dans deux sens différents, soit pour désigner une classe d'adjectifs (*adjectifs relationnels* ou *adjectifs de relation*), soit pour désigner une propriété des signifiés et des expressions. Rappelons que les concepts relationnels sont des signifiés d'expressions non saturées, généralement des verbes et des adjectifs.

Cette idée est développée de manière encore plus directe chez Marengo (2011) qui remplace la notion d'« emploi » ou de « lecture » par celle de « lexie », celle-ci étant « soit un mot pris dans une acception bien spécifique (= un lexème), soit une locution prise dans une acception bien spécifique (= un phrasème complet ou un quasi-phrasème) » (*ibid.* : 16). Marengo considère qu'« un même vocable peut contenir des lexies qualificatives et des lexies non qualificatives » (*ibid.* : 17), et il parle lui aussi de « polysémie, voire d'homonymie » (*ibid.*).

Or, selon nous, parler de polysémie réduit la question du signifié adjectival à la seule dimension lexicale inhérente, et efface ou ignore la dimension relationnelle de l'adjectif. La notion d'emploi développée par Gaston Gross, en revanche, prend en compte cette dimension, car l'emploi est toujours associé à un schéma d'arguments spécifique (et à l'ensemble de propriétés impliquées par ce dernier). Prandi (2011) observe que le concept d'emploi est investi d'une fonction négative lorsqu'il n'est vu que comme un sélectionneur de composants virtuellement présents dans la structure sémantique du lexème. En revanche, l'emploi est chargé d'une fonction positive lorsqu'il a « un rôle actif, de constructeur et organisateur d'acceptions distinctes » (*ibid.* : 42). La notion de lexie adoptée par Marengo (2011) nous paraît incarner le premier type de vision. Dans le modèle de Gross, en revanche, l'emploi est ce qui définit et organise les différentes acceptions d'un adjectif, la polysémie n'étant déterminée qu'à *posteriori*, à partir des emplois de l'adjectif. Cela veut dire que pour étudier le sémantisme d'un adjectif, il faudra d'abord dresser la liste de tous ses emplois.

Or, la notion d'emploi telle qu'elle est définie dans la Théorie des classes d'objets soulève au moins trois points qu'il peut être intéressant de discuter.

Le premier est le rapport entre l'emploi et la dimension extra-linguistique, plus particulièrement la dimension conceptuelle. Il nous semble, en effet, que sur ce point la perspective de Gaston Gross a évolué au fil du temps. En 2001, en présentant le projet de constitution du dictionnaire électronique des adjectifs Gaston Gross observait que :

Il y a [...] des prédicats qui ont pour argument n'importe quel substantif. C'est le cas, par exemple, de *intéressant*. Il y a ensuite des prédicats qui sélectionnent de grandes classes sémantiques comme les humains (*réfléchi*), les animaux (*sauvage*), les végétaux (*fané*), les concrets (*volumineux*), les locatifs (*vaste*) ou les noms de temps (*indéfini*). Il y a enfin des prédicats qui ont des arguments appartenant à des sous-classes plus précises que nous appelons des classes d'objets [...]. Ils décrivent avec précision la nature des arguments compatibles avec un prédicat donné.

[...] Nous voudrions signaler que ces définitions reposent sur des propriétés syntaxiques et ne sont en aucune façon des *primitifs sémantiques*<sup>12</sup>.

(Gross *et al.*, 2001 : 6)

Comme la citation le montre, la Théorie des classes d'objets à cette époque était réfractaire à l'inclusion de la composante conceptuelle dans sa dimension formelle, la distinction entre les humains, les animaux, les végétaux, les concrets, les locatifs – qui, dans cette théorie, sont appelés *hyperclasses* – étant uniquement basée « sur des propriétés syntaxiques ».

Cette position est à nouveau présente dans Gross (2012 : 75), où l'auteur observe que « la construction des classes [d'objets] se fonde sur des propriétés grammaticales et non sur une classification présyntaxique ». Les classes d'objets désignent, donc, « des ensembles de mots, et non des entités extra-linguistiques » (*ibid.*).

Quelques années plus tard, dans un article écrit avec Fasciolo (Fasciolo & Gross, 2014), l'optique évolue. Les auteurs observent que :

définir la signification d'un substantif signifie le mettre en relation avec un ensemble de prédicats dans le cadre d'un schéma prédicatif. En ce sens, notre méthode est strictement linguistique. [...] Si l'on devait expliquer ce qu'est un *miroir*, par exemple, il serait tout à fait naturel d'énumérer les actions que l'on peut faire avec un miroir. Répondre à la question ontologique, extralinguistique *Qu'est-ce que c'est?* s'avère alors, en dernière analyse, analogue à dresser une liste d'opérateurs appropriés. Ces opérateurs sont linguistiquement codés, mais leur codage est motivé par des distinctions conceptuelles ontologiquement ou culturellement saillantes.

(Fasciolo & Gross, 2014 : 138)

La dimension ontologique semble ainsi se frayer un chemin dans la caractérisation des opérateurs constitutifs d'un schéma prédicatif. Les retombées sur le plan méthodologique sont importantes : « Nous suggérons », soulignent les auteurs dans une affirmation à caractère programmatique, « que la voie la meilleure pour accéder au domaine extralinguistique [...] est la démarche intralinguistique » (*ibid.* : 138). En effet, « bien qu'une telle démarche soit purement linguistique, elle ne construit pas un concept *en langue* distinct de ce qui est dans le monde » (*ibid.* : 139)<sup>13</sup>. Il nous semble donc que la théorie fondée par Gaston

<sup>12</sup> C'est nous qui soulignons.

<sup>13</sup> Cf. également Fasciolo (2023, chapitre 14).

Gross a su évoluer de manière significative vers l'intégration de la composante conceptuelle dans son modèle de description du lexique, se rapprochant davantage d'une architecture formelle reposant sur des fondations conceptuelles.

Le deuxième point que nous aimerions souligner concerne une limite de la notion d'emploi adjectival. L'observation est simple : s'il est vrai que l'emploi se définit comme une signification (adjectivale) incarnée par un schéma d'arguments spécifique, cette notion ne saurait s'appliquer qu'aux adjectifs prédicatifs. Comme on l'a vu précédemment, en effet, les adjectifs non prédicatifs ne peuvent pas être considérés comme des prédicats, ce qui empêche en principe la possibilité d'une description par le biais des schémas d'arguments, donc suivant la notion d'emploi. C'est la raison pour laquelle Gross *et al.* (2001 : 4) affirment que les adjectifs non prédicatifs sont exclus du dictionnaire électronique et seront traités ailleurs. Seuls sont gardés les adjectifs prédicatifs, qui peuvent avoir les trois fonctions habituellement attribuées aux adjectifs : attribut, épithète, apposition. L'impossibilité de prendre en compte, dans le dictionnaire électronique, les adjectifs non prédicatifs représente sans aucun doute une anomalie dans un dictionnaire, qu'il soit électronique ou non : on ne voit pas pourquoi, en effet, ces adjectifs devraient être exclus et traités à part. Si l'on accepte que les adjectifs non prédicatifs sont des termes insaturés ayant besoin d'un substantif, comme les autres adjectifs, on pourrait essayer d'utiliser le modèle des classes d'objets pour leur description et les réintégrer dans le dictionnaire. Par exemple, l'adjectif non prédicatif *soi-disant* pourrait avoir deux emplois distincts selon qu'il s'utilise avec des humains (*les soi-disant amis du père Grandet*) ou des inanimés (*les soi-disant cadeaux que vous m'avez faits ; la soi-disant liberté de pensée*). Dans le premier cas, il signifie « Qui se dit, qui prétend être tel » (Petit Robert), dans le deuxième, il signifie « qui n'est pas ce qu'on en dit, qui n'est pas vraiment » (Petit Robert), et il a comme synonyme l'adjectif *prétendu*. Dans la traduction italienne, les deux emplois se différencient par le fait que seul le premier peut se traduire par *supposto*, alors que le deuxième n'a pas cette possibilité. Les deux emplois pourraient donc être décrits comme suit :

*soi-disant/N0: hum/ It: sedicente, cosiddetto, supposto*

*soi-disant/N0: ina/S: prétendu/ It: sedicente, cosiddetto*

Une fois de plus, on s'aperçoit qu'il n'y a pas de coïncidence entre « non saturé » et « prédicatif » : d'une part, les noms saturés peuvent avoir des emplois prédicatifs (*Max est un chien*) ; d'autre part, il existe des adjectifs non saturés qui n'admettent pas d'emploi prédicatif : c'est le cas de *soi-disant*. Ainsi faudrait-il analyser les classes d'objets et les hyperclasses décrites par Gaston Gross non pas

comme de possibles arguments d'un prédicat, mais comme des expressions saturant des expressions non saturées.

Le troisième point que nous souhaitons mettre en évidence consiste dans le rappel que toute la description du lexique, y compris des adjectifs, menée au sein de la Théorie des classes d'objets se base sur l'idée que « l'emploi d'un lexème est sa distribution à l'intérieur d'un modèle de phrase au signifié cohérent et approprié » (Prandi, 2011 : 42). Concernant l'adjectif, cela veut dire que l'emploi circonscrit les différentes valeurs que celui-ci reçoit dans le cadre d'une relation cohérente avec le substantif. La notion d'emploi ne peut donc pas être utilisée pour la description des adjectifs figurant dans une connexion conflictuelle avec le substantif, et recevant une interprétation figurée. La notion d'« emploi figuré » chez Gaston Gross n'existe purement et simplement pas<sup>14</sup>. Ce choix est parfaitement justifié par la nature lexicologique et lexicographique du travail de recherche de Gross. Ces deux branches de la linguistique ne sont en effet censées être soumises qu'à la seule contrainte de la cohérence, l'étude des figures étant de pertinence de la rhétorique ou de la grammaire. Néanmoins, la description des emplois telle qu'elle est conçue dans la Théorie des classes d'objets pourrait avoir d'intéressantes retombées et applications dans le domaine de la détection automatique des figures de rhétorique, ou du moins des figures issues d'un conflit (que ce soit de nature conceptuelle, cognitive ou lexicale). Puisque les emplois isolent des schémas prédictifs cohérents, leur application aux textes pourrait consentir le repérage automatique de connexions conflictuelles, dégageant de possibles candidats au statut de figure de rhétorique.

## 5. Conclusion

Les réflexions que Gaston Gross a menées sur l'adjectif ont permis de faire avancer les études sur cette partie du discours. L'alliage qu'il a prôné entre lexique et syntaxe a permis de dégager une vision de la signification complètement soustraite à la composante contextuelle et contingente de la communication. Sa notion d'emploi et sa méthode de description des signifiés de l'adjectif restent parmi ses contributions majeures à l'étude de l'adjectif, constituant un héritage important pour tous ceux qui seront en mesure de saisir l'importance de ce travail.

---

<sup>14</sup> Précisons que les catachrèses et les métaphores conceptuelles sont considérées comme des emplois cohérents.



En même temps, le bref aperçu que nous avons tenté d'offrir dans cet article nous donne également une idée du travail qu'il reste à accomplir. D'une part, il reste le problème du positionnement des adjectifs non prédicatifs dans le cadre théorique développé par Gross. D'autre part, et c'est là le défi majeur, il reste à achever le dictionnaire des adjectifs, défi qui sera, espérons-le, relevé par ses disciples. En tout état de cause, les études de Gaston Gross restent un instrument de connaissance important et témoignent de la passion sans limite qu'il a toujours nourrie pour l'étude du langage et la compréhension de ses mécanismes complexes.

## Références citées

- Aristote (2001). *Catégories* (Texte établi et traduit par R. Bodéüs). Les Belles Lettres.
- Arnauld, A. & Nicole, P. (1662[1992]). *La logique ou l'art de penser*. Gallimard.
- Berman, A. (1974). Adjectives and Adjective Complement Constructions in English. *Report No. NSF-29 to the National Science Foundation*. Department of Linguistics, Harvard University.
- Blinkenberg, A. (1960[1969]). *Le problème de la transitivité en français moderne*. Munksgaard.
- Bolinger, D. (1967). Adjectives in English: attribution and predication. *Lingua* 18, 1–34.
- Cinque, G. (2010). *The Syntax of Adjectives. A Comparative Study*. MIT Press.
- Cinque, G. (2014). The Semantic Classification of Adjectives. A View from Syntax. *Studies in Chinese Linguistics* 35(1), 1–30.
- Croft, W. (1991). *Syntactic Categories and Grammatical Relations*. The University of Chicago Press.
- Emonds, J. (1976). *A Transformational Approach to English Syntax*. Academic Press.
- Fasciolo, M. & Gross, G. (2014). Classifications linguistiques vs classifications ontologiques. *Travaux de linguistique* 69(2), 129–144.
- Fasciolo, M. (2023). *Les présuppositions repensées*. Classiques Garnier.
- Forsgren, M. (1988). Apposition adnominale: déterminants et ordre des constituants. *Travaux de linguistique* 17, 137–157.
- Frege, G. (1891). *Funktion und Begriff*. Hermann Pohle.
- Goes, J. (1999). *L'adjectif. Entre nom et verbe*. Éditions Duculot.
- Goes, J. (2011). Les adjectifs de repérage temporel: un troisième type ou un troisième emploi? Dans J. Goes et al. (éds), *Temps, aspect et classes de mots: études théoriques et didactiques* (179–200). Artois Presses Université.
- Gross, G. (2005a). Introduction. *Cahiers de Lexicologie* 86(1), 5–10.

- Gross, G. (2005b). Un dictionnaire électronique des adjectifs du français. *Cahiers de Lexicologie* 86(1), 11–33.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses Universitaires du Septentrion.
- Gross, G. et al. (2001). Un dictionnaire électronique des adjectifs prédicatifs sur une base sémantique. *Communication au Colloque international du CRISCO, L'Adjectif en français et à travers les langues, 28-30 juin 2001*. Université de Caen, 1–21.
- Guillaume, G. (1973). *Principes de linguistique théorique*. Klincksieck/Presses Universitaires de Laval.
- Le Goffic, P. (1993). *Grammaire de la phrase française*. Hachette.
- Levi, J. (1975). *The syntax and semantics of non-predicating adjectives in English*. Thèse de doctorat, University of Chicago.
- Lyons, J. (1977). *Semantics*. Cambridge University Press.
- Marengo, S. (2011). *Les adjectifs jamais attribués. Syntaxe et sémantique des adjectifs constructeurs de la référence*. De Boeck–Duculot.
- Motsch, W. (1967). Können attributive Adjektive durch Transformationen erklärt werden?. *Folia Linguistica* 1, 23–48.
- Nowakowska, M. (2004). *Les adjectifs de relation employés attributivement*. Wydawnicwo Naukowe Akademii Pedagogicznej.
- Orlandi, A. (2020). *Le paradoxe de l'adjectif*. De Boeck Supérieur.
- Prandi, M. (2004). *The Building Blocks of Meaning*. John Benjamins.
- Prandi, M. (2011). Les mots entre forme et substance : la dimension relationnelle du lexique. *Cahiers de lexicologie* 99(2), 35–48.
- Prandi, M. (sous presse). La distinzione tra nomi, verbi e aggettivi: morfologia, sintassi, strutture concettuali e pragmatica. Dans A. Hornung & R. Zanin (éds), *Abenteuer Sprache und Sprachen. Festschrift fuer Hans Drumbl zum 80. Geburtstag*. Stauffenburg.
- Rastier, F. (1987). *Sémantique interprétative*. Presses Universitaires de France.
- Rastier, F. (1994). Sur l'immanentisme en sémantique. *Cahiers de linguistique française* 15, 325–335.
- Rastier, F. (2001). *Arts et sciences du texte*. Presses Universitaires de France.
- Riegel, M., Pellat, J.-C. & Rioul, R. (1994[2006]). *Grammaire méthodique du français*. Presses Universitaires de France.
- Riegel, M. (1985). *L'adjectif attribut*. Presses Universitaires de France.
- Riegel, M. (1993). Grammaire et référence : à propos du statut sémantique de l'adjectif qualificatif. *L'information grammaticale* 58, 5–10.
- Ruwet, N. (1967). *Introduction à la grammaire générative*. Plon.
- Siegel, M. E. A. (1976). *Capturing the Adjective*. Thèse de doctorat, University of Massachusetts.
- Strawson, P.-F. (1959[1973]). *Les individus. Essai de métaphysique descriptive*. Éditions du Seuil.

Strawson, P.-F. (1985). *Analyse et métaphysique*. Vrin.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.


Winter, W. (1965). Transforms without Kernels. *Language* 41, 484–489.

Wittgenstein, L. (1953[2004]). *Recherches philosophiques*. Gallimard.



*José A. Pascual Rodríguez*

Real Academia Española, Madrid  
España

 <https://orcid.org/0000-0003-3887-0215>

# La caracterización lingüística de los textos. El criptoaragonesismo *servar* como pretexto

Linguistic characterization of texts. “*Servar*”, a crypto-aragonesism, as a pretext

## Abstract

This article shows the need to linguistically characterize texts that constitute a corpus in order to enable the historical study of words. Examples about the challenges found in the process are provided by showing the treatment given to fifty documents. These texts allow us to explain the Aragonese origin of the “*servar*” Latinism used by the Marquis of Santillana.

## Keywords

Philology, history of words, corpus

## 1. Introducción

Hace un tiempo que Gaston Gross me propuso pronunciarme en esta misma revista sobre el futuro que le esperaba al tipo de microcirugía filológica aplicada a los textos con la que me entretenía y trato aún de entretenerme. Sin mucha seguridad me referí a que, frente a la dificultad que habíamos tenido en el pasado los filólogos positivistas para ir allegando algunos datos, ficha a ficha, por el único expediente de extraerlos de nuestras lecturas, nos encontrábamos ahora en los corpus —y no digamos nada, en la red— con una masa de ejemplos como caídos del cielo, cuyo problema mayor era la dificultad para manejarlos. Pero estos, que hasta hacía poco podíamos tomar como indicios que nos orientaban para

comprender mejor la historia del léxico, se han convertido en argumentos que nos permiten ser mucho más precisos (Pascual 2021: 2). En esta situación no me ha resultado difícil explicar la apertura del marqués de Santillana y sus seguidores al aragonés y al catalán, mostrando con ello en sus escritos su atención a esta realidad lingüística no castellana (Pascual 2023a y, sobre todo, 2023b). Me consta que con ello no he alterado los cauces por los que ha discurrido la historia del castellano, pero soy consciente también de que he entendido un poco mejor algunos matices del color de la lengua literaria del Cuatrocientos.

Claro está que estos matices orientan sobre el discurrir en el tiempo de una lengua que dista mucho de haber estado constreñida en su larga historia por la autarquía, como podría parecer que ocurre, por poner un ejemplo, cuando en el repertorio de voces españolas que contiene el *Diccionario histórico de la lengua española* de la Real Academia Española de 1960–1996 (DHLE) aparece la palabra *aljez*, registrada casi exclusivamente en el ámbito del aragonés y de sus áreas de influencia (lo que ocurre igualmente en el DECH, s. v. *yeso*, aunque aquí por falta de datos).

Que las cosas las han facilitado mucho los datos de los que ahora disponemos, lo muestra el hecho de que, mientras que aquellos con que contábamos en el DECH (s. v. *conservar*) no podían llevarnos más lejos que a suponer que el cast. *servar* se tomara en el S. XV del lat. *Servare*; no resulta ahora arriesgado formular una hipótesis más precisa, sustentada en los 167 ejemplos<sup>1</sup> en casi medio centenar de textos que proporciona el CDH de textos orientales u orientalizantes de los siglos XIV y XV, frente a los 9, en 4 documentos, que proceden de textos castellanos. De este modo, me he atrevido a no quedarme en la condición escueta de esta voz vista como un latinismo que ocultaba (de ahí que lo haya calificado de criptoaragonesismo) que se había tomado, con toda probabilidad en el caladero del aragonés, donde se empleó en los siglos XIV y XV, preferentemente en los textos jurídicos, administrativos o históricos (Lleal 1995, Lleal 2016 [y bibliografía allí citada, a la que se deberían añadir algunos trabajos de su discípulo M. Raab, posteriores a su tesis doctoral]). Y ante esto, como un hecho secundario, no me parece ahora que sea casual la relación que se puede establecer con el cast. *servar*.

Para moverme por estos terrenos ejemplificaré con *servar*, que era una más entre las voces de Santillana que he estudiado, la necesidad de partir del conocimiento de la situación dialectal de los textos del CDH en que aparecían estas palabras. Para ello tuve que examinar someramente los documentos en que las

---

<sup>1</sup> En el apartado 3.5., tomo como 3.5.1. los 5 ejemplos de los textos de poesía, en que distintos cancioneros repiten el mismo ejemplo de J. de Dueñas (3.5.1.); igualmente lo hago con los 3 de 3.5.3., en que ocurre lo mismo con un ejemplo de F. Pérez de Guzmán.

encontré; luego las distribuí en dos grupos: en uno, las proclives al orientalismo, y las que no lo eran, en otro. Establecí en los primeros casos una gradación que iba de un nivel alto de lo que podríamos llamar masa dialectal oriental (lo ejemplifico con los documentos 3.1.2., 3.1.6.<sup>2</sup>), a uno medio (lo ejemplifico con los documentos 3.1.4., 3.2.4.) o bajo (lo ejemplifico con los documentos 3.1.1., 3.1.5. y gran parte de 3.1.14.).

La cantidad de tiempo que dediqué a esta tarea hizo que me preguntara si la llamada inteligencia artificial no podría ayudarnos a los filólogos a marchar por estos caminos con una cierta comodidad y hasta contribuir a responder a algunas preguntas sobre la historia de las palabras que ni siquiera nos hemos atrevido a formular. Sé muy poco de las posibilidades que nos brinda esa nueva disciplina, pero, por lo que intuyo, ganaríamos mucho si se aplicara a la comprensión de la historia del léxico de una lengua.

Para conseguirlo habría que contar con esa caracterización lingüística previa de los textos a que acabo de referirme, que acabo de hacer a mano. Supongo que las máquinas, aunque lo hagan a su manera, podrían llegar a someterlos a una marcación diatópica y que la forma de realizarlo consistiría en hacer pasarlos por filtros construidos por la combinación de una selección de rasgos gráficos, morfológicos, sintácticos y léxicos; aparte de aprovechar las informaciones sobre la procedencia del escritor o copista o la persona a la que se dirige (situación de los documentos que se contienen en el ítem 3.1.5. de este artículo). Por otro lado, cada paso que se diera en esta dirección conseguiría que se fuera refinando paulatinamente el propio filtro del que se partiera, que es lo que ha ocurrido con esos textos que presentaré en el apartado 3.5., que no había previsto en un principio que fueran proclives al orientalismo.

¿Podremos liberarnos de un trabajo manual, para el que Juan Antonio Frago (1977: 314) —en lo que yo no disenta— creía que estábamos condenados los filólogos a recurrir a «la paciente elaboración de monografías y el indefectible despojo de los fondos documentales»?

## 2. La caracterización diatópica de los textos

El comportamiento de la voz *servar*, que es la que he seleccionado de las estudiadas en el artículo citado sobre el marqués de Santillana (Pascual 2023b) con el

---

<sup>2</sup> En el apartado 4 se explica a qué corresponde este tipo de referencias.

fin de explicar por medio de ella la necesidad de situar diatópicamente los textos, encajaba bien en los esquemas que se podían dibujar de otras voces del *CDH*, que aparecían de un modo preferente en textos orientales u orientalizantes de los siglos XIV y XV, y solo en contados casos en textos castellanos del XV (esto último se podría deber a que la voz se hubiera institucionalizado finalmente en algunos de los niveles de nuestra lengua o que hubieran llegado a esos textos por otros caminos).

*Servar* en concreto aparece en el *CDH* en medio centenar de textos situados en una horquilla cronológica que va de principios del siglo XIV a los últimos del s. XV, algunos de ellos, como es el caso de 3.1.5. y 3.1.11., llegan a los primeros del XVI, por lo laxo de su fechación en el corpus y mi tardanza en acceder a los libros de los que se tomaron estos documentos. De los 280 registros del lema *servar* hube de dejar de lado formas como *servo* ‘siervo’, *servir*, *Serván*, *Servando*, *serba*, etc., así como unas cuantas latinas, del tipo *servando modo*, tanto en documentos latinos como romances. Prescindi también de los ejemplos de Santillana (§ 142, § 143, § 149, § 150, § 151, § 126, § 127<sup>3</sup>), y no porque aparecieran repetidos en unos cuantos cancioneros —normalmente anónimos—, sino porque al comparar en mi trabajo sus usos en la *Comedieta de Ponça* con los del resto de los textos, eran solo estos los que tenía que tomar en consideración. Pasados los documentos por este primer filtro, los organicé en esos dos grupos a los que me he referido antes.

Algunos argumentos externos (como el título de las obras), junto a otros internos (gráficos, fonéticos, morfológicos, sintácticos y particularmente léxicos) me permitieron considerar marcadas como orientalizantes una serie de obras, que había que situar en una gradación que diera cuenta de la que podíamos llamar su masa dialectal, distinguiendo las que podían considerarse castellanas aragonizadas de aquellas otras aragonesas o que mantenían un buen número de rasgos aragoneses. Junto a ellas coloqué las que siendo castellanas delataban en alguna etapa de su transmisión la intervención de un traductor o copista aragonés y podían contener, por tanto, algunos restos orientales (Pascual 2016: 59<sup>4</sup>). Hube de añ-

<sup>3</sup> Vid. nota 1.

<sup>4</sup> He ejemplificado esta condición (Pascual 2012: 220 y Pascual 2016: 59) con el códice escorialense que contiene la traducción del *Decamerón* cuatrocentista, del Escorial, o con la traducción del *De proprietatibus*, de Bartolomé Ánglico. Me atrevería a añadir las *Ystorias de Roma de Paulo Eurosio* (que se conserva en el Fitzwilliam Museum de Cambridge, McClean, 180), de Alfonso González de Zamora en su traducción al castellano de 1439, hecha por encargo del marqués de Santillana. Se mantienen en ella restos procedentes de la versión aragonesa de la *Historia romana*, patrocinada por Fernández de Heredia, que se tuvo en cuenta en esta traducción, como es el caso de *gridar*, al que pudo recurrir el traductor dejándose llevar por el italiano del texto de que partía, junto a otras voces que no necesitaban de tal impulso, como *el señal* o construcciones como *alegre-*

dir finalmente algunas obras que estaban en una situación textual con la que no había contado en principio: la de aquellas que, sin poderlas situar en este ámbito dialectal, mostraban una apertura al aragonés, que las acercaba a las anteriores, como ocurría con las poesías de muchos poetas del siglo XV (que he colocado en el apartado 3.5.) o de Arragel (en el apartado 3.4.). El resto de los textos, que he considerado no marcados (claramente castellanos), aparecen en el apartado 3.6.

El tiempo que a causa de esto hube de dedicar a hurgar en unas cuantas obras me ha convencido de la ventaja que supondría que cada filólogo no se viera obligado a marcar por su cuenta diatópicamente los materiales de que dispone en un corpus de gran tamaño (no digamos si ha de dotarlo de todas las demás marcas que se requirieren para poderlos someter a comparación) y se pudiera conseguir, en cambio, llegar a caracterizar estos materiales de una manera automática.

### 3. Medio centenar de textos caracterizados

Entro ya en la caracterización que he conseguido hacer en un tiempo razonable (si dos meses y medio lo son) de los documentos del *CDH* en los que aparecían *servar* y las otras 21 palabras de Santillana de las que me estaba ocupando. Conté con los fondos de la biblioteca de la Real Academia Española para los casos en que, al no ser muy experto para manejar por la red, no llegaba a acceder a determinados libros por internet. Este trabajo me ha llevado a organizar en cinco grupos textuales aquellos que podríamos considerar diatópicamente marcados: *Documentos administrativos orientalizantes*, *Obras del escriptorio de Fernández de Heredia*, *Textos de finales del S. XIV y del S. XV traducidos del catalán o escritos por escritores catalanes o aragoneses*, *Arragel*, *Textos de poetas castellanos del XV*. Y, frente a ellos, en el grupo sexto, los castellanos sin más.

En cada uno de esos grupos he numerado los textos con dos dígitos, separados por un punto (el primero se refiere al apartado en que aparece el documento y el segundo, el número que lleva en el lema *servar*, tal y como aparecía en el *CDH* cuando lo consulté, buscando por lemas, en los meses de abril, mayo y parte de junio de 2023). A lo que añado la secuencia de la numeración de los ejemplos que están en ese documento (en negritas y precedidos del signo §). Le sigue la fecha del texto en que aparece la voz en el corpus, salvo en algunos casos en que

---

*mente e dulce o no se glorificaron punto* (Cacho Blecua 2002–2004: 1896, 1987). Aquí mismo, en el apartado 3.3.6. atribuyo esta misma condición a la *Teseida*.



la bibliografía me ha convencido de la pertinencia de cambiarla. Continúo con el título que ese documento o colección documental tiene en el corpus, a veces con algún pequeño cambio basándome también en la bibliografía. Añado después entre corchetes la referencia bibliográfica de dónde se ha tomado el texto. Y, finalmente, en los cinco primeros grupos me sirvo de las palabras de un especialista como prueba de la caracterización dialectal que he aceptado del texto, y/o lo explico a menudo añadiendo algunos ejemplos que he recogido en él (los organizo empezando por los hechos gráfico-fonéticos, siguiendo por los morfológicos y sintácticos y terminando por los léxicos).

### 3.1. Documentos administrativos orientalizantes

En bastantes casos el propio título muestra la condición dialectal aragonesa o navarra de un texto (si bien a veces lo confirmo por medio de algún comentario o ejemplo). En este primer grupo de textos contamos con 73 casos de *servar* contenidos en 14 documentos o repertorios documentales<sup>5</sup>, pertenecientes al ámbito jurídico (p. ej. el doc. 3.1.9.) o administrativo (p. ej. el doc. 3.1.11.).

La mayor parte de estos textos, que se escanearon (y se corrigieron después) o transcribieron en la Academia, se volcaron en el *CORDE*. Pertenecían a la nómina de los que se estaban empleando en la redacción del *DHLE*, tras pasar al fichero académico, seleccionados en bastantes casos, por medio de un subrayado a lápiz en los libros.

3.1.1. § 9. «Embajada del rey de Aragón al de Portugal acerca de los asuntos de Castilla». 1313 [Giménez Soler, *Don Juan Manuel, biografía y estudio crítico*. Zaragoza: Tipogr. La Académica, 1932 (se trata del documento 259, que ocupa las pp. 420–421)].

Este texto forma parte de una serie documental de escaso tinte aragonés; pero no es, como otros de la colección, ajeno a Aragón: está dirigido al rey don Jaime II, suegro de don Juan Manuel, quien, por otro lado, «tanto [tenía] que ver con ciudades muy importantes [...] como son la propia Murcia, Alicante, Lorca, Orihuela, Cartagena y tantos otros territorios de frontera de los reinos de Murcia, Valencia y Granada, [que] conocía muy bien» (Díez de Revenga 1997: 111).

---

<sup>5</sup> Tomo como documentos también colecciones documentales de los siglos XIV y XV, aunque sean de distintas épocas, si los materiales que las integran proceden de una misma zona dialectal o están escritos por una misma persona o relacionados con ella; disminuye de este modo el peso de lo oriental, de forma que no forzaríamos con ello una estadística a favor del aragonés.

3.1.2. § 13. «Carta de hermandad entre Ansó y las poblaciones de Cillas, Navasal y Huértolo pertenecientes al monasterio pinatense». 1338 [A. I. Lapeña Paul, «Documentos en romance del monasterio de San Juan de la Peña (1325–1399)», *Alazet*, 11 (1999): 295–314, pp. 302–306].

Documento escrito por el notario público de la Val de Ansó, Guillelm Sánchez, quien mantiene muchos usos orientales: *concello*, *melloramientos* ~ *milloramientos*, *acullir*, *dreitos*, *paxer* ‘pacer’, *bienga* ‘venga’, *rienda* ‘rinda’, *dira* ‘diera’, *eligiran* ‘eligieran’ ~ *eslieran*, *vint* ‘20’, *judes*, *dita*; *entrant octubre*, *obidient*; *aquellos de qui somos procuradores* ~ *aquellos de quien somos procuradores*, *los lures* ‘sus’ *terminos*; *dius* ‘yuso’, *sinse* ‘sin’, *non paguen res*, *de aqui adebant*, *debante los juezes*, *ende podrá*, *entro a* ‘hasta’; *bovalar* ‘sitio donde pastan los bueyes’, *pujar* ~ *puyar* ‘subir’, *trobar* ‘encontrar’, *bedaler* [cat. *vedaler*] ‘guarda’, *fusta de maiestria*, *escaxar* (cf. cat. *catxar*), *yerra* ‘yerro’.

3.1.3. § 79, § 80, § 81, § 82, § 83, § 84, § 85, § 86, § 87, § 88, § 89. «Ordinación dada a la ciudad de Zaragoza por el rey don Juan I». 1391 [M. Mora Gaudó «Ordenanzas de la ciudad de Zaragoza, Zaragoza 15 de agosto de 1391». Zaragoza: Imprenta de Mariano Escar, 1908: 223–279].

En unas rápidas calas de la pp. 225 a la 228 encuentro: *dita*, *clamar*, *esleydo*, *dito* *yes*; *las casas comunas*; *sines* ‘sin’, *aprés*, *el jurado non hy aurá*; *gitados*, *siquier puestos en un vaxiello pleno de aygua*, *si clamado algunt passant por la carrera*, *discort en esleyr*.

3.1.4. § 90, § 91, § 92. «Ordinaciones y paramientos de la ciudad de Barbastro». 1396 [M. de Pano y Ruata, «Ordinaciones y paramientos de la ciudad Barbastro», *Revista de Aragón*, 3 (1902). (Se publican en numerosas entregas en los números 3 (1902): 4/1, 4/2 (1903) y 5 (1904), que ocupan normalmente de 2 a 6 páginas)].

Tras hacer varias calas en distintos documentos, extraigo unos cuantos ejemplos del primero de ellos: *concello*, *meyllor*, *dita*, *metge*, *sean parellados con las armas*, *priesten* ‘presten’; *los otros qui han armas*; *sienes*, *encara*; *senblantes* ‘semejantes’, *bregua o pelea*. En el tomo IV (1903) de la revista, en el documento que ocupa las páginas 212–214 encuentro: *los ueydores*, *vinyuegalos* [cf. cat. *vinyòvol*] o *uedaleros*, *siquier guardas*. El comienzo de otro documento de estas *Ordinaciones* (V/1 [1904], pp. 191–193: 191), merece citarse por extenso: «Muytas uegadas por dar auiso temerario et permissio que dentro la Ciutat las gentes vayan con armas

se ha seguido en los tiempos passados bregas feridas scandalos et encara muertes. Por esto queriendo prouedir et remediar en lo sobredito por euitar scandalos e males Statuyoron e ordenoron que los jurados de la dita Ciudat en cada un anyo en principio de su officio por voz de crida publica feyta por la dita Ciudat por los lugares acostumbrados vieden e prohibescan...».

3.1.5. § 175, § 176, § 177, § 178. «Adiciones a las Ordinaciones de Barbastro». c1450–1525 [M. de Pano y Ruata, «Ordinaciones ciudad Barbastro», *Revista de Aragón*, 5 (1904): 4].

Muestra el documento un colorido aragonés muy desvaído, que se reduce a casos como *feyto*; *el quizenno de Septiembre primero veniente, alguno de nos no discrepant ni contradicent; ensemble*. Es notable la distancia que mantiene en los rasgos dialectales con el documento anterior. Lo escribió en parte, y en parte lo mandó escribir, mossen Thomas Foncillas, público notario, el 13 de octubre de 1521. Si, a pesar de la fecha, no prescindí de este documento se debe a que, como en el caso del doc. 3.1.11., tardé tiempo en dar con la revista de donde se había tomado, de forma que para las calas que había hecho en él había tenido que acudir a los ejemplos del corpus por el procedimiento de ampliar en mi consulta varias veces la extensión de estos.

3.1.6. § 99. «Renovación de un contrato de Guillén de Leví para pintar un retablo en el convento de dominicos de Calatayud». 1396 [M. Serrano y Sanz, «Documentos relativos a la pintura en Aragón durante el siglo XV, [I]», *Revista de Archivos, Bibliotecas y Museos*, 32 (1915): 147–166, pp. 148–149].

Basta con ojarlo por encima para encontrar rasgos léxicos claramente aragoneses: *mi pertreyt, adueyto, a coneximiento, segunt dito yes, yes a saber; finas colores; attendient e considerant; tantost luego pora comprar fusta, dius ciertas formas, d'alli adelant, meter hi oro e colores muyt finas; posado e asentado en el altar, de buena fusta e sequa*.

3.1.7. § 110, § 111, § 113, § 114, § 115, § 116, § 117, § 118, § 119. *Ordinación dada a la ciudad de Zaragoza por el rey don Fernando I (el de Antequera)*, II. 1414 [M. Mora Gaudó. Zaragoza: Imprenta Mariano Escar, 1908: 311–479].

Documento muy castellanizado, si bien conserva rasgos aragoneses, como *feytas, consello, jutge; de suso ditos; concernescen, punezca*.

3.1.8. § 131, § 132, § 133, § 134. § 135, § 136, § 137, § 138, § 139. *Modificaciones y corroboraciones de las Ordenanzas de Zaragoza por el rey don Alonso V, con las confirmaciones del rey don Juan, II. 1429–c1458* [M. Mora Gaudó. Zaragoza: Imprenta Mariano Escar, 1908: 485–514].

Documento muy castellanizado, que conserva, no obstante, rasgos aragoneses, como *jutge, seyer; apres*.

3.1.9. § 164, § 165, § 166. «Concesiones establecidas por don Juan Mur [...], señor de la baronía de Aljafarín». 1446 [P. Longás, «Un documento sobre los mudéjares de Nuez (Zaragoza) Siglo XV», *Al-Andalus*, 281 (1963): 431–443, pp. 435–443].

Documento muy castellanizado, que conserva, no obstante, rasgos aragoneses, como *feytos, ito; diuso; sauieza*.

3.1.10. § 182. *Documentación medieval de la Corte del Justicia de Ganaderos de Zaragoza*. 1484 [J. A. Fernández Otal, *Documentación medieval de la Corte del Justicia de Ganaderos de Zaragoza [1472–1492] (y Ordinaciones de la Casa de Ganaderos del año 1511)*. Zaragoza: Institución Fernando el Católico, 1995].

Corresponde a la primera parte del título del libro de J. A. Fernández Otal, y el documento de 1484 aparece en la p. 81. Está muy castellanizado, pero, no obstante, se encuentran en él casos como *feyto, traye* o *apellidant*.

3.1.11. § 202, § 203, § 204, § 205, § 206. *Ordinaciones de la Casa de Ganaderos del año 1511*. 1511 [J. A. Fernández Otal, Id. que en 3.1.10.].

Corresponde a la segunda parte del título del libro de J. A. Fernández Otal: las *Ordinaciones* de 1511, que tiene aún menos rasgos aragoneses que el anterior (3.1.10.). Los ejemplos del corpus están en las pp. 142 (f. 294v), 143 (f. 296) [2 veces], 151 (f. 302), 155 (f. 306). Si, a pesar de la fecha, no prescindí de este documento, ello se debe a que, como en el caso del doc. 3.1.5., tardé tiempo en dar con la fecha en el libro de donde se había tomado.

3.1.12. § 179, § 180. «Retablo para el altar mayor de la iglesia de Santa María de Jesús, de Zaragoza». 1456 [M. Serrano y Sanz, «Documentos relativos a la pintura en Aragón durante el siglo XV», *Revista de Archivos, Bibliotecas y Museos*, 32 (1914): 147–166, pp. 155–159].

Muestra el documento numerosos rasgos aragoneses: *yes feyto, yes feyt, drecho, colliran, allyos* ‘ajos’, *fillos, seyer, clamados, bien enguixado*; *la present carta, los vasallos [...]* *qui son; drectament ni indirecta; deuant él, encara, el notario dius nonbrado; façer cridar por voz de corredor, un semblant retaulo, en semble, gueyto dias, devallamiento de la cruz, capción* ‘captura’, *mes de janero, sleyran, setze, mazonerías e obras.*

3.1.13. § 192. «Retablo para la cofradía de Todos los Santos, en la Seo de Zaragoza». 1487 [M. Serrano y Sanz, «Documentos relativos a la pintura en Aragón durante los siglos XIV y XV», *RABM*, 33 (1915): 41–48].

No he logrado dar con el documento, pero su título y otros documentos del mismo tipo me hacen pensar que estará castellanizado y conservará bastantes aragonesismos.

3.1.14. «Correspondencia de Fernando II sobre asuntos relacionados con las tareas de gobierno en sus reinos». [A. de la Torre, *Documentos sobre relaciones internacionales de los Reyes Católicos*. Barcelona: CSIC. 1949–1950]. Pertenecen al tomo I los ejemplos: § 184, § 185 1479; § 186 1479; § 187 1480; § 188 1481; § 189 1486; § 190 1486; § 191 1486; § 194 1488; § 197 1491; § 196 1491; § 198 1491; § 207 1493; § 211, § 212 1494. Pertenecen al tomo II los ejemplos: § 213, § 216, § 217, § 218 1495; § 219 1496; § 223 1498; § 224 1498.

Se agrupan en esta compilación documentos redactados en distintos momentos y por diferentes escribanos, con notables diferencias en cuanto a sus rasgos orientales. He encontrado, al ojearlos con toda rapidez, algunos rasgos aragoneses, como *guiaje* y *ensemble* en dos documentos del I tomo relacionados entre sí, el § 211 y el § 212, pero otros, como el § 191 y § 213 son una muestra de la escasez de este tipo de rasgos. Lo cual ocurre sobre todo en los del segundo tomo, por más que haya algunos elementos aragoneses, como *quan afectuosamente e estrecha* en § 218.

## 3.2. Obras del escritorio de Fernández de Heredia

El gran número de obras procedentes del escritorio de Fernández de Heredia desequilibra la proporción entre textos castellanos y aragoneses del *CDH*, tanto en relación con los aragoneses anteriores, como con los castellanos del s. XIV. Esto hubiera debido llevarme a hacer correcciones estadísticas, pero me parecieron innecesarias, dada la escasez de textos castellanos en que aparece el verbo.

Estas obras patrocinadas por Fernández de Heredia no se comportan lingüísticamente de una manera uniforme, pues no responde su escritura a la variedad aragonesa (e incluso catalana o castellana) propia del traductor o copista que realiza el trabajo, sino que surge de una compleja y variada génesis lingüística de cada uno de ellos (en que interviene la lengua de partida de quien realiza la traducción, la lengua o las lenguas que este conoce, su procedencia y formación). De ahí que los escritos heredianos se muevan en un arco que va de un extremo que ocupan aquellos que podemos considerar castellanos con rasgos aragoneses (*Las vidas semblantes*), pasando por los que se sitúan en el centro (*Secreto secretorum*), hasta llegar, en el otro extremo, a los que pertenecen a un aragonés más alejado del castellano (Álvarez Rodríguez 2009: XXX, XXXI); aparte de que a lo largo de una obra pueden cambiar los usos lingüísticos. No se puede simplificar, pues, la situación ni atribuyendo las peculiaridades de los distintos textos heredianos a la variación lingüística debida al espacio del que proceden sus distintos autores ni tampoco a que estén originados porque unos tengan más deudas que otros con el catalán.

La aparición en 1982, en microfichas de una serie de obras de Fernández de Heredia publicadas por el HSMS, supuso un cambio importante en la nómina de obras del *DHLE*. Luego, su edición electrónica en un CD Rom en 1992 (y reeditadas en 1995 y 1997), permitió incluirlas con relativa rapidez en el *CORDE* y en la ampliación de este por medio del corpus «nuclear»<sup>6</sup>. Esto supuso contar con un fondo importante de textos referentes al aragonés, en ediciones fiables (Octavio de Toledo & Rodríguez Molina 2017: 11, n. 8), revisadas o corregidas para su incorporación al corpus académico, en distintas universidades españolas (Cacho 2002: 696; Lagüens 2010: 337 y n. 10). Y no fue menos importante la ventaja que suponía para la explotación filológica del corpus contar, no con ediciones críticas, sino con una transcripción semipaleográfica de manuscritos—aunque fue necesario dejar al margen los problemas codicológicos que pudieran conllevar—.

---

<sup>6</sup> La precipitación con que se hubo de comenzar la redacción del que se empezó llamando *Nuevo diccionario histórico del español* originó la repetición de determinados textos de la misma procedencia, adaptados para el *CDH* por otras personas, como es el caso del texto 3.2.12., *Obra sacada de las crónicas de San Isidoro...* para el que hay dos versiones en la nómina del *CDH*, hechas en el MSMS, tomadas del mismo manuscrito: una de J. Cacho y otra de R. af Geijerstam y C. M. Wasick, titulada *Historia de los godos de San Isidoro*; lo que ocurre también con el texto 3.3.7., *Traducción del Tratado de la Phisonomía en breue summa contenida, de Mondino dei Luzzi*, inc. de Hurus de 1494, para la que contamos, junto a esta edición de M. N. Sánchez, HSMS 1987, con otra de M. T. Herrera & M. E. González de Fauve, HSMS, 1997, titulada Anónimo, *Tratado médico (Tratado de la patología general)*, tomada, en cambio, del ms. I51 BNE.

Así, en la Universidad de Zaragoza, J. M. Cacho ha corregido basándose en los manuscritos correspondientes, las siguientes ediciones electrónicas de J. Nitti & Ll. A. Kasten publicadas digitalmente por el HSMS: el *Rams de flors*; *Eutropio y Paulo diácono, Compendio de historia romana y longobarda* (de la que hubo una edición anterior a la de Nitti & Kasten, de P. P. Conerly, en su tesis doctoral de 1979; cf. Herráiz 2005–2006); la *Historia contra paganos*, de Paulo Orosio; el *De secreto secretorum*; las *Vidas semblantes de Plutarco, I y II* (tomando en este caso en consideración la edición que preparaba A. Álvarez Rodríguez, publicada en 2009). J. M. Cacho ha corregido también la edición que David Mackenzie & A. Lutrell hicieron en 1999 para el HSMS del *Libro de los fechos et conquistas del principado de la Morea* y ha adaptado al formato del CORDE la edición electrónica que R. af Geijerstam & C. M. Wasick del HSMS hicieron de la *Obra sacada de las crónicas de San Isidoro, de Don Lucas, Obispo de Tuy* publicada en 1988 (en la comparación de las lecturas de los dos primeros folios de esta obra en el repositorio del HSMS y en el CORDE no encuentro diferencias que me lleven a utilizar en este caso la palabra corrección).

El *Tucidides romanizado* lo editaron J. M. Cacho & M. Sanz basándose en el ms. 10 801 de la BNE. J. M. Cacho editó *Las Vidas semblantes de Plutarco, III* del códice 72 del fondo español de la BNF y la *Grant crónica de España, III* del ms. 10 134 de la BNE.

Es distinta la situación de otros dos libros: uno procede de la edición electrónica de 1997 del HSMS de la *Grant Crónica d’Espanya, I*, de R. af Geijerstam (en que hizo correcciones a la edición previa que había publicado en microficha en 1982); el otro, las *Gestas del rey don Jayme de Aragón*, tomado de la edición de R. Foulché-Delbosc de 1909, que aparece en 3.2.13., que supongo que se transcribió en la RAE, obra para la que hoy disponemos de una edición con más garantías de José Martínez Roy.

No sería necesario decir que los textos publicadas en el HSMS puestos a disposición de los filólogos en el CDH, no son solo los del escritorio de Fernández de Heredia; son bastantes los que están en los apartados 3.4. y 3.5.

Vayamos ya a este segundo grupo de textos.

3.2.1. § 68, § 69, § 70, § 71, § 72, § 73, § 74, § 75, § 76, § 77, § 78. La *Grant crónica d’Espanya, I*, ms. 10 133 BNE. 1385 [R. af Geijerstam HSMS, 1995].

«[E]n los libros I y II el lenguaje presenta numerosos rasgos aragoneses y catalanes, tanto en lo que se refiere a la grafía como en morfología, sintaxis y vocabulario [...], por otra parte, las soluciones fonéticas son puramente castellanas» (Geijerstam 1964: 129).

3.2.2. § 16. *Gran crónica de España, III*, ms.10 134 BNE. 1376–a1391 [J. M. Cacho, Universidad de Zaragoza, 2003].

«[E]l lenguaje ostenta rasgos claramente catalanes» (Geijerstam 1964: 43).

3.2.3. § 17, § 18, § 19, § 20, § 21, § 22, § 23, § 24. *Rams de flors (Libro de actoridades)*, ms. Z-I-2 RBE. 1385–1393 [J. M. Cacho, Universidad de Zaragoza, 2003].

La lengua del texto es aragonesa; en cuanto a la procedencia del autor o copista no son provenzales (Guardiola Alcover 1998, *apud* Cacho 2000: 251). Para el cambio de fecha y título me baso en Guardiola Alcover 1998.

3.2.4. § 25, § 26. *De secreto secretorum*, ms. Z.I.2. RBE. 1376–1396 [J. M. Cacho, Universidad de Zaragoza, 2003].

Es de las obras más alejadas del castellano, junto con el *Libro de los emperadores* (Álvarez Rodríguez 2009: LXXX). Con todo, está muy castellanizada. Procede de una traducción de un texto árabe al latín y de ahí al aragonés, quizá a través de un intermediario catalán.

3.2.5. § 27, § 28, § 29. Orosio, *Historia contra paganos. Valencia [ms. V] Pontificia (Patriarca), olim Corpus Christi*. 1376–1396 [J. M. Cacho, Universidad de Zaragoza, 2003].

La versión del manuscrito de Valencia está más elaborada que la anterior del ms. 10 200 de la BNE. Esta (de a1377) se basaba en la traducción al italiano de Bono Gianboni; la que volvió a encargar después Fernández de Heredia se tradujo directamente de latín, aunque se tuvo en cuenta la traducción anterior (Cacho Blecua 1999: 249). El texto que aparece en el *CDH*, del ms. V, se puede considerar, pues, la versión definitiva, en la que «revela una voluntad clara de aproximar el romance a [un] latín», que se adelanta al que a finales del XIV y XV aparece en varios textos aragoneses y castellanos, por más que no busque ser ciceroniano (Romero Cambrón 2005–2006: 62). Hoy contamos con una edición de este texto de M. Á. Romero Cambrón & I. J. García Pinilla (2008).

3.2.6. § 30. *Libro de los fechos et conquistas de principado de la Morea*, ms. 10 131 BNE. 1393 [J. M. Cacho, Universidad de Zaragoza, 2003].



«La lengua de la crónica de Morea es fundamentalmente castellana. Hay rasgos típicos aragoneses, pero faltan algunas características más marcadas de este dialecto» (Hódcroft 1963–64: 101) se refiere también a algunos catalanismos léxicos (p. 102). Para los galicismos originados por la traducción intermedia en francés, *vid.* J. Pomer (2016: 141, 142). Para la fecha, *vid.* Romero Cambrón (2016: 358).

3.2.7. § 31, § 32, § 33, § 34. *Eutropio y Paulo diácono, Compendio de historia romana y longobarda*. París, Arsenal, ms. 8324. 1377–1399 [J. M. Cacho, Universidad de Zaragoza, 2003].

Se trata de un aragonés muy castellanizado, en que hay una proporción insólita de catalanismos, posiblemente porque un catalán tradujera el texto del italiano al aragonés (Herraiz Pareja & Álvarez Rodríguez 2018: CVII; *cf.* Geijerstam 1989: 509). Adopto el título del libro de M. J. Herraiz Pareja & A. Álvarez Rodríguez (2018), más acorde con el hecho de que el texto contiene el *Breviarium ab urbe condita* de Eutropio, refundido y ampliado por Paulo Diácono en su *Historia romana*, así como un resumen en seis libros de la *Historia longobardorum*.

3.2.8. § 48, § 49, § 50, § 51 § 52. Plutarco, *Vidas semblantes, I*, códice 70 del fondo español de la BNF. 1379–1384 [J. M. Cacho, Universidad de Zaragoza, 2002].

Se trata de «un aragonés fuertemente castellanizado», cuyos «rasgos occidentales se aprecian también en el léxico, donde las formas *después, fallar, favlar* alternan con las orientales *aprés, trobar, parlar*» (Álvarez Rodríguez 2009: XXX, XXXI). Tomo la fecha de J. M. Cacho (2000b: 113) y adopto el título de la edición de Álvarez Rodríguez (2009).

3.2.9. § 45, § 46, § 47. Plutarco, *Vidas semblantes, II*, códice 71 del fondo español de la BNF. 1379–1384 [J. M. Cacho, Universidad de Zaragoza, 2002].

*Vid.* 3.2.8.

3.2.10. § 36, § 37, § 38, § 39, § 40, § 41, § 42, § 43. Plutarco, *Vidas semblantes, III*, Códice 72 del fondo español la BNF. 1379–1384 [J. M. Cacho, Universidad de Zaragoza, 2002].

*Vid.* 3.2.8.

3.2.11. § 58. *Tucídides*, ms. 10 801 BNE. c1384–1396 [J. M. Cacho & M. Sanz, Universidad de Zaragoza, 2003].

«un aragonés castellanizado» (Álvarez Rodríguez 2007: XXXIII).

3.2.12. § 59, § 60, § 61, § 62, § 63, § 64, § 65, § 67. *Obra sacada de las crónicas de San Isidoro, de Don Lucas, obispo de Tuy*. Stockholm, Kungliga, ms. D 1272a, de finales del s. XIV. 1385–1396 [J. M. Cacho, Universidad de Zaragoza, 2003].

Encuentro en los dos primeros folios: *aflaquexe, resplandexe; mas tost, diusmetido* ‘sometido’, *encara, sauiaient e piadosa; en ren ‘en nada’ non fallesca; sauieza, esguart*. Cito el siguiente ejemplo por extenso, pues es revelador del tipo de lengua del texto: «Aquest prologo he dauant posado porque en la primera fruent del uolumpne aprenquan los principes por claro negocio d’alto linatge gobernar».

3.2.13. § 93, § 94, § 95, § 96, § 97, § 98. *Gestas del rey don Jayme de Aragón*, en que se sigue un manuscrito de la BNE contemporáneo a la redacción del texto. a1396 [R. Foulché-Delbosc, *Gestas del rey don Jayme de Aragón*. Madrid: Bibliófilos Madrileños, 1909].

Se trata del «libro XVII de la Tercera parte de la compilación titulada *La grant Coronica de los Conquiridores*, hecha por inspiración de Johan Fernández de Heredia, Gran Maestre de la Orden de San Juan de Jerusalén. El señor Foulché dice en el Prefacio que el compilador aragonés, en el citado libro XVII, arregló a su manera el texto catalán de la Crónica de D. Jaime el Conquistador. El objeto de aquel al publicarlo ha sido reproducirlo con la mayor fidelidad posible y sin más variación que la de deshacer las abreviaturas y añadir entre corchetes las letras omitidas por el copista. El texto hallábase inédito hasta ahora, con excepción de sus diez y seis primeras páginas, que fueron publicadas por míster G. W. Humphrey en un artículo titulado ‘Aragonese texts’ que vio la luz en la *Revue Hispanique*, t. XVI, p. 244» (Puyol 1931: 1042).

Es un texto muy castellanizado, aunque hay rasgos aragoneses, como *fillo, filla, muller, veyer, aplego sus conpanyas; lures; e apres vinose ende a Taraçona, por res del mundo; decibidos, la sobida o puyada*.

### 3.3. Textos de finales del XIV y del XV, traducidos del catalán o debidos a autores catalanes o aragoneses

No justificaré aquellos en los que no vea necesario mostrar su apertura al orientalismo, como es el caso de marqués de Villena, bilingüe en castellano y catalán, la traducción del *Libro de las donas* o la *Crón. de Aragón de Fr. Fabricio Gauberte de Vagad*.

3.3.1. § 56. Traducción del *Libro de Palladio*, ms. 10 211 BNE. 1380–1385 [P. Sánchez-Prieto Borja, Universidad de Alcalá de Henares, 2004].

*Vid.* Lleal (2016) y Pascual (2020).

3.3.2. § 130. E. de Villena, *Traducción y glosas de la Eneida. Libros I–III*. 1427–1428 [P. M. Cátedra, Madrid: Turner Libros, 1994].

3.3.3. § 101. *Traducción del Compendio de la humana salud de Johannes de Ketham*, ms. I-51 BNE. 1494 [M. T. Herrera y M. E. González, HSMS 1997; hay una edición impresa, previa, de M. T. Herrera (1990)].

«[E]n el plano léxico son abundantes los vocablos aragoneses y aun catalanes» y se citan *albellón*, *abet*, *beçones* ‘mellizos’, *bletas* ‘acelgas’ *botons* ‘turmas’, *bresca* ‘panal’, *cadira* ‘silla’, *culturar* ‘cultivar’, *cuquos* ‘gusanos’, que justifica las sustituciones que hace el editor de Burgos, de 1495, frente a la de Zaragoza, de 1494 (Herrera 1990; *vid.* también Jiménez Ríos 1992: 97 ss.).

3.3.4. § 168, § 169, § 170, § 171, § 172. *Traducción del Libro de las donas de Francesc Eiximenis*. a1448 [G. Lozano López, HSMS, 1992].

3.3.5. § 173 *Cuatro oraciones a la República de Florencia*. 1450 [C. Parrilla, Madrid: Gredos, 1995].

En el libro citado de C. Parrilla encuentro *ornamiento* (p. 26); *la honor ~ el honor* (p. 31); *semblantes actos* (p. 32), *luengas de contar* (p. 29), *luengos tiempos* (p. 31); *de cosas se contenta onbre en tiempos de estrecha necesidad* (p. 30).

3.3.6. § 174. *Traducción de la Teseida de Boccaccio*, ms. 1537 BNE (de la segunda mitad del XVI). c1450 [M. Gerli, HSMS, 1991]. Hay una edición posterior de V. Campo y M. Rubio (1996), que transcribe el ms. 1537 BNE, pero que tienen en cuenta también el ms. 7553 BNE en lugares deturpados o problemáticos (p. 43).

Vayan por delante las siguientes palabras que me escribió A. Octavio de Toledo el 28 de noviembre de 2021: «el *descriptus* quinientista de la *Teseida delle nozze dell'Emilia* boccacciana, cuyo testimonio cuatrocentista es tan oriental como el *Decameron* del Escorial...», para darle la razón confirmándolas con los siguientes ejemplos de orientalismos, que tomo de las primeras ciento diez páginas de la edición aquí citada: *estol* ‘flota’ (p. 67), *estol o flota* (p. 73); *grande era allí la grita que las dueñas fazían y los marineros* (p. 75); *proa* (p. 77); *Teseo, que resguardaba e vido la flaca Gatelle* (p. 78); *non temían punto a los griegos* (p. 78), *a las dueñas non punto temían* (p. 78); *la tu honor* (p. 82), *les fuese fecha grande honor* (p. 86); *la reina alta e pujante* (p. 83), *las pujantes saetas* (p. 98); *las dueñas aten[e]nsas* (p. 91); *plañir* (p. 93); *palud* ‘laguna’ (p. 94), *remor* (pp. 97, 98); *el esguarde manso* (p. 109). En este texto era esperable que aparecieran bastantes italianismos, algunos de los cuales coincidían con una posibilidad aragonesa, como *nave nin leño, nuestros leños fueron luego a la ribera o nuestro sangre*, que P. Sánchez Prieto-Borja (1996: 282) interpreta como italianismos.

Merece una atención especial, por lo problemática, la forma *pugnir* con el significado de *punir* ‘castigar’: el *aborrecible crimen pugnir* (p. 67), confusión que el *DEt-Cat* (s. v. *pena*) registra en catalán y explica así: «Per confusió amb *punyir* ‘punxar, apunxalar’»; en ese diccionario se señala que aparece en el *Vidal Mayor* con el significado de ‘castigar’, lo que se confirma en el *CDH* donde, además del dato del *Vidal Mayor*, encontramos *puñir ~ punyr ~ pugnir* alternando en este sentido con *punir*, en el siglo XIV en Fernández de Heredia y en el XV fundamentalmente en textos aragoneses y en dos escritores proclives al orientalismo, como el Arcipreste de Talavera o R. Sánchez de Arévalo (que usa voces como *estol* y *conrear*). No se me oculta la posibilidad de que haya habido problemas en la representación gráfica de *punir* y *puñir* (y variantes interpretadas como palatalizadas) por parte de copistas o editores. En cualquier caso, a la variación *punir ~ puñir ~ pugnir* es la misma que se da entre *punar ~ puñar ~ pugnar*. Este comportamiento de dos derivados del latín, uno de *punire* y otro de *pugnare*, pudo deberse a una doble procedencia, popular y culta.

### 3.3.7. § 209, § 210 Traducción del Tratado de la Phisonomía en breue summa contenida, de Mondino dei Luzzi, inc. de Hurus. 1494 [M. N. Sánchez, HSMS 1987].

N. Sánchez González de Herrero & M. C. Vázquez de Benito (2009) escriben sobre este incunable: «Por lo que se refiere a la variedad lingüística, podríamos caracterizarla como castellana con un ligero tinte oriental» (p. 6), «la presencia [...] de voces del tipo *carmini*, *lora* o *meca*, que enlazan con el catalán [DCECH, s.vv. *carmín*, *loro II* y *mequetrefe*], parece sugerir una traducción oriental, o bien

la intervención de un copista en algún momento del proceso de transmisión del texto, que desconocemos cuál fue» (p. 7).

Añado algunos rasgos de este texto, que serían impensables sin contar con una capa aragonesa en su transmisión, como en el caso del género de *el señal*, y de *el costumbre* o de *aljez* ‘yeso’, *botones* ‘turmas’.

3.3.8. § 225 F. Gauberte de Vagad, *Corónica de Aragón*, inc. de Hurus. 1499 [J. C. Pino Jiménez, HSMS, 2002].

3.3.9. § 236, § 237, § 238. *Strategematon de Sexto Julio Frontino*, ms. 10 198 BNE. a1500 [E. Jiménez Ríos, HSMS, 1995].

Aparecen en la obra ejemplos como *collir* ~ *coger*; *defensar*, *desferra* ‘despojo ‘botín’, *nafra*; «los enemigos escampados e corrijentes las tindas donde por diligencia de algunos pocos muchos fuegos fazientes semeiante de grant muchadumbre allj stante seruando tomo los e los mato».

### 3.4. La particularidad de Arragel

Tenemos un único ejemplo en un solo texto.

3.4.1. § 121. Mosé Arragel de Guadalfajara, *Traducción y glosas de la Biblia de Alba, II*. Bibl. Casa ducal de Alba. c1422–1433 [A. Paz y Melia, Madrid: Imprenta Artística, 1920–1922].

En muchas ocasiones Arragel se comporta como un escritor aragonés.

### 3.5. Poesía

Muchos de los poetas castellanos del Cuatrocientos compartieron con el marqués de Santillana sus usos orientales. Fue una aventura en la que lo acompañaron Juan de Mena, Gómez Manrique y bastantes poetas más, entre ellos los tres que aparecen en este apartado por haber empleado *servar*. La poesía cortesana del reino de Castilla no se podría explicar sin contar con la Corona de Aragón, en ese «trasiego entre cortes, [que] fue norma común en el comportamiento del poeta cancioneril» (Casas Rigall 1995: 25), de forma que el «entorno de los Trastámara aragoneses, en particular la corte castellana y navarra del rey Juan y la napolitano-aragonesa del Magnánimo y de su hijo Ferrante» fue decisivo para la producción de los poetas castellanos (Beltrán 2001: 87), en una situación en que

«Don Enrique de Aragón, marqués de Villena, presidió varios certámenes de la gaya ciencia. Don Iñigo López de Mendoza, futuro marqués de Santillana y vinculado a Fernando I de Aragón y al Magnánimo, mantuvo contactos con Jordi de Sant Jordi, a quien recuerda en la *Carta-Prohemio* y en la *Coronación*, y quizás también con Pere de Torroella, Pedro de Urrea, Guillén de Bergueda» (Beceiro Pita 2016: 240).

Esta relación tuvo una consecuencia en la forma con que nos ha llegado la producción de aquellos poetas, en gran medida en unos cuantos cancioneros aragoneses. El *Cancionero de Íxar* refleja el trasiego de poetas entre las cortes de Castilla y Aragón (Alganza Roldán 2010: 33), igual que el *Cancionero de Stúñiga*, ligado, como el anterior, a la corte napolitana de Alfonso el Magnánimo (Beltrán 2011: 411, 427, 428; Alganza Roldán 2010: 34). En cuanto al *Cancionero de Palacio*, algunas de cuyas partes se han relacionado con uno de los círculos poéticos del Marqués (López Drusetta 2017b: 136) contiene la poesía producida en la corte castellana entre 1441–1444, pero también en el entorno del rey Alfonso V de Aragón y su hermano Juan I de Navarra (para la huella que el Oriente peninsular deja en el cancionero, *vid.* Tato 2003: 510–513); círculos literarios en que se movieron poetas aragoneses o castellanos que emigraron y se adaptaron a la realidad aragonesa (López Drusetta 2017: 277, 278). En estos y en otros cancioneros (relacionados entre sí, a los que se hará referencia un poco más abajo) aparecen poetas que tienen una fuerte relación con Aragón, como es el caso de Juan de Dueñas (López Drusetta: 2017b: 125) o de Pedro de Santa Fe, poeta al que M. R. Lida consideraba *medio catalán*, según escribe la filóloga a su maestro Amado Alonso, que tiene rasgos claramente aragoneses (Tato 2012: 313, 314 se refiere a su aragonesismo y a la utilización de voces catalanas, como *trasmuntana*, *serena* ‘brisa nocturna’, *penedir* ‘arrepentirse’). Incluso F. Pérez de Guzmán, más alejado de Santillana (Lapesa 1957: 275), muestra, además de *servar*, otros aragonesismos que estudio en el trabajo complementario a este (Pascual 2003).

3.5.1. § 109. Juan de Dueñas: «seruando iusto timón», en «La nao que fizo de Amor mosén Juan de Duennas», *Cancionero de Estúñiga*. c1407–1463 (1460–1463) [N. Salvador, Madrid: Alhambra, 1987: 236]. El ejemplo se repite en: § 144. *Cancionero de Salvá* (PN13), ms. Esp. 510 BNF. c1430–a1480 [F. Maguire, HSMS 1995], § 141. *Cancionero castellano y catalán de París* (PN4), ms. Esp. 226 BNF. a1430–a1494 [R. G. Black, HSMS, 1995], § 146. *Cancionero castellano de París* (PN5), ms. Esp. 227. 1430–1470 [F. Maguire HSMS, 1995], y § 148. *Cancionero castellano de París* (PN12), ms. Esp. 313 BNF. 1434–c1470 [J. Coca, HSMS, 1995].

3.5.2. § 120. Santa Fe: «Allí non servastes ley / de senyor...», en «Poesías», *Cancionero de Palacio*. a1419–a1440 (c1439) [A. Pellitero, Salamanca: Junta de Castilla y León, 1993].

3.5.3. § 125. F. Pérez de Guzmán: «en tu coraçon seruando / quanto dulçor fue gustado», en «Ymno a los Go[ç]os de Nuestra Señora», *Cancionero de Juan Fernández de Íxar*, ms. 2882 BNE. 1470 [J. M. Azáceta, Madrid: CSIC, 1956]. El ejemplo se repite en § 145. *Cancionero castellano de París* (PN5), ms. Esp. 227 BNF. 1430–1470 [F. Maguire, HSMS, 1995] y en § 147. *Cancionero castellano de París* (PN9), ms. Esp. 231 BNF. c1430–1460 [F. Maguire, HSMS, 1995].

3.5.4. § 122. Varios ejemplos en prosa en el *Cancionero de Juan Fernández de Íxar*, ms. 2882 BNE (en la traducción del italiano de la *Flor de virtudes e de costumbres*, atribuido a Tommaso Gozzadini). 1470 [J. M. Azáceta, *op. cit.*]: «San Geronimo dize: fuerte cosa es en la riqueza seruar castidad» (p. 741), al que le siguen los ejemplos siguientes: § 123: «el horden que se deue seruar en las otras cosas» (p. 748), y § 124: «non sabe domar nin seruar a non a[v]er contraste con otri» (p. 750).

### 3.6. Documentos castellanos

Una ojeada poco cuidadosa y muy apresurada no me ha llevado a considerar que estos documentos estén abiertos al aragonesismo.

3.6.1. § 140. P. de Corral, *Crónica del rey don Rodrigo, postrimero rey de los godos (Crónica sarracina)*. c1430 [J. D. Fogelquist, Madrid: Castalia, 2001].

3.6.2. § 153, § 154. «Carta de juramento» (en Documentos del Archivo Municipal de Guadalajara. a1200–a1492 [No he logrado saber la procedencia real del documento, que quizá proceda de la tesis doctoral de J. M. López Villalba, *Diplomática municipal medieval de Guadalajara* [Recurso electrónico]. Guadalajara: Diputación de Guadalajara, 2006].

3.6.3. § 158, § 159, § 160, § 161, § 162. Conde de Haro (Pedro Fernández de Velasco), *El Seguro de Tordesillas*. 1439 [N. F. Marino, Valladolid: Secretariado de Publicaciones de la Universidad de Valladolid, 1992].

Se trata de un documento jurídico, con usos extrañísimos, como *fascer*, *fisciére*, en el que en unas catas que he hecho, me he topado con un *guiaje*, que podría ser un catalanismo.

3.6.4. § 208. *Exemplario contra los engaños y peligros del mundo*, ms. I/1.194 BNE. 1493 [F. Gago Jover, HSMS, 1989].

## 4. Conclusión

Llegado aquí, le habrá llamado la atención a un lector ajeno a este tipo de trabajos el esfuerzo que supone solo dar unos pocos pasos como estos para poder describir la historia del léxico español siguiendo la pista a unas cuantas palabras que han llegado al castellano desde los romances hispánicos que convivían con él. Si lo que he escrito muestra de algún modo ese esfuerzo, no es con el fin de hacer autobiografía mostrando algunos de mis apuros, sino para llamar la atención de que no es deseable ni previsible que podamos seguir trabajando así en el futuro. Lejos quedan aquellos tiempos en que Gaston Gross llamó mi atención sobre la importancia de la técnica informática en el trabajo filológico. En este momento en que intuyo la revolución que se nos viene encima, no he olvidado sus lecciones, dadas en sus trabajos, en que sabía perder todo el tiempo necesario para preparar los rieles por los que debían discurrir estos. Lo mismo que nos cumple lograr a nosotros para que cuando llegue ese gran cambio tecnológico que nos espera sepamos preparar el camino. La utopía —razonable, no obstante— no consistiría ya en disponer de una adecuada lematización del corpus, sino de una caracterización lingüística de los textos que aparecieran en él. Lo cual obviamente habría de permitir quedar abierto a una permanente corrección.

Quiere esto decir que no estaría de más ayudar al futuro, por medio de una cuidadosa planificación de todo aquello que la filología ha de proponer a las máquinas. De lograrlo, un corpus sería no solo imprescindible para la lexicografía histórica, sino que podría convertirse en el recurso en que se centralizara el trabajo filológico. Aparte de que podría quedar como un resto de pasado el prestigio del *odium philologicum* al que se refiere G. Steiner:

El hábito de dedicar las horas que está uno despierto al cotejo de un manuscrito, a la recensión de marcas de agua en dibujos antiguos, la disciplina de investigar los sueños propios en la siempre vulnerable dilucidación de abstrusos problemas accesibles solo a unos cuantos colegas entrometidos y rivales puede segregar un singular veneno en el espíritu. El *odium philologicum* es una enfermedad bien conocida. Los estudiosos arremeterán unos contra otros con desenfrenada malevolencia sobre cosas que a los profanos les parecen puntos en debate minúsculos, a menudo risibles.

(Steiner 2009)



## Referencias bibliográficas

- Alganza Roldán, M. (2010). Elegíaca: notas de la tradición clásica a dos defunciones del Cancionero de Íxar. *Florentia Iliberritana*, 21, 3–64.
- Álvarez Rodríguez, A. (2007). *Tucídides, Discursos de la guerra del Peloponeso. Versión aragonesa de la Historia de la guerra del Peloponeso, patrocinada por Juan Fernández de Heredia*. Prensas de la Universidad de Zaragoza.
- Álvarez Rodríguez, A. (2009). *Vidas semblantes I-II: Versión aragonesa de las Vidas Paralelas de Plutarco, patrocinada por Fernández de Heredia*. Prensas de la Universidad de Zaragoza.
- Beltrán, V. (2001). El aprendizaje de una antología. Un estado de la cuestión para la poesía de cancionero. In C. Parrilla, P. Botta, & J. I. Pérez Pascual (Eds.), *Canzioneri iberici I* (pp. 77–104). Toxosoutos.
- Beltrán, V. (2011). Morfología del cancionero: Los cancioneros castellanos. In L. Leonardi (Coord.), *La tradizione della lirica nel Medioevo romanzo: problemi di filologia formale* (pp. 409–437). Atti del Convegno Internazionale, Firenze-Siena, 12–14 noviembre 2009. Edizioni del Galluzzo per la Fondazione Ezio Franceschini.
- Beceiro Pita, I. (2016). El contacto con el libro de la nobleza peninsular bajomedieval. *Discurso, memoria y representación en la nobleza peninsular en la Baja Edad Media* (pp. 213–242). Gobierno de Navarra.
- BNE: Biblioteca Nacional de España.
- BNF: Bibliothèque Nationale de France.
- Cacho Blecua, J. M. (1999). Las traducciones aragonesas de Orosio patrocinadas por Fernández de Heredia, un folio recuperado. In *Aragón en la Edad Media. Homenaje a la profesora Carmen Orcástegui Gros, I* (pp. 243–262). Universidad de Zaragoza.
- Cacho Blecua, J. M. (2000). La reconstrucción del *Rams de flores* de Juan Fernández de Heredia: A propósito de una nueva edición. *AFA*, 59, 247–264.
- Cacho Blecua, J. M. (2000b). La lengua de Fernández de Heredia: Una carta del castellán d'Amposta. *Aragón en la Edad Media*, 16, 111–128.
- Cacho Blecua, J. M. (2002). § 96 Juan Fernández de Heredia. In Alvar & Lucía (Eds.), *Diccionario filológico de literatura medieval española: textos y transmisión* (pp. 696–717). Castalia.
- Cacho Blecua, J. M. (2002–2004). La traducción de la *Historia romana de Paulo Diácono* (Cambridge, Fitzwilliam Museum, Mc Clean, 180). *AFA*, 59–60, 1881–1899.
- Campo, V. & Rubio, M. (1996). Giovanni Boccaccio, *La Teseida: Traducción castellana del siglo XV*. Vervuert–Iberoamericana.
- CDH: *Corpus del Diccionario histórico de la lengua española*. <https://apps.rae.es/CNDHE/view/inicioExterno.view;jsessionid=9664BA6FF93E5E88C504D72AC5FAEE2B>.

- CORDE: *Corpus diacrónico del español*. <https://corpus.rae.es/cordenet.html>.
- Casas Rigall, J. (1995). *Agudeza y retórica en la poesía amorosa de cancionero*. Universidad de Santiago de Compostela.
- DHLE: Real Academia Española, *Diccionario histórico de la lengua española*. <https://www.rae.es/dhle/>.
- Díez de Revenga, F. J. (1997). Don Juan Manuel: realidad textual y originalidad creadora. *Anales de la Universidad de Alicante. Historia Medieval*, 11, 111–120. <https://doi.org/10.14198/medieval.1996-1997.11.06>.
- Frago, J. A. (1977). Una perspectiva histórica sobre la relación entre el léxico navarroaragonés y el del área occitana. *RLR*, 41, 302–338.
- Geijerstam, R. af (1964). *Juan Fernández de Heredia: La Grant crónica de Espanya, libros I y II*. Edición según el ms. 10 133 de la Biblioteca Nacional de Madrid. Almqvist & Wisells.
- Geijerstam, R. af (1989). Juan Fernández de Heredia, transmissor de catalanisms lèxics a l'aragonés-castellà. In A. Ferrando (Ed.), *Segon Congrés Internacional de la Llengua Catalana, VIII* (pp. 499–511). Institut de Filologia Valenciana.
- Geijerstam, R. af & Wasick, C. M. (Eds.). (1988). *Lucas de Tuy, Obra sacada de las crónicas San Isidoro, arcebispo de Sevilla. [...] of Kungliga Biblioteket, Stockholm MS D 1272a*. HSHS (edición en microfichas).
- Guardiola Alcover, C. (1988). *Rams de flores o libro de actoridades [...]*. Edición del Manuscrito Z.I.2, de la Real Biblioteca de El Escorial. Institución Fernando el Católico.
- Herráiz Pareja, M. J. (2005–2006). Sobre la necesidad de una nueva edición de Eutropio. *AFA*, 61–62.
- Herráiz Pareja, M. J. & Álvarez Rodríguez, A. (2018). *Eutropio y Paulo diácono, Compendio de historia romana y longobarda*. Prensas de la Universidad de Zaragoza.
- Herrera, M. T. (1990). *Johannes de Ketham, Compendio de la humana salud*. Arco Libros.
- Hódcroft, F. (1963–64). Nota sobre la Crónica de Morea: Fonética, *AFA*, 14–15, 83–102. *HSMS: Seminary of Medieval Spanish Studies de la Universidad de Wisconsin-Madison*. <http://www.hispanicseminary.org/history-en.htm>.
- Jiménez Ríos, E. (1992). Historia del léxico a través de las variantes que aparecen en los textos médicos. In J. Villegas (Coord.), *Actas de XI Congreso de la Asociación Internacional de Hispanistas, I*, (pp. 196–113). Universidad de California.
- Lagüens Gracia, V. (2010). Regina af Gejerstam (1918–2010), *AFA*, 66, 331–339.
- Lapesa, R. (1957). *La obra literaria del marqués de Santillana*. Ínsula.
- Lleal, C. (1995). El secretario, el nuncio y la difusión del latinismo en el siglo XV. *Lletres Asturianas, Boletín oficial de l'Academia la Llingua Asturiana*, 56, 19–34.
- Lleal, C. (2016). Etimología e historia en el léxico de español. In M. Quiros *et al.* (Coords.), *Estudios ofrecidos a José Antonio Pascual* (pp. 873–884). Iberoamericana / Vervuert.

- López Drusetta, L. (2017). *Poetas del Cancionero de Palacio (SA7): Diego Hurtado de Mendoza, García de Pedraza y mosén Moncayo. Edición y estudio de su poesía* [Tesis de Ddoctorado Universidad de Coruña]. [https://ruc.udc.es/dspace/bitstream/handle/2183/19449/LopezDrusetta\\_Laura\\_TD\\_2017.pdf?sequence=2](https://ruc.udc.es/dspace/bitstream/handle/2183/19449/LopezDrusetta_Laura_TD_2017.pdf?sequence=2).
- López Drusetta, L. (2017b). El entorno poético de García de Pedraza en el *Cancionero de Palacio* (SA7). In V. Dumanoir (Ed.), *De lagrymas fasiendo tinta... Memorias, identidades y territorios cancioneriles* (pp. 125–137). Casa de Velázquez.
- Octavio de Toledo, J. & Rodríguez Molina, J. (2017). La imprescindible distinción entre texto y testimonio y los criterios de fiabilidad lingüística. *Scriptum digital: Revista de corpus diacrònics i edició digital en llengües iberoromàniques*, 6, 5–68.
- Pascual, J. A. (2012). Filología y lexicografía: La marcación diatópica en los corpus históricos. In *Dizionario e ricerca filologica. Atti della Giornata di studi in memoria di Valentina Pollidori*, Firenze, 26 ottobre 2010 (Supplemento III al *Bollettino dell'Opera del Vocabolario Italiano*) (pp. 199–228). Edizioni dell'Orso.
- Pascual, J. A. (2016). La filología en vago y en vilo entre los datos. In E. Blanco (Ed.), *Grandes y pequeños de la literatura medieval y renacentista* (pp. 55–84). SEMYR.
- Pascual, J. A. (2020). Notas léxicas sobre el aragonés: A propósito de *La Agricultura de Palladio al castellano*. *Revista de Investigación Lingüística*, 23, 1139–1146.
- Pascual, J. A. (2021). De los datos léxicos y de los textos que los contienen: A propósito del futuro próximo de la filología. *Neophilologica*, 33, 1–21.
- Pascual, J. A. (2023a). Una nota sobre la voz *pomelado* de la *Comedieta de Ponça* del marqués de Santillana. In C. Quijada van den Berghe, B. Alonso Pascua, F. Escudero Paniagua, C. Martín Gallego, & G. B. Garrido Vilchez (Coord.), *De Estepa a Salamanca miradas en torno a la lengua* (pp. 619–624). Publicaciones de la Universidad de Salamanca.
- Pascual, J. A. (2023b). Catalanismos y aragonesismos en la *Comedieta de Ponça* del marqués de Santillana. In *Actas del XII Congreso de la Asociación de Historia de la Lengua Española*.
- Pomer, J. (2016). Les fons grega e francesa de la versió aragonesa de la Crònica de Morea. In J. J. Pomer Monferrer & J. Redondo (Eds.), *La traducció (in)directa dels clàssics a les literatures hispàniques, entre més* (pp. 135–229). Adolf M. Hakkert – Publisher.
- Puyol, J. (1930). Biliografía de R. Foulché-Delbosc (1864–1929). *BRAH*, 97, 963–1125. *RBE: Real Biblioteca de El Escorial*.
- Romero Cambrón, M. Á. (2005–2006). Latinismos sintácticos en la lengua herediana: a propósito de las *Historias contra los paganos* (ms. V-27). *AFA*, 61, 57–84.
- Romero Cambrón, M. Á. (2016). La datación tardía de los facticios heredianos. *eHumanista/IVITRA*, 9, 355–367.

- Romero Cambrón, M. A. & García Pinilla, I. J. (2008). *Paulo Orosio, Historia contra paganos. Versión aragonesa patrocinada por Juan Fernández de Heredia*. Prensas de la Universidad de Zaragoza.
- Sánchez González de Herrero, N. & Vázquez de Benito, M. C. (2009). Tratado de fisonomía: Tratado de la generación de la criatura. In *Salamanca DLE. Artículos del Departamento de Lengua Española*. Salamanca.
- Sánchez-Prieto Borja, P. (1999). [Reseña a] Giovanni Boccaccio, *La Teseida (Traducción castellana del siglo XV)*: Edición, introducción y notas de Victoria Campo y Marcial Rubio Áquez. *Revista de literatura medieval*, 11, 280–283.
- Steiner, G. (2009). *George Steiner en The New Yorker* (R. Boyers, Ed.; M. Condor, Trad.). Siruela.
- Tato, C. (2003). El *Cancionero de Palacio* (SA7), ms. 2653 de la Biblioteca Universitaria de Salamanca (I). In J. L. Serrano Reyes & M. Alvar López (Eds.), *Cancioneros en Baena, I. Actas del II Congreso Internacional Cancionero de Baena. In memoriam Manuel Alvar* (pp. 495–524). Ayuntamiento de Baena.
- Tato, C. (2012). Prolegómenos a la edición del *Cancionero de Palacio* (SA7). In P. Lorenzo Gradín & S. Marcenaro (Eds.), *El texto medieval: de la edición a la interpretación* (pp. 299–318). Anuario de Verba.



*Magdalena Perz*

Université de Silésie à Katowice  
Pologne

 <https://orcid.org/0000-0003-1446-6920>

# Sur la synonymie dans le lexique adjectival – questions de sémantique

**On synonymy in the adjectives – object classes and semantic issues**

## **Abstract**

The purpose of the present article is to discuss, through adjectival uses, a number of difficulties that affect the synonymy relation. The article reviews various approaches to synonymy and contain a theoretical and applicative dimension. The notion of the object class (G. Gross, W. Banyś) is used to describe the semantics of adjectives in order to disambiguate them. The linguistic approach taken here to describe synonymous units is that synonymy is relation between words in use. A contextual approach to synonymy can overcome the shortcomings of classical dictionaries of synonyms.

## **Keywords**

Adjectives, synonymy, object-classes, co-text, polysemy

*La description des adjectivaux  
est une œuvre de longue haleine.  
(G. Gross, 1991 : 175)*

## **1. Introduction**

L'objectif de cet article est de discuter, par le biais des emplois adjectivaux, quelques aspects qui alimentent les discussions sur la notion de synonymie. Notre étude des faits de synonymie va comporter une dimension théorique et applicative. Dans un premier temps, nous allons faire quelques réflexions sur le concept de

la synonymie afin de rapprocher les divers aspects qu'elle convoque. Cette notion renvoie à une pratique linguistique largement reconnue qui demeure pourtant difficile à appréhender sous un angle sémantique. C'est ce statut équivoque de la synonymie qui nous a amenés à (re)discuter cette notion et à montrer son influence sur la description des unités lexicales. Nous illustrerons notre propos à partir des emplois adjectivaux – *fixe* et *stable* que nous avons retenus, d'une part, par leur polysémie significative, et d'autre part, en raison de leur synonymie approximative en langue. Enfin, nous rapprocherons quelques difficultés qui émergent de nos analyses.

## 2. Notion de synonymie

La synonymie, sous ses diverses formes<sup>1</sup>, fait partie de la pratique linguistique, très importante dans l'acquisition des compétences langagières. La reformulation, le choix du mot juste ou la recherche d'une unité de sens proche pour éviter la répétition demeure une pratique linguistique courante. En outre, la synonymie est largement utilisée comme méthode lexicographique, faisant partie de la définition des unités linguistiques. À titre d'illustration, citons le *Dictionnaire de l'Académie Française*, qui définit l'adjectif *stable* : *qui est assuré, ferme, durable, permanent*<sup>2</sup>. Tout cela prouve que la relation de synonymie est appréhendée de façon intuitive – les synonymes sont des unités qui appartiennent à la même catégorie grammaticale et qui ont un sens proche.

Rappelons que la synonymie est une notion très ancienne, remontant à la période antique. Les premières considérations sur la synonymie ont été discutées sous deux angles différents : l'un dit onomasiologique où l'on partait du sens donné afin de trouver les mots qui puissent l'exprimer et l'autre suivant lequel les mots constituaient le point de départ des analyses, nommé sémasiologique. Or, comme le constate G. Doualan (2014 : 410), les études de la synonymie sont arrivées à la maturité avec les écrits de l'abbé Girard. La synonymie demeure de nos jours un phénomène complexe et interprété de multiples façons.

---

<sup>1</sup> O. Duchaček (1964), dans son article « Différents types de synonymes », distingue deux catégories de synonymes : stylistiques (syntactico-praséologiques, expressifs, fonctionnels, spéciaux, caractérisés par la couleur locale) et sémantiques. Synonymie discursive vs lexicale discutée par D. Śliwa (2015).

<sup>2</sup> Dictionnaire de l'Académie Française, disponible en ligne : [dictionnaire-academie.fr](https://dictionnaire-academie.fr), consulté le 6.12.2023].

L'un des problèmes majeurs auxquels la relation de la synonymie se heurte, c'est la diversité de ses acceptions. Les définitions fournies varient d'un auteur à l'autre en introduisant différents postulats et critères en sémantique (critère sémantique, critère référentiel)<sup>3</sup>. Par conséquent, les discussions récentes sur la notion de synonymie ont fait émerger deux positions antagonistes – une interprétation restrictive, puisée dans la logique qui veut que les deux unités synonymiques affichent une relation logique d'équivalence, et une interprétation plus large selon laquelle la synonymie s'instaure entre certains fragments du sémantisme des unités. À un extrême, nous retrouvons la synonymie dite *totale* ou *absolue* qui postule que les unités soient substituables dans tous les contextes sans aucun changement dans leur portée cognitive et affective. Une telle vision de la synonymie a été contestée en linguistique par beaucoup de chercheurs comme étant exceptionnelle et n'ayant aucune utilité langagière<sup>4</sup>. À l'autre extrême, on trouve des chercheurs comme G. Kleiber ou B. Victorri et S. Ploux qui vont à contre-courant de la vision stricte de la synonymie et construisent leurs postulats en sémantique sur la notion du contexte. G. Kleiber précise qu'en synonymie il faut se placer :

au seul niveau de l'acception (ou du sémème) et non au niveau de toute la pluralité du sens. Et par conséquent aussi, la substitution ne doit être envisagée qu'au niveau de chaque acception et non pour l'ensemble des différents sens relevés.

(G. Kleiber, 2009 : 20)

Nous rejoignons sur ce point les observations de G. Kleiber. Cette acception de la synonymie requiert un contexte et la situe à un autre niveau – celui du discours. Notons pourtant qu'une telle prise de position conduit à admettre que les unités qui passent difficilement comme synonymiques dans l'usage courant, comme *stable* vs *immuable* ou *fragile* vs *succinct* sont synonymiques parce qu'on trouve pour ces couples l'environnement linguistique dans lequel ils peuvent commuter sans affecter le sens de l'énoncé. Une telle affirmation nous rend compte à quel point cette relation est relative et difficilement identifiable.

Tout linguiste voit dans la synonymie une équivalence plus ou moins large de sens entre deux unités lexicales différentes. Les chercheurs qui ont adopté une acception très stricte de la synonymie ont été amenés à reconsidérer la notion

---

<sup>3</sup> Pour plus de détails, cf. P. Cadiot (2009), B. Śmigielka (2007).

<sup>4</sup> (...) for absolute synonyms, and that is that if they exist at all, they are extremely uncommon (A. Cruse, 1986 : 270). Pour plus de détails, nous renvoyons aux travaux de M.-L. Honeste (2007), J. Picoche (1993), A. Polguère (2008), Lehmann F. & Martin-Berthet (1998).

elle-même. Elle a été réfutée en faveur de l'acception plus large dite *synonymie partielle*, *parasynonymie*, *quasi-synonymie*, *synonymie occasionnelle*, une optique qui sous-entend des degrés de synonymie. Une définition consensuelle de la synonymie est fournie, entre autres, par I. Tamba qui la définit ainsi :

Dans l'usage courant, on appelle *synonymes*, des mots de forme différente mais de sens identique ou semblable et de même statut morphosyntaxique.

(I. Tamba, 1988 : 80)

Postuler la relation de synonymie entre deux unités lexicales ne fait qu'affirmer qu'il existe un espace commun où leurs emplois se chevauchent. Il y a un environnement linguistique dans lequel les deux unités apparaissent pour exprimer le même contenu. Qui plus est, les dictionnaires encouragent ce type de démarche en fournissant les définitions à l'aide de listes de mots qui ont des sens voisins. Une acception plus stricte de la notion de synonymie exclura les cas qui vont être traités dans notre analyse.

### 3. Dictionnaires des synonymes

Ce qui a servi de tremplin pour nos discussions, c'est, entre autres, le cas des dictionnaires de synonymes tel que *le Dictionnaire Électronique des Synonymes (DES) du CRISCO*, *le Dictionnaire des synonymes et contraires Larousse*<sup>5</sup>. Ces dictionnaires rattachent les partenaires synonymiques aux différentes acceptions de l'unité en question sans donner des commentaires. Les listes de synonymes qui y sont dénombrés ne sont exploitables que par des usagers capables de s'imaginer les contextes dans lesquels les deux unités en question affichent une équivalence de sens. Ils ne permettent pas aux lecteurs d'accéder rapidement au synonyme précis. En outre, pour certains renvois censés être synonymiques, on a du mal à récupérer l'environnement garantissant la substituabilité. C'est le cas, par exemple du couple synonymique *stable* et *immuable*. Une telle approche n'est éclairante que dans la mesure où elle présente le potentiel sémantique de l'unité, mais elle se désintéresse de la délimitation des emplois. La question de la délimitation est de taille parce que, d'une part, elle est liée à la substituabilité des unités – le remplacement d'un élément lexical par son synonyme ne se fait que dans une

<sup>5</sup> <https://www.larousse.fr/dictionnaires/synonymes>, consulté le 06.11.2023.



situation spécifique et, d'autre part, sans procédures de vérification, le choix des synonymes reste voué à l'intuition<sup>6</sup>.

La connaissance des différences entre les emplois de deux unités est l'une des difficultés que rencontrent les usagers dans le domaine lexical. Les normes d'emploi propres à chaque langue conditionnent la sélection du mot. C'est une faculté cruciale du point de vue de la compétence lexicale. Penchons-nous maintenant sur la catégorie dont nous avons puisé les exemples pour notre analyse.

#### 4. Spécificités des adjectifs

Nous avons choisi les adjectifs *stable* et *fixe* – les adjectifs à emploi qualificatif, qui plus que les relationnels (majoritairement dénominaux), peuvent revêtir une multitude de sens différents<sup>7</sup>. Servant à exprimer diverses propriétés des substantifs qu'ils caractérisent, les adjectifs qualificatifs effectuent un procédé de restriction de l'extension sur le substantif. Cette opération de restriction peut s'effectuer de manières différentes. À titre d'exemple, prenons l'adjectif *stable*, qui entre en relation avec divers accompagnateurs : <humains> : *un homme stable* ; <concrets> : *chaise stable, échelle stable* ; <abstraites> : *sens stable, position stable* ; <temps> : *heure stable*.

En parcourant le sémantisme de ces groupes nominaux, nous constatons que les rapports conceptuels qui s'établissent entre adjectif et son nom-tête sont de nature diverse.

*une échelle stable = qui est en équilibre*

*un caractère stable = qui ne change pas*

*un emploi stable = qui est permanent*

*une personne stable = qui est constant(e) dans ses idées ; dont le caractère ne change pas*

*un environnement stable = qui est peu susceptible de transformations*

*un vent stable = qui est constant ; dont la direction est invariable*

*une température stable = qui n'est pas sujette à des variations*

*un pays stable = qui est équilibré sur le plan économique*

<sup>6</sup> Les usagers d'une langue ont une intuition naturelle résultant du savoir épilinguistique de pouvoir juger quels mots sont synonymes.

<sup>7</sup> Pour plus de détails sur le classement des adjectifs, cf. I. Bartning (1980), J. Goes (1999), Noailly (1999), Murphy (2003).

La multiplicité des relations sémantiques qui se construisent entre adjectif et nom dépend, premièrement, du type de substantif qualifié et du sémantisme de l'adjectif en question. Le sens de l'adjectif varie en fonction du nom en activant une ou plusieurs de ces caractéristiques. Les adjectifs qualificatifs sont susceptibles de plusieurs emplois et comme l'a justement remarqué M. Riegel (1985 : 97), ils ne possèdent pas d'autonomie référentielle. Ils se caractérisent par une « incomplétude » ou « dépendance » référentielle. R. Moon (1987 : 179), à son tour, précise que « (...) they are often heavily context-dependent and flexible, taking on as many meanings as you like or have space for ». Nous soulignons ici une des majeures particularités des adjectifs qualificatifs, à savoir leur distribution quasi infinie. Les types de substantifs avec lesquels un adjectif qualificatif peut entrer en combinaison sont de toutes sortes. Ainsi, étant donné sa remarquable souplesse sémantique, ce type d'adjectif accepte difficilement la hiérarchisation et résiste souvent à une définition cohérente.

Étant donné la polysémie des adjectifs qualificatifs, il faut traiter chaque adjectif qualificatif comme un faisceau de significations possibles. L'analyse de son sémantisme implique la prise en compte d'autres éléments linguistiques avec lesquels un adjectif entre en interaction. Cette prise de position nous conduit à admettre que la description des adjectifs et par conséquent la représentation des relations sémantiques doit être contextuelle. Elle se rapporte aux mots en emploi, ce qui souscrit à l'idée selon laquelle la synonymie est un phénomène propre au discours.

En cherchant à identifier les ressemblances entre les emplois de deux adjectifs, il est nécessaire d'indiquer et de regrouper les substantifs les régissant pour mettre en évidence leurs espaces communs. Soulignons toutefois que, les unités synonymiques auront toujours un fragment de signification qui leur est propre.

Pour mener à terme notre analyse, nous proposons d'adopter une procédure en deux étapes. Premièrement, une description qui permettra de répertorier les sens de l'adjectif en question, ce qui va permettre sa désambiguïsation et la structuration de ses acceptions. Deuxièmement, le repérage, pour chaque ensemble d'éléments retenus, son substitut (son synonyme) pour indiquer dans quelles conditions les deux adjectifs pris en couple affichent une relation d'équivalence. Ainsi, si les deux adjectifs sont substituables, ils peuvent être traités comme synonymes.

## 5. Classes d'objets

Pour décrire le contenu sémantique des adjectifs, nous recourons à des approches linguistiques dont G. Gross et W. Banyś ont donné un aperçu global. La description en termes de *classes d'objets* (G. Gross, 1994, 1998 ; G. Gross & A. Clas, 1997) ainsi que l'approche orientée-objets (W. Banyś, 2002, 2005) postulent que la description des unités lexicales devrait être co-textuelle, c'est à dire faite à travers leurs emplois<sup>8</sup>. Par l'emploi d'une unité lexicale, il faut comprendre les autres éléments linguistiques avec lesquels l'unité analysée entre en interaction. Ainsi, le sens du mot s'étudie dans sa relation avec d'autres unités en présence. L'indication de cet ensemble d'éléments, va ensuite permettre leur regroupement en ensembles homogènes de point de vue linguistique, nommées *classes d'objets*. Comme le souligne G. Gros (1994 :4), les classes d'objets sont « le lieu naturel de la description de la synonymie ».

## 6. Stable vs fixe

Nous avons choisi les adjectifs *fixe* et *stable*, qui sont des couples de synonymes approximatifs en discours à large spectre d'utilisation. Les dictionnaires confirment que les deux adjectifs constituent des synonymes selon les données fournies par le CRISCO (Centre de Recherches Inter-langues sur la Signification en Contexte). Nous ne proposons pas une analyse très approfondie de ces deux adjectifs. Nous tenons à fournir les éléments d'information nécessaires pour comprendre le fonctionnement sémantique de ces adjectifs. Pour parvenir à une description, il nous a fallu repérer, dans les limites de nos possibilités<sup>9</sup>, les éléments du co-texte. En simplifiant, on peut ranger les éléments de *stable* en classes suivantes :

---

<sup>8</sup> La première condition d'une description synonymique est de pouvoir délimiter correctement les emplois (G. Gross, 1994 : 4)

<sup>9</sup> Chaque analyse visant des ensembles lexicaux n'est pas exhaustive. Les classes d'objets sont toujours ouvertes à d'éventuels enrichissements.

## STABLE (adj)

stable <appareils> : appareil, caméra, machine, antenne, prothèse  
 stable <aspects> : caractère, nature, personnalité, vie  
 stable <collectivités> : bibliothèque, électorat, main d'œuvre, personnel, population, couple  
 stable <compositions chimiques> : composé, solution, corps, structure, suspension  
 stable <constructions> : construction, structure, installation, bâtiment  
 stable <conditions> : climat, décor, cadre, condition, environnement, lieu, milieu, paysage, temps  
 stable <coûts> : prix, revenu, financement, budget, dépenses, rendement  
 stable <dispositifs> : échelle, escabeau, escalier, marchepied, support, trépied  
 stable <événements> : phénomène, vent  
 stable <humains> être, quelqu'un, personne  
 stable <jugements> : classement, conception, position, situation, point de vue, opinion  
 stable <lieux> : terrain, territoire, lieu, endroit, surface  
 stable <meubles> : bibliothèque, chaise, étagère, fauteuil, table, tabouret  
 stable <moyens de transport> : bateau, navire, vélo, voiture  
 stable <occupations> : emploi, état, occupation, poste, travail  
 stable <rapports> : connexion, équilibre, union, lien, paix, rapport, relation, sens  
 stable <règles> : convention, doctrine, loi, norme, règle, proposition  
 stable <systèmes> : démocratie, économie, établissement, logiciel, pays, marché, famille  
 stable <valeurs> : chiffre, étendu, poids, température, pourcentage, proportion, valeur, niveau

On adopte la même structuration pour l'adjectif *fixe*.

## FIXE (adj)

fixe <appareils> : téléphone, appareil, barre, caméra, châssis, machine, antenne, prothèse, roue, pédale  
 fixe <astres> : astre, étoile  
 fixe <coûts> : capital, frais, impôt, loyer, montant, prix, salaire, somme, rémunération, revenu, tarif  
 fixe <collectivités> : population, collection, ensemble  
 fixe <directions> : sens, direction, trajet  
 fixe <habitations> : adresse, domicile, emplacement, lieu  
 fixe <occupations> : emploi, état, occupation, poste, travail  
 fixe <ouvrages> : figure, statue, sculpture  
 fixe <rapports> : sens, valeur  
 fixe <règles> : décision, direction, formule, norme, loi, procédure, règle, résolution  
 fixe <regard> : œil, regard, vue, yeux  
 fixe <temps> : calendrier, date, fête, jour, horaire, heure, terme  
 fixe <valeurs> : température, valeur, proportion  
 accent fixe  
 douleur fixe  
 feu fixe  
 idée, pensée fixe  
 point fixe  
 sourire fixe

L'analyse des éléments lexicaux de deux adjectifs nous conduit au cœur de leurs ressemblances et leurs différences d'emploi. Le sens de l'adjectif *stable* se superpose partiellement sur le domaine collocatif de son synonyme *fixe*. Pour s'en convaincre, il suffit de considérer les éléments de la classe <coûts> ou celle <collectivités> qui constituent les accompagnateurs récurrents de deux adjectifs.

Il convient de faire quelques commentaires sur la procédure d'établissement des classes sémantiques. Force est de constater qu'en recourant à la notion de classes d'objets, on est en mesure de structurer davantage les acceptions retenues pour l'adjectif analysé.

Dès qu'on procède à la structuration des unités polysémiques, on s'aperçoit que la démarche adoptée fait émerger quelques difficultés. Conformément aux principes, les classes d'objets reposent sur les critères linguistiques, mais il semble difficile de proposer une distribution catégorique. Par conséquent, les classes d'objets sont des structures hétérogènes – elles sont, d'un côté systématiques et, de l'autre côté, négociables. À titre d'exemple, citons le lexème *sens*, qui a été catégorisé comme <rapports> parce que si nous parlons du *sens stable*, nous nous référons à un certain type de relation qui n'est pas sujette à des changements. Mais, rien n'empêche de ranger ce lexème dans la classe <valeurs>. Par conséquent, parmi les différentes classes retenues, il est possible d'opérer les regroupements en fonction des propriétés linguistiques qui motivent leur fondement.

La difficulté est encore accrue par le fait qu'il existe un nombre considérable de cas intermédiaires, qui n'autorisent pas d'attribuer un substantif à une classe particulière. Cela est dû, entre autres, à la polysémie des substantifs. Pour s'en convaincre, citons le cas du nom *pays* qui, accompagné de l'adjectif *vert* dénote un territoire et ainsi fait partie de la classe <lieux>, accompagné de l'adjectif *riche*, *avancé*, il est associé à la classe <collectivités> et il peut référer à un système économique et par conséquent faire partie de la classe nommée <systèmes>, comme dans la séquence *pays stable*.

Or, une description de ce genre peut avoir un grand intérêt pratique parce qu'elle nous donne l'accès à la structure conceptuelle des substantifs. Un substantif aura autant d'emplois qu'il existe de classes dont il fait partie. La difficulté pour le linguiste est de réussir à s'abstenir de la représentation conceptuelle du mode qu'il a à sa disposition et adopter une démarche linguistique.

Bien que les ensembles proposés soient linguistiquement pertinents, mais, il faut le dire, leur existence est souvent motivée par le découpage conceptuel. En ce qui concerne les adjectifs, ils servent à exprimer une propriété concrète ou abstraite du substantif. Cette propriété est souvent attribuée par les locuteurs en fonction du statut ontologique ou conceptuel de l'unité en question. Déjà, par rapport à la situation extralinguistique, qui est souvent arbitraire certaines suites

seront qualifiées différemment. Ainsi, Noël est qualifié en tant qu'une fête *fixe, stable* et Pâques est une fête *mobile*.

Soulignons que la distinction des emplois différents des unités ainsi que l'affectation à des classes sémantiques exige une grande précaution de la part du lexicographe. Cela dépend, entre autres de la complexité de l'adjectif en question, de la mesure dans laquelle il dépend de son support nominal pour sa signification.

Pour mener à bien cette analyse et préciser les environnements dans lesquels les deux adjectifs peuvent commuter il est nécessaire de proposer pour chaque classe d'objets retenue son partenaire synonymique.

#### STABLE (adj)

stable <appareils> : appareil, caméra, machine, antenne, prothèse – SYN : **fixe**

stable <aspects> : caractère, nature, personnalité, vie – SYN : **équilibré**

stable <collectivités> : bibliothèque, électorat, main d'œuvre, personnel, population, couple – SYN : **fixe ; invariable**

stable <compositions chimiques> : composé, composition, corps, solution, structure, suspension – SYN : **durable**

stable <constructions> : construction, structure, installation, bâtiment – SYN : **solide ; résistant ; inébranlable**

stable <conditions> : climat, décor, cadre, condition, environnement, lieu, milieu, paysage, temps – SYN : **invariable ; permanent**

stable <coûts> : prix, revenu, financement, budget, dépenses, rendement – SYN : **fixe**

stable <dispositifs> : échelle, escabeau, escalier, marchepied, support, trépied – SYN : **statique ; équilibré**

stable <événements> : phénomène, vent – SYN : **cyclique**

stable <humains> être, quelqu'un, personne – SYN : **équilibré**

stable <jugements> : classement, conception, position, situation, point de vue, opinion – SYN : **fixe**

stable <lieux> : terrain, territoire, lieu, endroit, surface – SYN : **assuré, solide**

stable <meubles> : bibliothèque, chaise, étagère, fauteuil, table, tabouret – SYN : **statique ; équilibré**

stable <moyens de transport> : bateau, navire, vélo, voiture – SYN : **statique ; équilibré**

stable <occupations> : emploi, état, occupation, poste, travail – SYN : **fixe ; permanent**

stable <rapports> : connexion, équilibre, union, lien, paix, rapport, relation, sens – SYN : **fixe ; permanent**

stable <règles> : convention, doctrine, loi, norme, règle, proposition – SYN : **ferme ; fixe ; invariable**, (la loi de la gravité, les lois de la nature – SYN : **immuable**)

stable <systèmes> : couple, démocratie, économie, établissement, logiciel, pays, marché, famille – SYN : **solide ; durable**

stable <valeurs> : chiffre, poids, équilibre, température, pourcentage, proportion, valeur, niveau – SYN : **constant ; fixe ; invariable**

personne stable – SYN : **sédentaire**

FIXE (adj)

fixe <appareils> : téléphone, appareil, caméra, machine, antenne, prothèse, roue, pédale – SYN :

**stable**

fixe <astres> : astre, étoile – SYN : **immobile**

fixe <coûts> : capital, frais, impôt, loyer, montant, prix, salaire, somme, rémunération, revenu,

tarif – SYN : **stable**

fixe <directions> : sens, direction, trajet – SYN : **déterminé**

fixe <habitations> : adresse, domicile, emplacement, lieu – SYN : **permanent ; stable**

fixe <occupations> : emploi, occupation, poste, travail – SYN : **stable**

fixe <ouvrages> : figure, statue, sculpture – SYN : **immobile**

fixe <rapports> : sens, valeur – SYN : **stable**

fixe <règles> : décision, formule, norme, loi, procédure, règle, résolution, orthographe – SYN :

**stable ; ferme**

fixe <regard> : œil, regard, vue, yeux – SYN : **figé, immobile**

fixe <temps> : calendrier, date, fête, jour, horaire, heure, terme – SYN : **défini ; déterminé**

fixe <valeurs> : température, valeur, proportion – SYN : **stable**

accent fixe – SYN : **stable**

douleur fixe – SYN : **incessante, précise**

feu fixe – SYN : **constant**

idée, pensée fixe – SYN : **obsessionnelle**

point fixe – SYN : **stable**

sourire fixe – SYN : **fixé**

Les analyses qui viennent d'être données montrent que les deux adjectifs sont occasionnellement synonymes. Les adjectifs *stable* et *fixe* donnent lieu à des séquences de sens proche (synonymiques) dans leurs emplois avec plusieurs classes sémantiques : <appareils>, <coûts>, <habitations>, <occupations>, <rapports>, <règles>, <valeurs>. Ils sont conceptuellement voisins, par conséquent une grande partie de leurs emplois se recouvre, ce qui se manifeste par un nombre significatif d'accompagnateurs en commun. Le format en termes de classes d'objets permet de préciser les environnements de leur substituabilité.

Soulignons que deux ou plusieurs adjectifs peuvent se trouver dans une relation de synonymie par rapport à l'objet décrit. Ils peuvent dénoter la même propriété, exprimer la même caractéristique. Cependant, il faut garder à l'esprit que dans la langue rien ne peut être strictement similaire.

## 7. Quelques observations

Les adjectifs qualificatifs qui passent pour synonymes se distinguent souvent par la sélection de leurs collocatifs (cf. P. Blumenthal, 2006) – certaines séquences retenues forment souvent des suites statistiquement significatives : *idée fixe*, *regard fixe*, *à l'heure fixe*, *sans domicile fixe*, *douleur fixe*. Pour les locuteurs dont la langue maternelle est le français, ces combinaisons sont prévisibles en raison de leur récurrence dans le discours. De tels syntagmes nominaux constituent une problématique réelle dans une lexicographie contrastive. La suite – *idée fixe* constitue une combinaison privilégiée pour laquelle on ne va pas proposer de synonyme. Le syntagme *téléphone fixe* est beaucoup plus récurrent en discours que *téléphone stable*. Ainsi, en substituant les deux adjectifs, même dans des co-textes spécifiques, il faut vérifier la validité d'une telle séquence pour ne pas obtenir des suites inappropriées dans une langue. Ces restrictions de sélection sémantique méritent un traitement lexicographique spécial, ce qui complique leur structuration lexicographique. La frontière entre ce qui est libre dans la langue et ce qui est préconstruit constitue un défi pour tout traitement linguistique<sup>10</sup>.

## 8. Question de substituabilité

Il est communément admis que la substituabilité est un critère pertinent pour postuler la situation de synonymie. Elle se réfère à la possibilité de remplacer une unité par une autre dans une phrase ou un énoncé sans altérer la grammaire et le sens de la phrase. Soulignons que les deux adjectifs ne peuvent pas être totalement identiques du point de vue de leur sens, mais ils peuvent être substituables quant à leur interprétation. Chaque lexème a son sens et son fonctionnement propres – selon le principe avancé par L. Bloomfield (1962 : 82) : « ...each linguistic form has a constant and specific meaning. If the forms are phonemically different, we suppose that their meanings also are different ».

La description en termes de classes d'objets considère des suites canoniques composées d'opérateur et de ses arguments. Elle ne prend pas en compte le critère grammatical. Les comportements combinatoires des adjectifs qui passent

---

<sup>10</sup> Pour les analyses qui soulèvent la question du phénomène collocatif, nous renvoyons aux travaux de A. Tutin et F. Grossmann (2002) et M. Perz (2016).



pour des synonymes peuvent diverger, par conséquent la substituabilité est un phénomène qui engage l'axe syntagmatique. Il s'ensuit que les synonymes adjectivaux, bien qu'appartenant à la même catégorie grammaticale, ne connaissent pas la même distribution et ne peuvent pas être interchangeables que dans un nombre restreint de contextes. C'est le cas du couple synonymique : *stable* et *résistant*. Dans la séquence :

*une construction stable = une construction résistante*

*un métal résistant à la température = \*un métal stable à la température*

La prise en compte de ce paramètre syntagmatique réduit considérablement la validité de substituabilité, mais n'exclue non plus la synonymie de ces unités. C'est parce que les unités ne sont interchangeables que dans certains contextes possibles que la notion de synonymie apparaît problématique. Pour les linguistes, la notion de contexte est cruciale, surtout quand les unités linguistiques s'ouvrent à des situations extralinguistiques. Rappelons que la définition du contexte peut aller au-delà de l'environnement immédiat. Même si les unités connaissent la même distribution, cela ne garantit pas qu'ils peuvent être partout substituables. Il est évidemment impossible d'aborder ici en détail tous les aspects de la question de l'interchangeabilité, mais disons que l'épreuve de la commutation convoque plusieurs facteurs et finalement conduit à admettre que la substituabilité totale n'est pas observable dans la langue. Comme le fait pertinemment remarquer Kleiber :

ce n'est pas seulement « un sens » ou « du sens » que l'on commute, mais bien des unités lexicales, c'est-à-dire des unités possédant une forme et un sens !

(G. Kleiber, 2009 : 20)

## 9. Quelques pistes à explorer...

La voie de recherche adoptée nous conduit aussi à une interrogation sur la dimension pragmatique de la notion de synonymie, à savoir quel est l'espace minimal commun que les deux unités devraient partager pour postuler qu'elles sont synonymes. Autrement dit, quelles unités ont un sens suffisamment proche pour être classifiées comme synonymes ? Une telle question nous pousse à émettre l'hypothèse selon laquelle il faut parler plutôt de plusieurs degrés de synonymie que de synonymie tout court. Quelles que soient les difficultés qu'éprouvent les linguistes

à en donner une définition, ce rapport sémantique entre les unités est ce qui permet aux usagers de formuler leurs jugements sur ce qui est synonymiques ou ne l'est pas. Dans la langue, il y a des mots lexicaux qui semblent être synonymes hors de l'emploi en contexte, mais qui ne seront pas interchangeables en chaque énoncé. C'est le cas, par exemple, des couples conceptuellement très proches tels que : *facile* vs *simple*, *complexe* vs *compliqué*, *négatif* vs *péjoratif*. Cette observation nous conduit à une autre question – si l'attention des analyses synonymiques devrait se focaliser sur les différences ou sur les ressemblances d'emploi ? La réponse à cette question ne peut pas être définitive – elle va dépendre du couple synonymique analysé. En faisant recours à des adjectifs qui sont conceptuellement très voisins, les usagers veulent savoir où réside la différence entre leurs emplois et même les spécialistes du domaine ont du mal à fournir un échantillon d'emploi qui les diffère. Il est à remarquer que le format de description discuté ci-dessus permet de préciser les deux sphères – les classes d'objets à vocation linguistique précisent les ensembles pour lesquels la substitution est pertinente et ceux pour lesquels elle ne se fait pas.

## 10. En guise de conclusion

Bien que les unités lexicales soient employées largement comme synonymes dans les actes langagiers, il persiste des polémiques autour des critères défini-toires de la synonymie. En tenant compte des différentes facettes de la notion, nous optons pour la synonymie qui laisse plus de place à la substitution parce que l'usage de la synonymie est une pratique courante. Opinion que l'on retrouve, entre autres, chez M. Honeste (2007) qui reconnaît la prédominance de la synonymie dite partielle, quasi-synonymie. Ce type de synonymie, reconnue comme proximité sémantique entre les unités, offre les possibilités de substitution des unités dans certains contextes. Qui plus est, les dictionnaires encouragent ce type de démarche en proposant les définitions à l'aide de listes de mots de sens voisins.

Nous espérons avoir montré que la description en termes des classes d'objets semble être le cadre pertinent pour rendre compte des phénomènes de synonymie et les résultats obtenus sont encourageants quoique non dépourvus de difficultés. Du point de vue applicatif, la structuration en termes de classes sémantiques constitue une approche originale et intéressante à explorer en vue d'amélioration de descriptions existantes.

Soulignons pour terminer que le format des classes d'objets est une approche de nature linguistique. Autrement dit, la langue est vue en tant qu'objet d'analyse – elle peut être examinée et analysée indépendamment des facteurs extralinguistiques<sup>11</sup>. On peut se demander, pour prolonger le débat, si une telle prise de position influence la théorie de la signification. Considérer la théorie de la signification en adoptant un point de vue ontologique ou un point de vue linguistique nous donne une vision différente du concept de la signification des mots.

## Références citées

- Ajdukiewicz, K. (1985). *Język i poznanie. Tom 1*. PWN.
- Banyś, W. (2002). Bases de données lexicales électroniques – une approche orientée-objets : Partie I et II. *Neophilologica* 15, 7–29 ; 206–249.
- Banyś, W. (2005). Desambiguation des sens des mots et représentation lexicale du monde. *Neophilologica* 17, 57–76.
- Bloomfield, L. (1962). *The Menomini Language*. Yale University Press.
- Blumenthal, P. (2006). De la logique des mots à l'analyse de la synonymie. *Langue Française* 150, 14–31.
- Buvet, P. A. & Grezka, A. (2009). Les dictionnaires électroniques du modèle des classes d'objets. *Langages* 176, 63–9.
- Cadiot, P. (2009). Couleurs des mots ou synonymie. *Pratiques* 141/142, 26–38.
- Clas, A. & Gross, G. (1997). Les classes d'objets et la désambiguïsation des synonymes, *Cahiers de lexicologie* 70, 27–40.
- Doualan, G. (2011). Introduction à une approche instrumentée de la synonymie. *Cahier du Crisco* 32, 15–42.
- Doualan, G. (2014). Éléments pour une lecture de l'histoire de la synonymie. *Congrès Mondial de Linguistique Française*. 15–42.
- Goes, J. (1999). *L'adjectif. Entre nom et verbe*. Duculot.
- Gross, G. (1998). Pour une véritable fonction « synonymie » dans un traitement de texte. *Langages* 131, 103–14.
- Gross, G. (1996). Une catégorie hétérogène : l'adjectif. *Studi Italiani Di linguistica Teoretica e applicata* 25, 447–460.

---

<sup>11</sup> Un tel postulat a été avancé par un grand linguiste et logicien polonais K. Ajdukiewicz (1985). La théorie de la langue constituait le centre de ses recherches.

- Gross, G. (1994b). Classes d'objets et synonymie. *Annales Littéraires de l'Université de Besançon. Série Linguistique et Sémiotique* 23, 93–102.
- Gross, G. (1991). Typologie des adjectivaux. *Analyse et synthèse dans les langues romanes et slaves* 347, 163–178.
- Giry-Schneider, J. (2005). Les adjectifs intensifs : syntaxe et sémantique. *Cahiers de lexicologie* 86, 163–178.
- Honeste, M.-L. (2007). Entre ressemblance et différence : synonymie et cognition, *Le français moderne* 75, 160–173.
- Lehmann, A. & Martin-Berthet, F. (1998). *Introduction à la lexicologie. Sémantique et morphologie*. Duclot.
- Kleiber, G. (1999). *Problèmes de sémantique. La polysémie en question*. Septentrion.
- Kleiber, G. (2009). La synonymie – identité de sens n'est pas un mythe. *Pratiques* 141/142, 9–25.
- Moon, R. (1987). Monosemous words and the Dictionary, *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at University of Leeds*, 173–182.
- Murphy, L. (2003). *Semantic Relations and the Lexicon*. Cambridge University Press.
- Ploux, S. & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement automatique des langues* 39, 161–182.
- Noailly, M. (1999). *L'adjectif en français*. Éditions Ophrys.
- Ploux, S. & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement automatique des langues* 39, 161–182.
- Perz, M. (2016). Quelques réflexions sur le traitement automatique des collocations. *Translation and Meaning* 2, 153–165.
- Riegel, M. (1985). *L'adjectif attribut*. PUF.
- Riegel, M. (1993). Grammaire et référence : à propos du statut sémantique de l'adjectif qualificatif. *L'Information Grammaticale* 58, 5–10.
- Śliwa, D. (2015). La synonymie des termes juridiques polonais et français. Dans W. Fijałkowskiej, M. Izert, A. Kieliszczyk & E. Pileckiej (éds), *Être philologue. Mélanges offerts à Teresa Giermak-Zielinska* (237–250). Wydawnictwo Uniwersytetu Warszawskiego.
- Śmigielska, B. (2007). Assertion et suggestion dans les relations synonymiques. *Neophilologica* 19, 204–216.
- Tamba, I. (1998). *La sémantique*. PUF.
- Tutin, A. & Grossmann, F. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue Française de Linguistique Appliquée* 7, 7–26.



*Michele Prandi*

Université de Gênes  
Italie

 <https://orcid.org/0000-0002-6223-6946>

# Les relations transphrastiques à la frontière entre la phrase et le texte

## Interclausal relations at the boundary between sentence and text

### Abstract

Interclausal relations are traditionally studied within the framework of the complex sentence of endocentric subordinative form. However, there is no essential correlation between interclausal linkage and complex sentence. Interclausal relations are not meanings of subordinate clauses but conceptual relations that bridge saturated processes of equal rank. According to this premise, complex sentence is only one option for the expression of interclausal relations, along with text. Moreover, if we compare the competing structures, it turns out that the unmarked form of expression is not complex sentence but text. While a complex sentence imposes a hierarchic syntactic structure on a symmetric conceptual structure, the structure of the text and the conceptual structure of interclausal relations are isomorphic. As a marked option, the choice of complex sentence requires a specific functional motivation, which does not involve the ideation of the link itself but the communicative perspective: its function is to impose a layered communicative perspective on the connexion, which is distributed between foreground and background information.

### Keywords

Interclausal relations, complex sentence, text, up-bottom structures, bottom-up structures

Le projet en commun avec Gaston Gross qui a abouti à la publication de la monographie *La finalité. Fondements conceptuels et genèse linguistique* (Gross & Prandi, 2004) se fondait sur l'hypothèse qu'une seule et même relation transphrastique – dans notre cas, le but – pouvait déboucher sur un nombre très élevé de formes d'expression linguistiques aux propriétés structurales et aux contenus sémantiques différents, ouvertes aux choix des locuteurs. Une telle approche présupposait l'indépendance des relations transphrastiques de leurs formes

d'expression linguistiques, et particulièrement la remise en question de l'un des piliers de l'approche traditionnelle. Dans la tradition grammaticale, en effet, les relations transphrastiques sont associées à la phrase complexe, et notamment à la structure endocentrique subordinative qui contient une proposition principale indépendante et une proposition subordonnée en position de marge<sup>1</sup> : par exemple, *Le mur s'est effondré parce qu'il a plu pendant une semaine*. Dans cette optique, l'étude des relations transphrastiques se réduit à l'étude d'une poignée de propositions subordonnées étiquetée chacune avec le nom de la relation qu'elle est censée exprimer.

Le premier pas de notre projet commun ne pouvait être qu'une définition des relations transphrastiques indépendante de leurs formes d'expressions changeantes, et donc une définition qui se plaçait explicitement sur le plan des structures conceptuelles. L'objectif, dans le cas particulier, n'était pas en premier lieu de décrire la structure des propositions subordonnées dites finales mais d'identifier la structure conceptuelle de la relation entre une action et son but pour décrire ensuite de façon exhaustive l'éventail de ses formes d'expression hétérogènes. Gaston Gross voyait dans les relations transphrastiques autant de prédicats de deuxième ordre, à savoir des prédicats qui prennent comme arguments non pas des référents mais des procès saturés. La cause, par exemple, est un prédicat qui prend comme arguments deux événements du monde des phénomènes naturels ; le but, pour sa part, prend comme arguments une action humaine et une action ou un événement du monde naturel. Dans l'idée de Gaston Gross, j'ai reconnu tout de suite la même charpente conceptuelle qui était à la base de ma propre conception de grammairien philosophe, qui voyait les relations transphrastiques comme autant de relations conceptuelles entre procès saturés. Dans les deux perspectives, les relations transphrastiques sont dissociées à jamais de l'idée de subordination : les procès connectés se placent au même niveau, sont indépendants l'un de l'autre et sont les deux dominés par une relation qui crée un véritable pont conceptuel. La structure conceptuelle n'est pas en tant que telle subordinative mais symétrique, et ce n'est que l'expression dans la phrase complexe subordinative qui la change en une structure sémantique asymétrique. Si la forme d'expression est à son tour symétrique, comme il arrive dans la coordination et dans la juxtaposition, toute trace de subordination disparaît. Si la coordination appartient, comme la subordination, à la syn-

---

<sup>1</sup> L'étiquette *marge* (Longacre, 1985(2006)) est préférable à celle traditionnelle de *circonstanciel* du fait que les rôles non argumentaux incluent des couches différentes, et notamment les marges extérieures du procès, qui sont des circonstanciels au sens strict, les marges du prédicat, et les modificateurs du verbe (Prandi, 2004 : 268–276).

taxe de la phrase complexe, la juxtaposition nous transfère dans la dimension du texte cohérent.

Dans ma contribution à cet hommage, je me propose de tracer un bilan sur la position de la connexion transphrastique à la frontière entre la phrase et le texte. Pour ce faire, je commencerai par une analyse de la phrase complexe et de la place de la connexion transphrastique parmi ses structures et fonctions différentes. Ensuite, je vais souligner une affinité d'élection profonde entre la structure de la connexion transphrastiques et la structure des textes, qui brise son association millénaire avec la subordination grammaticale. Finalement, le tournant textuel pousse à revenir sur la phrase complexe dans une optique renversée : si la phrase complexe n'est pas le moyen d'expression d'élection des relations transphrastiques mais un choix marqué et, en tant que tel, motivé par une fonction spécifique, il faut s'interroger sur les raisons qui peuvent pousser un locuteur ou un écrivain à prêter à une structure symétrique une forme d'expression subordinative.

## 1. Deux structures et deux fonctions pour la phrase complexe

La phrase complexe subordinative est une phrase qui contient au moins une phrase, généralement appelée proposition, parmi ses constituants (Riegel, Pellet & Rioul, 1994 : 469). Dans les exemples suivants, la proposition subordonnée occupe les positions de sujet (1), de complément (2), de marge du prédicat (3), de circonstanciel externe au procès (4), de modificateur du verbe (5) et de modificateur du nom (6) :

1. C'est dommage que tu sois arrivé en retard
2. Luc craint qu'un orage va arriver
3. Luc a envoyé un message à Anne pour se féliciter avec elle
4. Quand il pleut, Marie ne sort jamais
5. Pierre marchait en trébuchant
6. La villa que tu vois au sommet de la colline a été projetée par Le Corbusier

Comme un constituant dans la phrase simple, une proposition subordonnée peut occuper une position d'argument ou de marge dans la phrase complexe. La distinction entre la proposition argumentale et la proposition marginale est tranchée, et se fonde sur des critères relevant de la structure syntaxique, du contenu

sémantique et de la fonction. Dans le cadre de mon argumentation, la différence de fonctions est stratégique.

Dans une phrase complexe qui contient une proposition subordonnée en position de marge, nous pouvons isoler un noyau – une proposition principale indépendante – et une proposition subordonnée non indépendante sur le plan syntaxique. Dans la phrase complexe *Le mur s'est effondré parce qu'il a plu pendant une semaine*, par exemple, il est possible d'isoler une phrase complète qui exprime un procès saturé indépendamment de la proposition subordonnée: *Le mur s'est effondré*. Dans une phrase complexe qui contient une proposition subordonnée en position d'argument, au contraire, il est impossible d'isoler une proposition principale indépendante. La raison est manifeste: la proposition qui exprime un argument est un constituant essentiel à l'intégralité conceptuelle du procès et à la bonne formation syntaxique de la phrase. Si de la phrase complexe *Luc craint qu'un orage va arriver* nous enlevons la phrase subordonnée, ce qui reste – *Lu craint* – n'est pas un noyau de phrase indépendant mais une séquence de mots dépourvue d'une structure en constituants.

Pour la même raison, une subordonnée marginale peut être séparée du noyau de la phrase principale et spécifiée dans une phrase indépendante qui forme avec la première un texte cohérent. Si la subordonnée exprime une circonstance externe, la reformulation contient une reprise anaphorique du noyau du procès en position de sujet du verbe *se passer*: la phrase complexe *Le mur s'est effondré parce qu'il a plu pendant une semaine*, par exemple, peut être reformulée dans le fragment de texte *Le mur s'est effondré. Cela s'est passé parce qu'il a plu pendant une semaine*. Si la subordonnée exprime une marge d'un prédicat d'action, la reformulation contient le pro-prédicat *le faire*, qui reprend le prédicat de l'action principale conservant son sujet: la phrase complexe *Luc a envoyé un message à Anne pour se féliciter avec elle*, par exemple, peut être reformulée comme *Luc a envoyé un message à Anne. Il l'a fait pour se féliciter avec elle*. Une proposition argumentale, au contraire, n'admet aucune des deux reformulations. La reformulation *George craint. Cela se passe qu'il va pleuvoir* échoue pour deux raisons: elle est incohérente parce qu'elle traite la séquence *Georges craint* comme si c'était un noyau de phrase saturé prêt à être repris par une anaphore saturée comme *cela*; dans la deuxième phrase, en outre, le verbe *se passer* n'arrive pas à intégrer la phrase subordonnée dans une structure bien formée et signifiante. La reformulation *George craint. Il le fait qu'il va pleuvoir* échoue pour des raisons semblables: le pro-prédicat *le faire* traite un verbe non saturé comme s'il était un prédicat et n'arrive à intégrer la phrase subordonnée ni dans une structure syntaxique bien formée ni dans une relation sémantique. La conclusion est incontournable: la phrase complexe qui contient une pro-



position subordonnée argumentale n'est pas l'issue de la combinaison de deux phrases<sup>2</sup>.

Sur le plan du signifié, la phrase complexe qui contient une subordonnée argumentale construit un seul procès complexe, formé d'un procès superordonné – dans notre exemple, la crainte d'une personne – qui contient comme argument un procès enchâssé : la pluie, objet de la crainte. Quand elle contient une subordonnée marginale, la phrase complexe contient deux procès saturés indépendants : dans nos exemples, l'effondrement du mur et la pluie, l'envoi du message et les félicitations.

La différence sémantique nous amène tout droit à la différence qui est pertinente pour notre sujet, à savoir la différence fonctionnelle. À la différence de ce qui se passe dans la phrase simple, les fonctions des phrases argumentales et des phrases marginales ne sont pas simplement différentes : elles sont incommensurables.

La fonction d'une phrase simple est la mise en œuvre d'un procès. À cette fonction répondent, de façon différente, tant les arguments que les marges. Les arguments saturent le verbe principal pour former le noyau du procès ; les marges expandent le noyau avec des relations conceptuelles cohérentes qui l'enrichissent. Entre les arguments et les marges, donc, il y a continuité fonctionnelle. Dans la phrase complexe, la continuité fonctionnelle est brisée.

---

<sup>2</sup> La définition de la phrase complexe comme combinaison de deux phrases est partagée par la tradition grammaticale et par la recherche typologique. La grammaire de Riegel, Pellat et Rioul (1994 : 469), par exemple, caractérise la phrase complexe comme l'une des stratégies disponibles pour combiner deux phrases, à côté de la juxtaposition, de la coordination et de l'insertion : « Les phrases complexes se distinguent traditionnellement selon leur mode de composition, c'est-à-dire selon la façon dont une ou plusieurs phrases constituantes  $P_1$ ,  $P_2$ , etc. [...] s'insèrent dans la structure globale d'une phrase constituée dite *matrice*  $P_0$  ». Les exemples de subordination donnés contiennent l'un une proposition subordonnée marginale – *Bien que les chiens aboient, la caravane passe* – et l'autre une subordonnée complétive – *J'attends que la caravane passe* – en contradiction avec l'affirmation qu'on lit quelques pages après, qui nie la présence d'une proposition principale indépendante dans la subordination complétive : « la subordonnée complétive est une partie de la principale au même titre que l'est le complément d'objet nominal ». Dans la recherche typologique, la subordination est généralement définie comme une forme de « linkage » entre une « main clause » et une ou plusieurs phrases subordonnées (Lehmann, 1988), ce qui implique que le terme *subordination* est un terme hyperonyme qui domine les hyponymes *subordination complétive* et *subordination adverbial* (Lehmann, 2018 : 30). Cristofaro (2003 : 2) définit toute les formes de subordination, et notamment « complement, adverbial and relative relations » comme « an asymmetrical conceptual/pragmatic relation between linked S[tates] o[f] A[ffair]s ». Une exception est Matthiessen, Thompson (1988) : à leur avis, dans la subordination argumentale « we do not have a clause combination [...] but simply a case of one clause functioning as a constituent, a complement, within another clause » (279).

La fonction des phrases argumentales est identique à celle des arguments dans la phrase simple, à savoir la mise en œuvre du noyau d'un procès grâce à la saturation de la valence du prédicateur<sup>3</sup>. Le procès n'est pas simple mais complexe, du fait qu'il contient un procès parmi ses arguments, mais il est unique, et sa construction n'est achevée qu'au moment où la subordonnée argumentale est intégrée dans sa structure : *George craint qu'il aille pleuvoir*. À la différence d'une marge d'une phrase simple, une phrase marginale n'a pas la fonction d'enrichir le procès mis en place par le noyau, mais de le relier à un procès logiquement indépendant dans une relation conceptuelle : la phrase complexe *Le mur s'est effondré parce qu'il a plu pendant une semaine*, par exemple, relie deux procès indépendants – l'effondrement du mur et la pluie – dans une relation de cause qui domine les deux. La conclusion est que la phrase complexe qui contient une phrase subordonnée marginale est l'issue de la combinaison de deux phrases indépendantes ; elle est donc un instrument de la connexion transphrastique à côté de la coordination et de la juxtaposition.

Après avoir identifié la fonction de la phrase subordonnée marginale avec la mise en place de relations transphrastiques, je vais maintenant analyser un échantillon représentatif de structures conceptuelles indépendantes, et notamment le micro-système formé par la cause, le motif et le but, avec un double objectif : démontrer l'indépendance et la priorité logiques des relations conceptuelles par rapport aux formes d'expression et remettre en question la corrélation électorale entre relations transphrastiques et phrases subordonnées.

## 2. La connexion transphrastique : de la phrase complexe au texte

### 2.1. Les relations transphrastiques comme structures conceptuelles

Les linguistes sont traditionnellement méfiants sur la possibilité de décrire avec exactitude les structures conceptuelles indépendamment de leurs formes d'expression linguistique ; cette idée, cependant, est dépourvue de fondement. À côté d'une grammaire des formes d'expression, l'usager de la langue maîtrise un système tout aussi solide de concepts cohérents partagés par-delà les frontières de chaque communauté linguistique, qui fournit sa matière première à la

---

<sup>3</sup> Le terme *prédicateur* (*predicator*) est utilisé par Lyons (1977 : 434) pour distinguer le pivot prédicatif d'une prédication de la relation grammaticale de prédicat (*predicate*).

mise en forme sémantique de la part des expressions linguistiques. Les expressions linguistiques ne fondent leurs signifiés ni sur la « nébuleuse », dont parle Saussure (1916(1978)), ni sur le « sable » de Hjelmslev (1943(1971)) mais sur un terrain solide de structures conceptuelles partagées. C'est la raison pour laquelle le « tournant linguistique » en philosophie (Rorty, 1967) doit être complété par un tournant philosophique en linguistique. Si le tournant linguistique en philosophie naît de la conscience que la voie d'accès privilégiée aux structures conceptuelles passe par une analyse fine de leurs formes d'expression linguistiques, le linguiste ne peut pas non plus oublier que la description des expressions linguistiques signifiantes demande un accès direct, indépendant des aléas du codage linguistique, à un système de concepts cohérents et à leurs conditions de cohérence : à une véritable grammaire des concepts. La description des formes d'expression linguistiques des relations transphrastiques demande comme condition préalable une description indépendante du contenu conceptuel de chacune e de ses conditions de cohérence. Pour fournir un exemple d'analyse conceptuelle appliquée aux relations transphrastiques, nous allons maintenant examiner un échantillon significatif, à savoir la constellation formée par la cause, les motifs et le but.

En premier lieu, il faut souligner que la cause et le but ne forment pas une opposition directe, comme le suggère l'association de chacune à une forme de phrase subordonnée, à savoir à la subordonnée dite causale et à celle dite finale. En effet, l'opposition pertinente sur le plan conceptuel s'établit entre la cause et les motifs qui poussent un agent à l'action, qui incluent le but. La cause est une relation qui relie deux événements réels du monde des phénomènes, à savoir un événement antérieur qui provoque la mise en place d'un événement successif : *Le mur s'est effondré parce qu'il a plu pendant une semaine*. Les motifs poussent un agent à accomplir une action (Daneš, 1985) : *Paul est sorti à pied parce qu'on lui a volé le vélo*. Si l'on s'arrêtait à la forme d'expression, on dirait que la structure de l'action motivée et de la cause sont parfaitement isomorphes, d'autant plus que la conjonction est la même. Les différences, cependant, sont essentielles et affleurent si l'on interroge les formes d'expression au jour de critères conceptuels portant à la fois sur le contenu des relations et sur leurs conditions de cohérence.

À la différence de la relation entre la cause et l'effet, la relation entre le motif et l'action n'est pas directe, mais comporte une décision. Pour que le vol du vélo pousse Paul à sortir à pied, il faut une décision de sa part, qu'Aristote, avec une métaphore suggestive inspirée par la vie de *la polis*, appelle « délibération » (*Étique à Nicomaque*, Livre III). Avant d'agir, Paul a passé en revue différentes options : par exemple, il aurait pu emprunter un bus, ou appeler un taxi. L'idée de décision n'est cohérente que si certains présupposés conceptuels sont satisfaits : le sujet d'une décision est un être humain libre et responsable de ses choix.

L'explicitation de la décision dans la forme linguistique est facultative, mais cohérente avec la structure conceptuelle du motif : *Paul a décidé de sortir à pied parce qu'on lui a volé le vélo*. Dans la structure conceptuelle de la cause, en revanche, il est clair qu'il n'y a aucune place pour une décision : poussé par l'eau qui a alourdi le terrain qu'il soutient, un mur ne peut pas choisir si s'effondrer ou pas.

La cause est cohérente avec les événements du monde des phénomènes ; le motif n'est cohérent qu'avec une action humaine. Cette différence peut être portée à la surface en manipulant l'expression linguistique, comme le suggère le tournant linguistique en philosophie, mais en même temps en évaluant les résultats au jour du critère de cohérence, qui est un critère conceptuel pertinent dans le cadre d'un tournant philosophique en linguistique. Les conditions de cohérence de la relation de cause affleurent grâce à une reformulation de la phrase complexe dans un fragment de texte qui se sert de la forme verbale *se passer*, dont les sujets cohérents sont les événements : *Le mur s'est effondré. Cela s'est passé parce qu'il a plu pendant une semaine*. Les conditions de cohérence de la relation de motif deviennent observables grâce à une reformulation de la phrase complexe dans un fragment de texte qui se sert du pro-prédicat *le faire*, qui n'est cohérente qu'en présence d'une action : *Paul est sorti à pied. Il l'a fait parce qu'on lui a volé le vélo*<sup>4</sup>. Une reformulation comme *\*Paul est sorti à pied. Cela s'est passé parce qu'on lui a volé le vélo*<sup>5</sup> est incohérente parce qu'un motif ne demande un événement quelconque qui se passe mais exige une action faite par un agent.

La structure temporelle de la cause est linéaire et n'inclut que deux moments : la cause précède l'effet dans le monde des phénomènes. La structure temporelle du motif n'est pas linéaire mais stratifiée et inclut trois moments. Dans notre exemple – *Paul est sorti à pied parce qu'on lui a volé le vélo* – les trois moments temporels, à savoir l'événement qui pousse Paul à agir, la décision et l'action, sont décalés : l'événement inspireur et l'action se situent dans le monde des phéno-

<sup>4</sup> *Le faire* est un prédicat à verbe support (Daladier 1978 ; Gross & Vivès, 1986 ; Gross, 1987, 1993, 1999, 2012 ; Giry-Schneider, 1987), l'équivalent de *faire cette action* : *Paul est sorti à pied. Il a fait cette action parce qu'on lui a volé le vélo*. Cette analyse, que je dois à mes discussions avec Gaston Gross, est confirmée par le comportement identique du pronom et du syntagme nominal. Quand ils occupent la position d'objet direct d'un verbe prédicatif, le pronom – *le* – et l'expression nominale – *cette action* – fonctionnent comme autant de reprises anaphoriques du procès antécédent, et de ce fait peuvent être remplacées par l'expression directe de l'antécédent en position de proposition objective *Paul est sorti à pied. Il m'a dit qu'il est sorti à pied parce qu'on lui a volé le vélo*. Quand les mêmes formes co-occurrent avec le verbe *faire*, elles n'admettent pas le remplacement par une proposition objective : *Paul est sorti à pied. Il a fait qu'il est sorti à pied parce qu'on lui a volé le vélo*. C'est la preuve que *faire* n'est pas utilisé comme verbe prédicatif qui prend un objet direct mais comme verbe support régi par le nom prédicatif ou par son substitut pronominal.

<sup>5</sup> Lastérisque signale l'incohérence conceptuelle de la combinaison.

mènes ; la décision, au contraire, se situe dans le monde intérieur de l'agent. Il s'agit de deux lignes temporelles logiquement distinctes. Si la temporalité externe est subie, la temporalité interne est le contenu d'une attitude intentionnelle maîtrisée par l'agent qui, au moment de la décision, peut aussi bien reparcourir le passé que prévoir ou projeter le futur. Si la relation de cause est contrainte dans la cage de la succession temporelle du monde externe, la structure de la temporalité interne affecte en profondeur la structure des motifs. À côté de motifs rétrospectifs, qui évaluent des faits du passé, s'ouvre l'espace logique pour des motifs prospectifs, fondés sur la prévision ou sur l'intention qui investissent le futur (Anscombe, 1956(1968)). Le motif que nous avons observé – *Paul est sorti à pied parce qu'on lui a volé le vélo* – plonge ses racines dans le passé : il est un exemple de motif rétrospectif. *Marc a pris la voiture parce qu'il craignait la pluie* est un exemple de motif prospectif fondé sur la prévision : ce qui pousse l'agent à la décision et à l'action n'est pas quelque chose qui s'est produit dans le passé mais le contenu d'une prévision. *Marc est allé au travail à pied parce qu'il voulait faire de l'exercice* est un exemple de motif prospectif fondé sur l'intention : ce qui pousse l'agent à la décision et à l'action n'est ni quelque chose qui s'est produit dans le passé ni le contenu d'une prévision, mais le contenu d'une intention.

La structure temporelle des motifs prospectifs se distingue de la structure temporelle des motifs rétrospectifs sur deux points. En premier lieu, le motif prospectif contient quatre moments distribués entre le monde externe et le monde interne de l'agent : le moment de l'intention ou de la prévision et le moment de la décision se situent dans le monde interne de l'agent ; le moment de l'action et de la réalisation du contenu de l'intention se situent dans le monde externe. En outre, et surtout, dans le motif prospectif ce n'est pas l'action qui clôt la séquence temporelle, mais la réalisation éventuelle du contenu de l'intention ou de la prévision. Il faut souligner par ailleurs que la réalisation du contenu prévu ou projeté est une donnée empirique qui n'affecte pas la cohérence de la relation. La seule condition nécessaire pour la cohérence de l'action, en effet, est la réalité de la prévision ou de l'intention.

La différence entre prévision et intention dans le cadre des motifs prospectifs nous mène au but : le but n'est rien d'autre qu'un motif prospectif qui coïncide avec le contenu d'une intention de l'agent (Gross & Prandi, 2004). La relation exprimée par la phrase complexe *Marc est allé au travail à pied parce qu'il voulait faire de l'exercice*, en effet, est la même que celle confiée à la structure dite finale *Marc est allé au travail à pied pour faire de l'exercice*.

L'analyse comparée de la cause, des motifs et du but nous montre que nous disposons d'un système de relations conceptuelles partagées, une véritable grammaire des concepts cohérents qui, avant de fournir sa matière première

à l'expression linguistique, est à la base de notre comportement cohérent. Nous tous sommes capables de lire dans le livre ouvert du monde les relations de cause et d'interpréter les motifs et les but de nos actions et de celles de nos semblables. Le socle rocheux sur lequel notre forme de vie fonde sa cohérence, ce « noyau massif, central de la pensée humaine qui n'a pas d'histoire, ou du moins n'est pas enregistré dans les histoires de la pensée » (Strawson, 1959(1964 : 9)), est stable dans notre patrimoine de ressources partagé auquel nous faisons entièrement confiance, bien plus stable que les structures lexicales et grammaticales des langues. En premier lieu, les langues changent dans le temps. La langue dans laquelle s'exprimait Aristote nous est étrangère et il nous faut un dur exercice pour l'apprendre et avoir un accès direct à ses pensées. Le contenu de ses réflexions sur les causes des événements et sur les buts des actions humaines, au contraire, est un patrimoine dans lequel, après deux millénaires et demi, nous nous reconnaissons sans difficulté. En outre, et surtout, l'expression des structures conceptuelles partagées est variable à l'intérieur d'un même état de langue ; dans quelques cas, comme nous pourrions le constater dans le prochain paragraphe, ses formes seraient même déroutantes, si un timon conceptuel indépendant et fiable ne nous aidait pas à garder toujours la route droite.

## 2.2. Relations transphrastiques et propositions subordonnées

Dans la tradition grammaticale (Riegel, Pellat & Rioul, 1994 : 503–518), mais aussi dans la description linguistique (Kortmann, 1997) et dans la recherche typologique (Cristofaro, 2003), les relations transphrastiques sont associées à la phrase complexe comme autant de signifiés de phrases subordonnées spécifiques : les relations de cause et de but, par exemple, sont associées à la phrase dite causale et à la phrase dite finale. Cette idée, cependant, est à la fois dépourvue de support empirique et déroutante sur le plan méthodologique. Si nous confrontons les relations conceptuelles que nous avons identifiées – cause, motif, but – avec les types de propositions subordonnées décrites dans la grammaire du français, il est facile de découvrir que la corrélation postulée par la tradition grammaticale n'est pas satisfaite.

En premier lieu, à côté des phrases causales et des phrases finales nous chercherions en vain des phrases « motivaes » : c'est le signe le plus direct d'un manque de correspondance plus général entre les relations conceptuelles et les formes de phrase subordonnée. D'une part, la forme d'expression dite causale – *p parce que q* – admet d'exprimer tant la cause (7) que les motifs (8, 9), inclus le but (10). D'autre part, la même relation conceptuelle de but, comme nous l'avons

déjà observé, peut être confiée à la forme causale (10) aussi bien qu'à la forme finale (11) :

7. Le mur s'est effondré parce qu'il a plu pendant une semaine
8. Pierre est rentré à la maison parce qu'il avait oublié les clés de la voiture
9. Pierre est rentré à la maison parce qu'il craignait un orage
10. Pierre est rentré à la maison parce qu'il voulait prendre les clés de la voiture
11. Pierre est rentré à la maison pour prendre les clés

Pour compléter le tableau, force est de constater que la forme d'expression en mesure de coder le but de façon adéquate n'est pas la forme finale mais la forme dite causale, qui contient nécessairement un verbe ou un prédicat relevant de la sphère de l'intention ou qui l'impliquent : l'exemple (10), par exemple, contient le verbe *vouloir*. La forme non finie de phrase finale introduite par la préposition *pour*, qui est probablement la plus utilisée, ne code pas le constituant conceptuel essentiel de la relation finale, à savoir l'intention de l'agent. Dans ces conditions, définir le but comme le contenu d'une phrase finale est tout à fait déroutant. Observons les exemples :

12. Le fleuve traverse la plaine en traçant d'amples méandres pour se jeter dans la mer quelques kilomètres plus loin
13. Le fleuve traverse la plaine en traçant d'amples méandres dans le but de se jeter dans la mer quelques kilomètres plus loin
14. Pierre s'est couché à neuf heures pour rester éveillé toute la nuit
15. Bien qu'il se soit couché à neuf heures pour se reposer d'une longue journée de travail, Pierre est resté éveillé toute la nuit.
16. Pierre s'est couché à neuf heures dans le but de rester éveillé toute la nuit
17. Pierre s'est couché à neuf heures pour se reposer d'une longue journée de travail
18. Pierre s'est couché à neuf heures dans le but de se reposer d'une longue journée de travail

La phrase complexe (12) montre que la forme ne code pas l'intention : en effet, elle n'attribue pas au fleuve l'intention de se jeter dans la mer, comme le fait la variante (13). À la différence de *pour*, la locution *dans le but de* code l'intention du référent du sujet comme une implicature conventionnelle (Grice, 1975), qui est une composante implicite mais non effaçable du signifié de l'expression : le résultat est un signifié conflictuel interprété comme une métaphore qui attribue au fleuve une intention, et donc la nature humaine. Les mêmes remarques valent

pour le couple de phrases (14) e (16). Du fait qu'elle ne code pas l'intention de l'agent, la première est interprétée comme l'expression d'une contre-finalité qui frustre l'intention de l'agent, comme dans une relation concessive (15). Du fait qu'elle code l'intention de l'agent, la phrase (16) lui attribue une intention contradictoire. Comme (17), la phrase (18) attribuée à Pierre l'intention de se reposer, mais pas aux mêmes conditions : dans (18), l'intention est codée, alors qu'en (17) elle est inférée à partir de la nature humaine du référent du sujet, qui le met en condition d'accomplir une action finalisée. Les différences examinées soulignent le rôle des signifiés complexes conflictuels comme observatoire privilégié sur les conditions formelles et conceptuelles de la mise en œuvre des signifiés complexes, et notamment sur l'équilibre entre codage et inférence. Dans le couple d'exemples (17) et (18), l'issue cohérente neutralise la différence entre codage et inférence. Les couples (12, 13) (14, 16), en revanche, mettent en relief les limites du codage : comme la cohérence de (12) et de (14) montre que l'intention n'est pas codée par la préposition *pour*, l'incohérence de (13) et le conflit dans l'intention exprimé par (16) ne se justifient que si l'intention est codée par la locution *dans le but de*.

La conclusion est incontournable. Les relations transphrastiques comme la cause, le motif et le but sont des structures conceptuelles partagées, indépendantes des aléas de leurs formes d'expression multiples et accessibles tant à la pensée cohérente de la part de l'utilisateur – à l'inférence – qu'à une analyse rigoureuse de la part du linguiste.

### 2.3. De la phrase au texte : structures descendantes et structures ascendantes

S'il est vrai que les relations transphrastiques sont des relations conceptuelles symétriques qui relient par un pont conceptuel des procès indépendants, la conclusion est obligée : relier avec des ponts conceptuels cohérents des énoncés indépendants et de même rang n'est autre chose que l'essence du texte, sa *quidditas*, comme le rappelle Conte (1988(1999 : 29)). Dans le fragment de texte cohérent *L'orage a provoqué un éboulement. La route est coupée*, en effet, nous reconnaissons la même relation de cause confiée à la phrase complexe *La route est coupée parce que l'orage a provoqué un éboulement*. Si l'approche traditionnelle voit l'expression des relations transphrastiques dans la phrase complexe comme allant de soi, notre approche l'envisage comme un choix marqué pour lequel il est nécessaire de fournir des raisons fonctionnelles. Pour ce faire, il faut s'interroger sur les relations, tantôt non marquées et tantôt marquées entre la structure



des relations transphrastiques d'une part, et la structure de la phrase complexe subordinative et la structure du texte d'autre part.

Si nous comparons la structure d'une phrase et la structure d'un texte, nous constatons qu'elles présentent des caractéristiques incommensurables. À la différence de la structure d'une phrase, unifiée par une charpente grammaticale unitaire, la structure d'un texte prend forme directement sur le plan du contenu, où elle construit sa complexité pas après pas. S'il est vrai que tant une phrase qu'un texte ont une structure, force est de reconnaître que les structures linguistiques rentrent dans deux types qui présentent des propriétés opposées : les structures que nous pouvons appeler descendantes et les structures ascendantes. Le passage de la phrase au texte nous conduit d'un type de structure à l'autre.

Dans les structures descendantes, le tout – la structure complexe – a la priorité logique par rapport aux parties : aux constituants. Les constituants, réciproquement, tirent leur valeur de la relation qu'ils entretiennent avec la structure complexe. Un exemple de structure descendante est la structure d'un bâtiment : les parties qui le forment acquièrent une valeur sur la base de leur fonction dans la statique de la construction. Dans une cathédrale gothique, par exemple, les piliers, les croisières, les contreforts et les arcs-boutants se définissent à partir de leur position dans la structure du bâtiment et de leur fonction de contribuer ensemble à garantir sa statique.

Dans les structures ascendantes, la relation entre le tout et les parties se renverse : la structure du tout résulte de la combinaison de parties constituantes qui jouissent chacune d'une priorité logique par rapport à la combinaison. Un exemple est la configuration urbanistique d'une ville, formée de bâtiments et d'espaces libres indépendants l'un de l'autre. Si un bâtiment est démoli pour faire place à un jardin, ou si une ancienne usine est transformée en une salle de concert entourée d'un parc, la physionomie d'une partie de la ville change, mais ce changement ne se répercute pas sur la structure de chacun des autres quartiers et bâtiments.

Parmi les expressions linguistiques, le cas paradigmatique de structure descendante est le noyau de la phrase nucléaire simple, alors que la structure ascendante trouve sa forme la plus typique dans le texte. Le noyau d'une phrase est une hiérarchie de relations grammaticales formelles qui se dessine au fur et à mesure que la phrase est analysée progressivement dans ses constituants jusqu'aux constituants ultimes. À l'intérieur de la hiérarchie, chaque relation grammaticale se définit à partir de sa position dans la structure complexe qui l'englobe. Dans une phrase transitive comme *Pierre admire la fresque*, par exemple, le sujet et l'objet direct ont la même forme – il s'agit d'expressions nominales – mais des positions différentes. Le sujet est un constituant immédiat de la phrase, la contrepartie du prédicat, alors que l'objet direct est un constituant immédiat du prédicat,

la contrepartie d'un verbe transitif. La structure d'un texte répond à un critère d'organisation opposé. La structure d'un texte est le résultat de la connexion d'au moins deux procès, signifiés de deux phrases indépendantes sur le plan grammatical, par un pont conceptuel qui est indépendant des deux. Une structure comme *L'orage a provoqué un éboulement. La route est coupée*, par exemple, forme un texte parce que les procès juxtaposés sont reliés sur le plan conceptuel par une relation cohérente de cause en l'absence de toute connexion grammaticale.

La distinction entre structures ascendantes et descendantes est un premier pas, nécessaire mais non suffisant, pour identifier la différence entre la phrase et le texte, car les structures linguistiques présentent une dimension supplémentaire : comme la langue en tant que système, toutes les expressions de chaque niveau, du signe simple au texte, sont des structures organisées sur deux plans – l'expression et le contenu (Hjelmslev, 1943(1971)) – qui ne sont pas isomorphes (Prandi, 2023 : 94–97). La présence de deux plans ajoute une dimension perpendiculaire à la distinction entre la phrase et le texte. La phrase est en premier lieu une hiérarchie de relations formelles, alors que le texte est un premier lieu un réseau de relations conceptuelles. Si nous prenons en charge les deux dimensions – l'opposition entre structures descendantes et structures ascendantes d'un côté, la priorité logique de l'expression ou du contenu de l'autre – nous pouvons conclure que le noyau de la phrase est une structure formelle descendante, alors que le texte est une structure conceptuelle ascendante.

### **3. La fonction de la phrase complexe subordonnée : la perspective communicative**

Si maintenant nous revenons aux relations transphrastiques et à leurs formes d'expression, une réflexion s'impose : la structure conceptuelle des relations transphrastiques est ascendante, leur forme d'expression est ascendante dans le texte, quand deux phrases indépendantes sont juxtaposées et reliées à *posteriori* par un pont conceptuel, et descendante dans la phrase complexe, quand une proposition subordonnée entre comme un constituant dans la structure unitaire d'une phrase superordonnée. Si cela est vrai, la conséquence est que les relations transphrastiques et leurs formes d'expression sont isomorphes dans le texte mais pas dans la phrase complexe. La question de l'isomorphisme entre expressions et contenus nous amène tout droit à la question de la motivation iconique des formes d'expression. L'isomorphisme entre expression et contenu, en effet, n'est

pas une condition nécessaire pour la signification des expressions complexes, mais il est une condition nécessaire pour leur motivation iconique.

La question sur la nature de la syntaxe, formelle et autonome ou instrumentale et iconique, traverse comme une véritable faille théorique toute la linguistique contemporaine. Traditionnellement, l'alternative entre autonomie et motivation de la syntaxe est vue comme exclusive : la syntaxe dans son ensemble est ou autonome (Bloomfield, 1933 ; Harris, 1946 ; Wells, 1947 ; Chomsky, 1957) ou instrumentale vis-à-vis des contenus conceptuels et donc iconique (McCawley, 1970(1971) ; Lakoff, 1971 ; Fillmore, 1968 ; Haiman, 1985 ; Langacker, 1987 ; Dik, 1997). Pour Chomsky (1957 : 17), « grammar is autonomous and independent of meaning », et « *uniquely* determines [...] semantic interpretation » (Chomsky, 1966 : 5). Pour Haiman (1985 : 2), « The linguistic form is a diagram of conceptual structure ». Les deux instances méthodologiques opposées partagent un présupposé commun : les structures de l'expression et les structures du contenu sont isomorphes. Que ce soit l'expression à modeler le contenu ou le contenu à modeler l'expression, la forme des deux ne peut qu'être la même. Or, si nous vérifions que dans la structure de chaque phrase des structures syntaxiques motivées par le contenu, et donc isomorphes, prennent le relais de structures syntaxiques autonomes non isomorphes, et que donc le présupposé de l'isomorphisme n'est pas toujours satisfait, l'idée de l'incompatibilité entre l'autonomie de la syntaxe et la motivation conceptuelle perd son fondement. Aucun obstacle logique n'empêche une syntaxe autonome d'interagir avec un système de concepts tout aussi autonome, avec un éventail d'issues hétérogènes qui peuvent être vérifiées dans la structure des formes d'expression. Si cela est vrai, la question sur l'autonomie de la syntaxe devient une question empirique qui ne porte pas sur sa présence ou absence mais sur ses limites.

Le passage d'une syntaxe autonome à une syntaxe motivée et la corrélation entre isomorphisme et motivation peuvent être observés d'une façon directe dans la structure de la phrase simple. Si nous comparons la structure formelle du noyau de la phrase – la combinaison entre un syntagme nominal, le sujet, et un syntagme verbal, le prédicat – avec la structure conceptuelle du noyau du procès, nous constatons que le présupposé de l'isomorphisme n'est pas satisfait, du moins dans les langues qui présentent un alignement nominatif-accusatif. La structure conceptuelle du procès a un centre : le prédicateur, typiquement un verbe, contrôle le nombre et le contenu de tous ses arguments, inclus l'argument confié au sujet. La structure distributionnelle de la phrase nucléaire, au contraire, n'a pas de centre : il s'agit d'une structure exocentrique<sup>6</sup> qui contient deux constituants

---

<sup>6</sup> Pour la distinction entre constructions endocentrique et exocentrique, voir Bloomfield (1933 : 194).

appartenant à des classes en distribution complémentaire – un syntagme nominal et un syntagme verbal – dont la combinaison produit une phrase, et donc une structure qui n'appartient à aucune des deux. L'absence d'isomorphisme soustrait à la motivation iconique son espace logique. Dans les langues qui partagent l'alignement nominatif-accusatif, le sujet grammatical a une forme propre qui code le premier argument de tout procès, indépendamment de la nature verbale ou nominale du prédicat et, à plus forte raison, de la valence (Tesnière, 1959(1965)) et du contenu du prédicateur. De ce fait, la structure sujet – prédicat est une structure formelle autonome qui ne reflète la structure d'aucun procès et est compatible avec tous.

À la différence de la relation entre le sujet et le prédicat, le prédicat présente une structure syntaxique qui est isomorphe à la structure conceptuelle du procès : le prédicat contient autant de compléments du verbe qu'il y a d'arguments excédant le premier, confié au sujet. Donc, sa structure formelle reflète comme un diagramme<sup>7</sup> la section correspondante du procès. La présence d'une motivation diagrammatique, et donc l'isomorphisme, n'implique pas l'iconicité au niveau de la structure formelle interne des différents constituants syntaxiques, mais ouvre sa possibilité logique. L'observation de la structure interne des compléments confirme cette hypothèse : au jour du paramètre de la motivation iconique, les compléments se comportent de deux façons opposées. La nature autonome ou iconique des compléments dépend de la présence ou de l'absence de relations grammaticales, et donc d'une propriété ouverte à l'observation empirique.

L'exemple par excellence de relation grammaticale autonome interne au prédicat est l'objet direct. Comme le sujet, l'objet direct se caractérise par des propriétés formelles exclusives. En tant que tel, il n'est pas associé à un argument spécifique mais il est prêt en accueillir plusieurs en fonction du contenu relationnel du prédicateur qui le contrôle. Dans les exemples suivants, le rôle de l'objet direct passe du patient (19), au résultat d'une action (20), à l'expérienceur passif (21), au stimulus (22) et au destinataire (23) :

19. Marc a réparé le vélo
20. Botticelli a peint cette Nativité
21. L'orage a terrorisé Béatrice
22. Béatrice craint les orages
23. Marc a informé les élèves de l'horaire des cours

---

<sup>7</sup> Parmi les icônes (Peirce, 1902(1978)), on distingue les images, qui se basent sur une ressemblance ponctuelle, des diagrammes (Jakobson, 1966), qui reflètent un réseau de relations.

Le cas le plus significatif d'une forme d'expression qui reflète comme une icône son contenu conceptuel est celui des arguments locatifs des verbes d'état, de mouvement et de déplacement. Les verbes de déplacement, par exemple, contiennent typiquement trois arguments : l'instigateur, l'entité qui bouge et la destination. À la différence des deux premiers arguments, confiés au sujet et à l'objet direct, la destination n'est pas associée à une relation grammaticale indépendante, mais est identifiée immédiatement comme une relation conceptuelle, et notamment comme une relation spatiale, dont le contenu motive la structure de la forme d'expression. La préposition, notamment, est choisie à l'intérieur d'un paradigme très vaste en fonction de sa capacité à dépeindre avec exactitude une relation spatiale donnée : par exemple, *Jean a envoyé son fils en face de la cathédrale / derrière la cathédrale / au sommet du clocher de la cathédrale / au fond de la nef de la cathédrale*, et ainsi de suite<sup>8</sup>.

Comme le montre le cas de la destination, la motivation iconique ouvre en même temps l'espace logique pour les options et les choix. En présence de relations grammaticales, la langue impose des règles rigides : un exemple est la sélection de la préposition qui introduit l'objet prépositionnel (Steinitz, 1969) des verbes intransitifs à deux arguments, qui n'est pas laissée au choix du locuteur mais est sélectionnée par le verbe : *renoncer*, par exemple, sélectionne *à*, *dépendre*, *de*, *compter*, *sur*, et ainsi de suite. Quand la forme de codage est iconique et donc motivée, la langue n'impose pas à l'usager une forme exclusive, mais lui offre, pour chaque relation conceptuelle, un répertoire d'options parmi lesquelles il est libre de choisir. À la différence de la motivation, cependant, l'ouverture d'un espace logique pour les options et le choix n'est pas nécessairement une conséquence de l'isomorphisme ; au contraire, il y a des cas où l'alternative entre une forme d'expression isomorphe et non isomorphe s'offre à son tour comme une option et donc comme l'objet d'un choix. Ce point nous approche de l'identification de la valeur fonctionnelle de la phrase complexe dans l'expression des relations transphrastiques.

L'alternative entre formes d'expression isomorphes et non isomorphes fait à son tour l'objet d'un choix dans deux cas : dans l'expression des relations conceptuelles marginales d'un procès simple et des relations transphrastiques. La spécification

---

<sup>8</sup> Je parle du codage relationnel pour la structure syntaxique du noyau de la phrase, qui est formée par une hiérarchie de relations grammaticales formelles indépendante des contenus conceptuels, et de codage ponctuel pour l'expression directe de relations conceptuelles indépendantes : par exemple, l'instrument ou les relations spatiales. Le codage ponctuel est confié directement à la forme d'expression isolée, et notamment à une préposition, sans la médiation d'un réseau de relations grammaticales indépendantes. En fonction du contenu de la préposition, le codage ponctuel est une valeur graduée qui s'étend du sous-codage au codage adéquat au surcodage (Prandi, 2004 : 61–62).

des arguments du prédicateur dans le noyau du procès ne peut qu'avoir lieu à l'intérieur de la structure de la phrase. Au-delà du noyau, comme nous l'avons constaté, l'expression des couches de relations conceptuelles marginales associées à un procès simple et des relations transphrastiques peut avoir lieu tant à l'intérieur d'une structure de phrase unitaire que dans la dimension textuelle. Le noyau d'une action comme *Jean a coupé le bois*, par exemple, est prêt à recevoir l'expression d'un rôle marginal comme l'instrument tant à l'intérieur d'une seule et même phrase – *Jean a coupé le bois avec cette hache* – que dans un fragment de texte cohérent grâce à des instruments de cohésion appropriés : *Jean a coupé le bois. Il l'a fait avec cette hache*. Également, la phrase complexe *La route est coupée parce que l'orage a provoqué un éboulement* et les fragments de texte *L'orage a provoqué un éboulement. La route est coupée ; La route est coupée : L'orage a provoqué un éboulement* expriment la même relation de cause. Or, comme le procès simple et les connexions transphrastiques d'une part, et la phrase et le texte d'autre part, n'ont pas la même structure, il est clair que le choix radical entre grammaire et texte remet en question l'isomorphisme entre formes d'expression et contenus conceptuels. Examinons les différentes options dans le détail.

Un procès simple élargi présente une structure conceptuelle descendante. Par conséquent, son expression est isomorphe dans la phrase, dont la structure est tout aussi descendante, mais pas dans le texte, dont la structure est ascendante. La connexion transphrastique présente une structure ascendante. Par conséquent, son expression est isomorphe dans le texte, dont la structure est tout aussi ascendante, mais pas dans la phrase complexe, dont la structure est descendante. L'absence d'isomorphisme entre la structure conceptuelle et la forme d'expression fournit un critère pour discriminer, au niveau de la fonction idéationnelle (Halliday, 1970), les formes d'expression non marquées des formes marquées : dans l'expression tant des rôles marginaux que des relations transphrastiques, nous pouvons considérer comme non marquée la structure isomorphe – qui reflète la structure conceptuelle – et marquée l'option non isomorphe. Dans l'expansion du procès simple, l'option non marquée est la phrase et l'option marquée le texte ; dans la connexion transphrastique, la relation est renversée : l'option non marquée est le texte et l'option marquée la phrase.

S'il est vrai que la disponibilité d'options multiples demande une motivation fonctionnelle pour chaque choix et que la motivation présuppose l'isomorphisme, la motivation des choix marqués, à savoir du texte dans l'expansion du procès simple et de la phrase complexe dans la connexion transphrastique, ne relève pas de la fonction idéationnelle, dans laquelle le présupposé de l'isomorphisme n'est pas satisfait. En effet, elle répond à la fonction que Halliday (1970) qualifie de textuelle, qui envisage les structures syntaxiques comme autant d'instruments au

service de la perspective communicative (Mathesius, 1928(1964) ; Daneš, 1974, 1974 ; Firbas, 1970, 1974, 1992) et notamment du dynamisme communicatif (Firbas, 1964) : au lieu de refléter la structure du contenu conceptuel, la structure de l'expression reflète des hiérarchies de poids communicatif.

En présence d'un procès simple, le choix marqué permet de fractionner l'expression d'une structure conceptuelle unitaire en unités communicatives distinctes, ce qui confère à chacune un poids communicatif plus élevé. En présence d'une connexion transphrastique, la phrase complexe impose une perspective communicative hiérarchique en haut-relief à une structure conceptuelle en tant que telle symétrique et aplatée. Pour saisir le point, confrontons une séquence de phrases juxtaposées (24), qui est isomorphe à la relation conceptuelle, et une phrase complexe subordinative (25), qui ne l'est pas :

24. Marie était un train d'écrire une lettre. Tout d'un coup, Jean entra dans la pièce

25. Pendant que Marie écrivait une lettre, Jean entra dans la pièce

Dans la juxtaposition (24), chaque phrase indépendante a sa perspective communicative et il n'y a pas de perspective unitaire qui inclut les deux. Comme dans une galerie de portraits, chaque tableau est fermé dans son cadre. La phrase complexe (25), au contraire, impose aux procès reliés un cadre syntaxique unitaire, et donc une perspective communicative à son tour unitaire. Comme dans un tableau de la Renaissance, le procès principal – l'entrée de Jean – occupe le premier plan sur le fond du procès subordonné. À la perspective interne à chaque phrase, qui présente une progression d'un thème à un foyer, s'ajoute une perspective unitaire de la phrase complexe, distribuée entre un procès de premier plan et un procès de fond. Dans la phrase complexe, la hiérarchie entre information de premier plan et information de fond reproduit en miniature la hiérarchie qui caractérise le texte narratif entre la progression des événements et la description des scènes qui les accueillent. Le facteur commun au macrocosme du texte narratif et au microcosme de la phrase complexe est la distribution complémentaire des temps verbaux de base autour desquels s'organisent les relations temporelles de la narration avec leurs implications aspectuelles (Weinrich, 1964) : le passé simple, qui est le temps de la narration des événements du premier plan, et l'imparfait, qui est le temps de la description des référents, des milieux et des procès placés sur le fond.

La conclusion est incontournable. La tradition grammaticale voit dans la phrase complexe l'instrument d'élection de la connexion transphrastique, et donc un instrument au service de la fonction idéationnelle. Or, l'analyse des données encourage l'hypothèse que le privilège de la phrase complexe réside plutôt dans

sa capacité d'imposer une perspective hiérarchique en haut-relief à une relation conceptuelle symétrique, valorisant le manque d'isomorphisme entre forme et contenu idéationnel comme une ressource précieuse au service de la perspective communicative.

## Références citées

- Anscombe, G. E. M. (1956(1968)). Intention. *Proceedings of the Aristotelian Society* 57, 321–332. (Réimpr. dans A. R. White (éd.), *The Philosophy of Action* (144–152). Oxford: Oxford University Press).
- Aristote (1959). *Éthique à Nicomaque*. Vrin.
- Bloomfield, L. (1933(1970)). *Language*. Holt, Rinehart and Winston. (*Le langage*. Payot).
- Chomsky, N. A. (1957(1969)). *Syntactic Structures*. Mouton. (*Structures syntaxiques*. Seuil).
- Cristofaro, S. (2003). *Subordination*. Oxford University Press.
- Daladier, A. (1978). *Problèmes d'analyse d'un type de nominalisation en français et de certains groupes nominaux complexes*. Thèse de doctorat, Université Paris VII.
- Daneš, F. (1974). Functional sentence perspective and the organization of the text. Dans F. Daneš (éd.), *Papers in Functional Sentence Perspective* (106–128). Mouton.
- Daneš, F. (1985). Some remarks on causal relationships in language and text. *Recueil Linguistique de Bratislava* 8, 151–157.
- Dik, S. C. (1997). *The Theory of Functional Grammar. Part I: The Structure of the Clause*, 2<sup>ème</sup> éd. Mouton De Gruyter.
- Fillmore, Ch. J. (1968). The case for case. Dans E. Bach & R. Harms (éds.), *Universals in Linguistic Theory* (1–88). Holt, Rinehart & Winston.
- Firbas, J. (1964). On defining the theme in functional sentence analysis. *Travaux Linguistiques de Prague* 1, 267–80.
- Firbas, J. (1970). On the interplay of means of FSP. Dans *Actes du X<sup>ème</sup> congrès international des linguistes, Vol. II* (741–745). Éditions de L'Académie de la République Socialiste de Roumanie.
- Firbas, J. (1974). Some aspects of the Czechoslovak approach to problems of FSP. Dans F. Daneš (éd.), *Papers in Functional Sentence Perspective* (11–37). Mouton.
- Firbas, J. (éd.) (1992). *Functional Sentence Perspective in Written and Spoken Discourse*. Cambridge University Press.
- Giry-Schneider, J. (1987). *Les prédicats nominaux en français. Les phrases simples à verbe support*. Droz.




- Grice, H. P. (1975). Logic and conversation. Dans P. Cole & J. L. Morgan (éds.), *Syntax and Semantics* 3 (41–58). Academic Press.
- Gross, G. (1987). *Les constructions converses du français*. Droz.
- Gross, G. (1993). Trois applications de la notion de verbe support. *L'information grammaticale* 59, 16–23.
- Gross, G. (1999). Verbes supports et conjugaison nominale. *Revue d'études francophones* 9, 70–92.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses Universitaires du Septentrion.
- Gross, G. & Prandi, M. (2004). *La finalité: fondements conceptuels et genèse linguistique*. De Boeck-Duculot.
- Gross, G. & Vivès, R. (1986). Les constructions nominales et l'élaboration d'un lexique-grammaire. *Langue française* 69, 5–27.
- Haiman, J. (1985). Introduction. Dans J. Haiman (éd.), *Iconicity in Syntax* (1–7). John Benjamins.
- Haiman, J. & Thompson, S. A. (1988). *Clause combining in grammar and discourse*. John Benjamins.
- Halliday, M. A. K. (1970). Language structure and language function. Dans J. Lyons (éd.), *New Horizons in Linguistics* (140–165). Penguin Books.
- Harris, Z. (1946). From Morpheme to Utterance. *Language* 22, 161–183.
- Hjemslev, L. (1943(1971)). *Omkring Sprogteoriens Grundlæggelse*. Bianco Lunos Bogtrykkeri. (*Prologomènes à une théorie du langage*. Minuit).
- Jakobson, R. (1966). À la recherche de l'essence du langage. *Diogène* 51, 22–38.
- Kortmann, B. (1997). *Adverbial Subordination*. Mouton De Gruyter.
- Lakoff, G. (1971). Presuppositions and relative well formedness. Dans D. D. Steinberg & L. A. Jakobovits (éds), *Semantics* (329–340). Cambridge University Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar, I*. Stanford University Press.
- Lehmann, Ch. (1988). Towards a typology of clause linkage. Dans J. Haiman & S. A. Thompson (éds), *Clause Combining in Grammar and Discourse* (181–225). John Benjamins.
- Lehmann, Ch. (2018). Linguistic concepts and categories in language description and comparison. Dans M. Chini & P. Cuzzolin (éds), *Tipologia, acquisizione, grammaticalizzazione – Typology, Acquisition, Grammaticalization Studies*. Franco Angeli.
- Longacre, R. E. (1985(2007)). *Sentences as combinations of clauses*. Dans T. Shopen (éd.), *Language typology and syntactic description. Vol. 2. Complex constructions*, 2<sup>ème</sup> éd. (235–286). Cambridge University Press.
- Lyons, J. (1977). *Semantics, Vol. I*. Cambridge University Press.
- Mathesius, V. (1928(1964)). On linguistic characterology with illustration from modern English. *Actes du premier congrès international des linguistes*, 56–63. Mouton. Réimpr.

- dans J. Vachek (éd.), *A Prague School Reader in Linguistics* (59–67). Indiana University Press.
- Matthiessen, Ch. & Thompson, S. A. (1988). The structure of discourse and ‘subordination’. Dans J. Haiman & S. Thompson (éds), *Clause Combining in Grammar and Discourse* (275–317). John Benjamins.
- McCawley, J. D. (1971). Where do noun phrases come from? Dans D. D. Steinberg & L. A. Jakobovits (éds), *Semantics* (217–231). Cambridge University Press.
- Peirce, Ch. S. (1902(1962)). *Logic and Semiotic: Theory of Signs*. Dans C. S. Peirce (éd.), *Philosophical Writings* (98–119). Dover.
- Prandi, M. (2004). *The Building Blocks of Meaning*. John Benjamins.
- Prandi, M. (2023). De la dimension relationnelle du lexique à la syntaxe : une frontière à retracer. Dans A. Roig & A.-G. Toutain (éds.), *Concert mondial de linguistique. Mélanges offerts à Franck Neveu* (91–103). ENS Éditions.
- Riegel, M., Pellat, J.-Ch. & Rioul, R. (1994). *Grammaire méthodique du français*. Presses Universitaires de France.
- Rorty, R. (éd.) (1967). *The Linguistic Turn: Recent Essays in Philosophical Method*. The University of Chicago Press.
- Saussure (de), F. (1916(1972)). *Cours de linguistique générale*. Payot.
- Steinitz, R. (1969). *Adverbial-Syntax*. Akademie Verlag.
- Strawson, P. F. (1959(1978)). *Individuals. An essay in Descriptive Metaphysics*. Methuen & Co.
- Tesnière, L. (1959(1966)). *Éléments de syntaxe structurale, 2<sup>ème</sup> éd.* Klincksieck.
- Weinrich, H. (1964). *Tempus. Besprochene und erzählte Welt*. Kohlhammer.
- Wells, R. (1947). Immediate constituents. *Language* 23, 81–117.



*Miroslaw Trybisz*

Universidad de Gdańsk  
Polonia

 <https://orcid.org/0000-0002-1514-2621>

# La actualización del nombre *COVID-19* (y sus sinónimos) en los portales de información de lengua española

**The actualization of the name *COVID-19* (and its synonyms) in the Spanish language information portals**

## **Abstract**

With the beginning of the COVID-19 pandemic, people needed information about the new disease. The journalists immediately facilitated it in the information portals. Due to lack of specialized knowledge, they frequently used metonymy very often, so as a result several popular synonyms acquired by readers appeared. In the text we analyze how the predicative name COVID-19 (also considering all possible synonyms) belonging to the category of <infectious diseases> is actualized in the texts of the Spanish language press online. We present the determinants and light verbs which are mentioned used by Spanish-speaking journalists, and this allows to know better not only the syntax of the name in question but also the combinatory inventory of all the predicative names of the same class. As the texts studied are characterized by a certain level of specialization, the results of the study may do not correspond to the analysis of other, more or less specialized, texts., more or less specialized.

## **Keywords**

COVID-19, graphic-phonetic variants, synonymy, actualization, determination, light verbs, Spanish language, information portals

El año 2020 empezó con amenaza de una crisis global. La revelación y descripción de una nueva enfermedad aguda en la ciudad de Wuhan en China en diciembre de 2019 provocó que el mundo empezara a prepararse al ataque de un brote mortal. El 30 de enero de 2020, la Organización Mundial de la Salud proclamó el estado de emergencia sanitaria de preocupación internacional y el 11 de marzo reconoció la enfermedad como una pandemia. Aquel día fueron confirmados los casos positivos en 114 países del mundo. Entre tanto, el 11 de febrero, la OMS asignó también la denominación oficial de la nueva enfermedad: *COVID-19*, realmente una sigla compleja formada a partir del nombre complejo de la lengua inglesa *coronavirus disease 2019*.

Este nombre de origen puramente científico, introducido a las lenguas naturales de una manera artificial, en su forma escrita empezó a aparecer en los medios de comunicación de masas del mundo entero. Por esta vía se propagó entre los usuarios de diferentes lenguas. Sin embargo, la construcción poco natural de la palabra, compuesta de una sigla inglesa y un número abreviado, provocó cierta oposición de los usuarios de las lenguas naturales, tanto los laicos, como los periodistas. Aparecieron variantes gráficas y, por consecuencia, gráfico-fonéticas, y muchas de ellas son abreviaciones del término original. Así, en español aparece el nombre *covid* y poco después *cóvid*, ambos escritos con minúsculas, sin o con tilde sobre la *o* según la pronunciación, como palabra llana o aguda.

Son estas formas, y en primer lugar la sin tilde, las que pone como lema el *Diccionario histórico de la lengua española (DHLE)* de la RAE (2013-), un diccionario que sigue desarrollándose. La forma *COVID-19* encabeza la subentrada, considerada variante de *covid* en español. En el *DLE, Diccionario de Lengua Española* de la RAE (2022), el lema tiene la forma escrita con mayúsculas, *COVID*. La forma oficial también encabeza la subentrada. En ambos diccionarios se considera la ambigüedad del género con preferencia para el femenino destacada en el *DHLE*. Este diccionario nota también la posible metonimia de *covid* para designar a los enfermos.

La descripción lexicográfica normativa de la Real Academia Española constituye el punto de partida para nuestras investigaciones. Queremos comprobar la combinatoria del nombre *COVID-19*, tomando en consideración sus variantes y sinónimos, en los textos de varios portales de información en español.

El corpus proviene de los textos mediáticos que se pueden consultar de una manera libre en las páginas web de los portales de información local, regional y nacional de casi todos los países hispanohablantes, pero también localizados en otros países, como la *BBC* en el Reino Unido o *Deutsche Welle* en Alemania. Los hemos recogido en una buscadora creada con la ayuda del motor de búsqueda *Google*, que ponemos a la disposición pública a través de la dirección

<https://cse.google.com/cse?cx=14f6f12a9478640a8>. El corpus tiene un carácter abierto y está sometido a las modificaciones según las operaciones realizadas en los portales de origen. Actualmente, se recogen casi cien direcciones correspondientes a casi cien portales de información con todas las páginas que almacenan. Tenemos la intención de seguir desarrollándolo para otros proyectos. El carácter del corpus no permite indicar el número de textos recogidos ni proceder a un análisis cuantitativo o cualitativo.

## 1. Variantes y sinónimos de *COVID-19*

Para recoger el corpus necesario, empezamos por establecer el conjunto de los signos lingüísticos escritos que refieren al mismo concepto que refiere el término oficial *COVID-19*. Está cierto que no podremos identificar todos los sinónimos posibles como no podemos comprobar todos los títulos de prensa en línea posibles.

Hay que observar que las autoridades intentan aprovechar la necesidad social de la información y buscan la ayuda de los medios de comunicación para propagar las actitudes que les parecen deseables. En su estudio sobre el proceso de la mediatización de la pandemia por gripe porcina de 2009 en diversos países, Daniel Hallin y otros (2020: 2) observan que, en los casos de enfermedades emergentes, el conocimiento público/mediático precede muy a menudo la consolidación del conocimiento médico/científico.

El “sentido común” indicaría que el conocimiento en el campo de la salud se desarrolla inicialmente en el ámbito de la medicina científica para luego ser transmitido por las instituciones y los profesionales a través de diversos canales, incluyendo los medios de comunicación. Sin embargo, las epidemias muchas veces se convierten en objeto de conocimiento público/mediático antes de consolidarse como objeto de conocimiento médico/científico.

(Hallin, Briggs, Mantini-Briggs, Spinelli, Sy; 2020: 2)

De esta manera se establece una divergencia notable entre ambos tipos de conocimiento: para los medios, las hipótesis no confirmadas que aparecen junto al aumento del interés social por un tema concreto suelen tener más valor que los resultados finales de las investigaciones, ya que para determinarlos se necesita demasiado tiempo desde el punto de vista de un ciudadano medio quien quiere

adquirir este tipo de conocimiento para poder comunicar con otros miembros de la sociedad y presentar su propia opinión. El grado de especialización depende en gran medida de la información (o desinformación) que recibe.

Podemos observarlo en las noticias que los portales de información del mundo entero emitieron, de una manera urgente, el 11 de marzo de 2020, al inicio de la pandemia:

- (1) *La OMS declara el brote de coronavirus pandemia global* (<https://elpais.com/sociedad/2020-03-11/la-oms-declara-el-brote-de-coronavirus-pandemia-global.html>, <https://www.heraldo.es/noticias/internacional/2020/03/11/oms-brote-coronavirus-convertido-pandemia-1363410.html>);
- (2) *La Organización Mundial de la Salud (OMS) ya considera el brote del nuevo coronavirus como una pandemia. “Hemos evaluado que el covid-19 puede caracterizarse como una pandemia”, dijo este miércoles Tedros Adhanom, director de la OMS.* (<https://www.bbc.com/mundo/noticias-internacional-51842708>);
- (3) *Ya es oficial: la Organización Mundial de la Salud (OMS) califica a la crisis sanitaria del coronavirus como pandemia, sin adjetivos.* (<https://www.elmundo.es/ciencia-y-salud/salud/2020/03/11/5e694c48fc6c8350428b4680.html>);
- (4) *OMS declara pandemia mundial por coronavirus [...] La Organización Mundial de la Salud (OMS) declaró oficialmente este miércoles la pandemia global de coronavirus, infección que hasta ahora acumula 121 mil contagiados en todo el mundo, 4.373 fallecidos y 118 países afectados, incluido Chile.* (<https://www.latercera.com/que-pasa/noticia/oms-declara-pandemia-al-coronavirus/2F2VP5KFDRG3RJA3DD6BCGC46U/>);
- (5) *Coronavirus es declarado una pandemia [...] La Organización Mundial de la Salud (OMS) anunció ayer que el coronavirus (COVID-19) ya puede definirse como una “pandemia” [...]* (<https://copades.com/monec/?p=41559>).

Al referirnos a la definición de *pandemia*, que la califica como una *enfermedad epidémica (DLE)*, y a la fecha de la publicación de las informaciones, que indica el inicio de la *pandemia por COVID-19*, podemos constatar que en los ejemplos desde (1) hasta (5), el *coronavirus* o *el brote de coronavirus* están considerados, a base de la metonimia de carácter popular, como sinónimos que se refieren exactamente a la misma enfermedad. Incluso, como observamos en (2), el uso del nombre compuesto *el brote del nuevo coronavirus* permite evitar la repetición del término *covid-19* que aparece en la citación de las palabras del director de la OMS. En (5), la denominación oficial de la enfermedad aparece entre paréntesis como explicación de *coronavirus* repetido.

En estos casos, el fenómeno de sinonimia tiene un carácter circunstancial, incluso aleatorio, ya que procede de la incomprensión del término especializado *pandemia*. La pregunta es si podemos considerar el signo *coronavirus* como sinónimo de *COVID-19* si (lo hemos advertido *supra*) se trata de un error de contenido desde el punto de vista especializado.

Aquel día, esta información, muy rebuscada por los lectores potenciales, necesitaba una gestión rápida para que se encontrara en seguida al inicio de la página principal de los portales. Los que la gestionaban y preparaban los textos finales no fueron especialistas del dominio de epidemiología ni tampoco de medicina general. Los términos especializados tienen un sentido específico y unívoco en el contexto de un dominio concreto y únicamente para la gente que posee un conocimiento perfecto de este dominio. Podemos observar que, si se cita a especialistas, como en (2), no se califica el *coronavirus* como *pandemia*. La misma tendencia, con algunas excepciones, puede observarse también en las páginas de los establecimientos universitarios.

A. Kacprzak y J.-P. Goudaillier (2014: 1–2) indican que se debe considerar la variación diastrática del léxico de la medicina en varios niveles. Ellos mismos destacan cuatro tipos de léxico usados en cuatro tipos de situaciones de comunicación: el puramente científico usado en la comunicación oficial de médicos, en congresos o simposios; el científico usado en la comunicación no oficial entre médicos; el profesional usado en la comunicación entre médicos y pacientes y el popular usado por pacientes.

Podemos considerar que la sinonimia popular es un fenómeno muy frecuente si nos referimos a las denominaciones de las especies biológicas, como plantas o animales. Cada comunidad cultural denominaba sus descubrimientos en este dominio de su propia manera y hay que admitir que el conocimiento popular muchas veces precedía el científico en estos dominios, sobre todo en lo que se refiere a las plantas medicinales. De todos modos, como lo confirman también A. Kacprzak y J.-P. Goudaillier (2014: 2), la existencia de la sinonimia popular, al lado de la sinonimia especializada, tiene por objetivo asimilar lo desconocido, sobre todo para intercambiar la información necesaria.

Los periodistas también, en los primeros días de la pandemia, carecían del conocimiento especializado. Tampoco podían consultar rápidamente a especialistas. Leían textos en inglés y, a base de sus conocimientos generales sobre la medicina, tenían que describir la situación en español, de una manera más detallada posible. A lo largo de los primeros días, e incluso las primeras semanas, cometieron muchos errores de contenido que fueron adquiridos por los lectores de sus textos, cuyos conocimientos especializados fueron aún menores y esperaban que pudieran aprender mucho al consultar los portales de información.

Claro está que estos nombres no son sinonímicos para un especialista del dominio. Sin embargo, podemos observar que, en varios contextos no especializados, los usuarios del español seguían denominando la consecuencia por la causa, la enfermedad por el brote que la provoca:

- (6) [...] *les resulta imposible por dedicar cuerpo y alma a arrebatarse de la muerte al prójimo que padece el SARS-CoV-2.* (<https://www.adelante.cu/index.php/es/a-fondo/25-entrevistas/20900-desde-zona-roja-en-angola-la-felicita-cion-para-annalieth-camaguey>);
- (7) *Donald Trump tiene la coronavirus.* (<https://quizlet.com/532384580/ya-lo-sabia-flash-cards/>).

No es un fenómeno nuevo, desconocido antes. Lo podemos observar también en el caso de otras dolencias, sobre todo de carácter infeccioso:

- (8) [...] *El palista sufre el virus Epstein-Barr [...]* (<https://www.lavozdegalicia.es/noticia/deportes/2021/04/28/tenia-cansancio-habia-dias-me-costaba-le-vantarme/00031619623299408112770.htm>);
- (9) [...] *pacientes en distintas situaciones clínicas de la enfermedad HIV.* (<http://www.scielo.org.ar/pdf/medba/v64n5/v64n5a06.pdf>);
- (10) *Diagnostican que padece de la bacteria sporothrix schenckii* (<https://www.elsoldetampico.com.mx/local/bacteria-le-impide-caminar-usa-muletas-desde-hace-16-anos-ahora-cynthia-tiene-una-esperanza-10193811.html>).

En todos estos casos, las enfermedades provocadas por los brotes evocados tienen denominaciones específicas en los contextos especializados: el virus de Epstein-Barr provoca la mononucleosis; el HIV, también un virus, provoca el sida y *Sporothrix schenckii* no es una especie de bacteria, sino de hongo dimórfico que puede provocar la esporotricosis (en contextos menos especializados adquiere el nombre de *enfermedad del manipulador de rosas*).

Tratamos pues todos los nombres que refieran de una manera metonímica a la *enfermedad infecciosa provocada por el virus SARS-CoV-2* como sinónimos de *COVID-19* aunque para un especialista puedan referir únicamente al brote. En este caso, tomamos en cuenta la relación entre un emisor y un receptor, ambos no especialistas, que entienden el signo de la misma manera. La arbitrariedad del signo lingüístico no se limita a la lengua, sino al habla, a un uso concreto entendido como una interacción en un acto de comunicación concreto. Tomamos en consideración también todas las definiciones que aparecen en los



textos de diverso nivel de especialización y las consideramos como sinónimos descriptivos.

El término *pandemia* también puede significar una dolencia. Sin embargo, casi no se lo usa para denominar la dolencia en cuestión, por lo menos en los portales de información, sino más bien en un sentido figurativo, para hablar de unas consecuencias de la pandemia por COVID-19, sobre todo sociales o económicas:

(11) *El sector audiovisual padece la pandemia.* (<https://noticias.imer.mx/blog/columna-el-sector-audiovisual-padece-la-pandemia/>).

Como podemos observar, los límites de la sinonimia de un signo lingüístico de origen especializado dependen muy a menudo del contexto en el que se describe el concepto. Se trata tanto del contexto textual como del nivel de especialización de los autores y lectores del texto.

Por sus rasgos definitorios, consideramos que el signo de base, así como todos sus variantes y sinónimos, pertenecen a la categoría de <enfermedades infecciosas>. La clase de <enfermedades> en francés fue descrita por P.-A. Buvet (1998). Este estudio nos servirá de base para hablar del nombre *COVID-19* (sus variantes y sinónimos) como predicado nominal de primer orden y para evocar algunos contextos en los que aparece en posición de argumento implicado por un predicado de orden superior. Así, vamos a observar los principales actualizadores y modificadores de este conjunto de nombres que refieren al mismo concepto.

## 2. Determinación

### 2.1. Alcance de la determinación

Para hablar de la determinación, tenemos que establecer su alcance.

Podemos hacer referencia al estudio de G. Gross (1996: 61–68) quien afirma que, al lado de la clase de determinantes formalmente simples, en el caso del español antepuesta al nombre, podemos observar el uso de la determinación compleja que puede tener la forma combinada de dos o más determinantes simples:

(12) *Todos los colores del otoño.*

o sea, compuesta de un predeterminante (determinante simple o complejo antepuesto) y un modificador (antepuesto o pospuesto):

(13) *Ponte la falda hawaiana.*

(14) *Ponte la falda que acortaste la semana pasada.*

(15) *María tiene una nueva falda.*

G. Gross (1996: 62) y P.-A. Buvet (1998: 93), entre otros, subrayan también que la determinación no depende únicamente del nombre, sino que hay que considerar como unidad de base toda la frase, ya que la determinación depende también de otros componentes como el tiempo y, sobre todo, el aspecto.

## 2.2. Restricciones de uso

En varios estudios se subraya la relación entre la determinación y la clase de objetos o predicados a la que pertenece un nombre, sobre todo en lo que se refiere a las restricciones de uso de los determinantes. Como el aspecto es uno de los componentes del sentido, los determinantes dependen mucho más del carácter de un predicado que de una clase de objetos. Al lado de P.-A. Buvet (1998: 92), lo indican también D. Le Pesant y M. Mathieu-Colas (1998: 17). Una descripción más detallada está presentada por G. Gross (2008, 2013). A base de sus análisis, la determinación, al lado de los verbos soporte y las construcciones de acontecimiento, está concebida como uno de los elementos que permite establecer los límites de la clase de predicados en su propuesta de un *Bescherelle de los nombres predicativos*.

## 2.3. Artículo definido

En primer lugar, podemos confirmar el uso del artículo definido, como lo hemos observado en los ejemplos desde (1) hasta (7). Es el actualizador más frecuente.

Nos parece interesante y justa la clasificación de X. Blanco y P.-A. Buvet (1999) quienes distinguen la categoría de los definidos de todas las demás que consideran como las relacionadas con determinantes indefinidos.

## 2.4. Artículo indefinido con modificador

El uso del artículo indefinido requiere un modificador, generalmente pospuesto:

- (16) [...] *nadie está exento de que un covid agudo se complique [...]* ([https://es.ara.cat/sociedad/explosion-contagios-jovenes-ahoga-cap\\_130\\_4040358.html](https://es.ara.cat/sociedad/explosion-contagios-jovenes-ahoga-cap_130_4040358.html));
- (17) *No se observó que una covid leve aumentara el riesgo de hemorragias internas.* (<https://www.bbc.com/mundo/noticias-61017731>).

## 2.5. Artículo cero con modificador

El modificador puede complementar también el artículo cero:

- (18) *Descubren coincidencias genéticas en pacientes que desarrollan coronavirus grave* (<https://es.ara.cat/-/descubren-coincidencias-geneticas-en-pacientes-que-desarrollan-coronavirus-grave>).

## 2.6. Modificador como elemento constituyente de una unidad fijada

Con una serie de modificadores observamos un cambio de significado o, mejor dicho, el modificador hace parte del nombre complejo, semifijado porque pueden aparecer varios adjetivos, junto con *COVID-19* (o sus variantes) y por ello aparece con mucha frecuencia con el artículo definido o el artículo cero:

- (19) *Estos son los síntomas más comunes de la covid persistente en adultos y niños* (<https://www.heraldo.es/noticias/salud/2022/06/27/covid-persistente-sintomas-adultos-ninos-1584329.html>);
- (20) *Alemania advierte sobre los efectos de covid-19 largo* (<https://www.dw.com/es/alemania-advierte-sobre-consecuencias-del-covid-19-de-larga-duraci%C3%B3n/a-62425949>).

Se trata de una dolencia considerada como consecuencia o incluso continuación de la infección por *COVID-19*, su forma crónica.

En general, no observamos la misma composición con la mayoría de los sinónimos populares que hemos evocado en la parte anterior. Hay excepciones, pero con poca frecuencia de uso:

- (21) *COVID prolongado: Vacunas Pfizer y Moderna reducen la aparición de coronavirus largo* (<https://www.elfinanciero.com.mx/salud/2022/12/30/covid-prolongado-vacunas-pfizer-y-moderna-reducen-la-aparicion-de-coronavirus-largo/>);
- (22) *¿Cómo saber si padezco de coronavirus prolongado?* (<https://www.tudn.com/edicion-digital-california/como-saber-si-padezco-de-coronavirus-prolongado-dolor-de-cabeza-y-decaimiento-entre-los-sintomas-video/embed>).

## 2.7. Ambigüedad en cuanto al género

Como podemos observar, *COVID-19* es un nombre ambiguo en cuanto al género. Tradicionalmente, los nombres de la clase de <enfermedades> son de género femenino. Una excepción apareció con el sida. Sin embargo, hay que observar que, desde el punto de vista especializado, el sida no es una enfermedad, sino un síndrome que puede causar un padecimiento más fácil de enfermedades. La RAE acepta únicamente el masculino, sin embargo, en el uso podemos encontrar también el femenino:

- (23) *La sida también puede contagiarse en el parto.* (<https://www.elmundo.es/ciencia-y-salud/salud/2020/07/14/5f0da57e21efa040708b45e8.html>)

Para la COVID-19, como lo hemos dicho, se aceptó el uso de una sigla prestada directamente de la lengua inglesa, una lengua en la que los géneros no juegan un papel primordial y en la que las enfermedades son de género neutro que no se usa en español. Además, una identificación frecuente de la enfermedad con el brote provoca esta vacilación de los usuarios nativos del español sobre el género:

- (24) *Vacuna contra la gripe y el covid [...]* (<https://www.abc.es/salud/enfermedades/coronavirus/>);
- (25) *Descubren una relación entre uno de los fármacos usados contra la covid y mutaciones del coronavirus* (<https://www.abc.es/salud/enfermedades/coronavirus/>);
- (26) *El príncipe Carlos de Inglaterra tiene el coronavirus* (<https://apnews.com/341d13331233443392b12bfa6da293b4>);
- (27) *Se espera el resultado de las pruebas para determinar si también tiene la coronavirus.* (<https://cholilaonline.ar/2020/04/murio-un-bebe-de-cuatro-dias-en-brasil-por-coronavirus.html>).

## 2.8. Determinante posesivo

De los usos poco frecuentes se nota sobre todo el de un determinante posesivo que indica una relación directa entre la enfermedad y el enfermo:

- (28) [...] *la mujer creía que su covid sería una anécdota más que una tragedia.* (<https://cubasi.cu/es/noticia/cantante-antivacunas-muere-tras-contraer-covid-proposito>).

## 2.9. Determinante numeral

Observemos también el uso del numeral ordinal:

- (29) *El tiempo entre el primer y el segundo covid es clave para entender esta reinfección [...]* (<https://www.redaccionmedica.com/secciones/sanidad-hoy/la-reinfeccion-covid-relativamente-comun-con-las-variantes-omicron-9324>).

## 2.10. Determinante indefinido

En los usos con el determinante indefinido complejo *todo el / toda la*, a veces combinado también con un partitivo complejo, podemos observar que COVID-19 cambia de referente y no se trata más de una enfermedad concreta, sino del conjunto de los enfermos de esta dolencia:

- (30) *Ahora mismo todo el covid que hay en el mundo es de Ómicron [...]* (<https://www.elindependiente.com/vida-sana/salud/2023/01/07/covid-ano-iii-china-la-subvariante-kraken-y-una-ola-despues-de-navidad/>);
- (31) [...] *la vacuna previno el 95% de toda la covid [...]* (<https://elpais.com/ciencia/2020-12-19/respuestas-para-las-principales-incognitas-sobre-las-vacunas-contra-la-covid.html>).

## 2.11. Determinante demostrativo

En varios casos, el determinante demostrativo tampoco actualiza el nombre de la enfermedad, sino que indica, junto con el nombre, una ola concreta de la pandemia por COVID-19:

- (32) *la cifra de personas fallecidas por esta COVID-19 alcanza las 233,622 víctimas* ([https://www.cronica.com.mx/notas-la\\_cifra\\_de\\_personas\\_fallecidas\\_por\\_esta\\_covid\\_19\\_alcanza\\_las\\_233622\\_victimas-1193900-2021.html](https://www.cronica.com.mx/notas-la_cifra_de_personas_fallecidas_por_esta_covid_19_alcanza_las_233622_victimas-1193900-2021.html));

pero en las transcripciones de la lengua hablada, el demostrativo subraya la determinación del nombre en su sentido principal y el demostrativo tiene un carácter enfático:

- (33) *Estamos bien y a orar [...] por todos los que tengan este Covid-19.* (<https://www.tvazteca.com/aztecauno/venga-la-alegria/notas-vla/notas/lupita-da-lessio-no-tiene-covid-19-nuera-charito-esposa-ernesto-dalessio-contagio>).

## 2.12. El plural

El uso del plural está muy restringido y también refiere a las olas de la pandemia:

- (34) *[...] de que las restricciones a los covidés podrían estar disminuyendo [...]* (<https://www.plus500.com/es-PL/NewsAndMarketInsights/Chinas-Lock-downs-Mute-Demand>).

Hay que destacar que no hemos encontrado en los portales de información recogidos ningún uso del plural con un numeral cardinal.

## 3. Verbos soporte

### 3.1. Delimitación de la clase

Los verbos soporte adoptan denominaciones diversas, como las enumera G. P. Comezaña (2022: 24): *verbos ligeros, livianos, vicarios, de apoyo* o incluso *verbos de soporte*. En todos estos casos se trata de una clase de verbos que sirven para actualizar los predicados no verbales (sobre todo nominales, pero también adjetivales) desde el punto de vista temporal y aspectual.

Según D. Le Pesant y M. Mathieu-Colas (1998: 12–17), lo que ha confirmado también G. Gross (2013: 187–200), los verbos soporte no constituyen una clase homogénea. Los verbos soporte generales se pueden aplicar a toda una clase semántica y los apropiados caracterizan una clase de predicados restringida.

Además, si la mayoría de los verbos soporte carece de contenido semántico, se usan también las extensiones aspectuales que añaden un matiz aspectual al concepto expresado por el predicado no verbal. Finalmente, los verbos de acontecimiento se usan en oraciones impersonales en las que se enfoca la existencia (o la no existencia) de la noción del predicado.

### 3.2. Verbos soporte general

En el caso de *COVID-19*, el verbo soporte general es *tener*:

- (35) [...] *las personas que han tenido una covid más grave tienen anticuerpos neutralizantes protectores que duran más de seis meses [...]* (<https://elpais.com/ciencia/2021-01-06/si-la-reinfeccion-de-covid-es-posible-como-sabemos-que-la-vacuna-va-a-funcionar.html>);
- (36) *Estamos bien y a orar por mi nuera Charito y por todos los que tengan este Covid-19.* (<https://www.tvazteca.com/aztecauno/venga-la-alegria/notas-vla/notas/lupita-dalessio-no-tiene-covid-19-nuera-charito-esposa-ernesto-dalessio-contagio>).

Como lo indica G. Gross (2013: 190), *tener* es el verbo soporte general de la clase de predicados nominales de estado.

### 3.3. Verbos soporte apropiados

El verbo soporte más apropiado para la clase de <enfermedades>, también las <enfermedades contagiosas>, es *padecer*:

- (37) [...] *información desagregada sobre la cantidad de mujeres, niños y hombres que padecieran el Covid [...]* (<https://www.elextremosur.com/nota/26519-comunidad-mapuche-considera-positivo-que-se-pregunte-el-origen-etnico-en-los-testeos/>).

Podemos añadir también el verbo *sufrir* que tiene a menudo un matiz durativo:

- (38) *Los asintomáticos contagian igual que los que sufren la Covid de forma evidente* (<https://www.elmundo.es/ciencia-y-salud/salud/2020/07/01/5efca8be21efa0e0728b457f.html>).

### 3.4. Extensiones aspectuales

#### 3.4.1. Extensiones aspectuales incoativas

Si hablamos de las extensiones aspectuales, podemos destacar una multitud de construcciones incoativas con los verbos *contraer*, *contagiarse*, *agarrar*, *coger* y *pillar*, el primero más formal y el último de registro coloquial:

- (39) *La aspirina a bajas dosis (75 mg) podría reducir el riesgo de contraer coronavirus (Covid-19) [...] (https://www.elsoldemexico.com.mx/doble-via/salud/aspirina-podria-reducir-riesgo-de-contraer-covid-19-segun-estudio-6479286.html);*
- (40) *Unos 74.000 enfermeros se han contagiado del Covid-19 [...] (https://www.abc.es/sociedad/abci-74000-enfermeros-contagiado-covid-19-202006161457\_video.html);*
- (41) *Según las autoridades, hay estudiantes que “juntan dinero en un bote y luego tratan de agarrar el COVID”. (https://www.dw.com/es/ee-uu-estudiantes-organizan-fiestas-covid-donde-quien-se-infecta-gana/a-54046369);*
- (42) *Así nuestros niños no cogerán ni pulmonía ni covid. (https://www.diariodesevilla.es/sevilla/alumnos-Sevilla-vuelven-colegio-mantas-abrigos-14-grados-aula\_0\_1537046820.html);*
- (43) *Porque por muy Diosa que pudiera parecer, es humana y ha pillado el COVID-19. (https://www.elperiodico.com/cuore/famosos/susi-caramelo-pi-bonexica-contagiada-pillado-84303333).*

#### 3.4.2. Extensiones aspectuales terminativas

Los verbos *curarse* y *superar* pueden considerarse como variantes terminativas:

- (44) *Crece el número de personas que se curan de COVID-19. (https://covid19mellilla.es/otras-tres-personas-se-curan-de-covid-19-y-aumenta-la-cifra-de-positivos-en-uno-mas/);*
- (45) *Un tercio de las 70.000 personas que superaron el covid lo hicieron este mes (https://www.elidealgallego.com/articulo/a-coruna/tercio-70000-personas-superaron-covid-hicieron-mes-3422314).*

#### 3.4.3. Extensiones aspectuales durativas

En las construcciones durativas, al lado del verbo *sentir* que, en algunos contextos, sobre todo con un modificador intensivo, introduce evidentemente este tipo de aspecto, aparece también *pasar*:



- (46) *Alertan del riesgo de coágulos tras pasar el Covid* ([https://www.ondacero.es/noticias/sociedad/alertan-riesgo-coagulos-pasar-covid-signos-alarma\\_202209276332bfc0502ad70001682733.html](https://www.ondacero.es/noticias/sociedad/alertan-riesgo-coagulos-pasar-covid-signos-alarma_202209276332bfc0502ad70001682733.html)).

### 3.5. Verbos soporte de acontecimiento

Los verbos de acontecimiento pueden provocar un cambio de sentido de *COVID-19* y eso concierne seguramente el verbo más general *haber*. No se trata simplemente de la existencia de la enfermedad en un lugar, sino del número de los enfermos, tanto más que se use generalmente una construcción comparativa o un determinante indefinido:

- (47) *Revilla la lía al explicar por qué hay más covid en el norte que en el sur* (<https://www.eldebate.com/sociedad/20220121/revilla-lia-explicar-hay-mas-covid-norte-sur-viento-trae.html>);
- (48) *[...] contrario de África donde no hay mucha covid [...]* (<https://hoy.com.do/la-ivermectina-de-cenicienta-a-reina/>).

*COVID-19* no tiene un carácter eventivo, pues carece de compatibilidad con los verbos de ese tipo. Lo que constituye un acontecimiento es la pandemia que muy a menudo está actualizada por el verbo soporte *haber*:

- (49) *Jared Leto, de casi morir a enterarse de que hay una pandemia por coronavirus* (<https://www.lainformacion.com/arte-cultura-y-espectaculos/coronavirus-jared-letto-se-entera-pandemia-doce-dias-despues/6552482/>).

Exactamente lo mismo concierne el verbo complejo *tener lugar*:

- (50) *Se enfatizó que en nuestros países la pandemia tiene lugar bajo sistemas de salud sumamente precarizados.* (<https://www.redeamerica.org/Noticias/Detalle/ArtMID/1370/ArticleID/2088/Ni241ez-pobreza-y-Pandemia-de-Coronavirus-191Cu225les-son-los-principales-impactos-y-desaf237os#gsc.tab=0>).

#### 3.5.1. Extensiones aspectuales de los verbos soporte de acontecimiento

Aunque no exista un verbo soporte de acontecimiento neutro que actualice *COVID-19*, podemos encontrar extensiones aspectuales.

Es, por ejemplo, el caso de *propagarse* que tiene un valor durativo e intensivo. En algunos contextos puede tratarse también de los casos positivos registrados o del brote, sin embargo, en otros se trata evidentemente de la enfermedad:

- (51) *Nueva variante de covid se propaga por EE UU en medio de un repunte de casos* (<https://es.wired.com/articulos/nueva-variante-de-covid-se-propaga-por-ee-uu-en-medio-de-un-repunte-de-casos>).

El verbo *incubar* también puede considerarse como una extensión aspectual, incoativa y durativa a la vez, que puede actualizar *COVID-19*:

- (52) *Un estudio con infectados voluntariamente muestra que el covid incuba en solo 42 horas* (<https://www.lainformacion.com/arte-cultura-y-espectaculos/coronavirus-jared-leto-se-entera-pandemia-doce-dias-despues/6552482/>).

La interpretación puede provocar algunos problemas porque, de nuevo, hay una oposición entre diversos contextos que dependen del grado de especialización del discurso. Así, lo que aceptaría un especialista, desde el punto de vista lingüístico, aparece en la primera frase del mismo artículo:

- (53) *La investigación muestra que este período es más corto, pues se creía que el coronavirus incubaba en 5 a 6 días* (<https://www.lainformacion.com/arte-cultura-y-espectaculos/coronavirus-jared-leto-se-entera-pandemia-doce-dias-despues/6552482/>).

Lo que incuba es el brote, pues *incubar* es un predicado verbal en (52). ¿Puede incubar una enfermedad? Según la RAE, que toma en cuenta sobre todo el uso general, es así, lo que muestra la definición de la segunda acepción de este verbo:

- (54) *Desarrollar una enfermedad desde que se contrae hasta que aparecen los primeros síntomas (DLE).*

A base de esta definición, podemos constatar que *incubar* es un verbo soporte en contextos generales, como (51).

Algunos verbos que parecen ser de acontecimiento de carácter intensivo no se aplican a la enfermedad como estado, sino a la situación (pandémica) aunque se usen con otros predicados nominales de la clases de <enfermedades infecciosas> como, por ejemplo, *gripe*:

- (55) *Se agrava el Covid-19 en Yucatán* (<https://yucatanahora.mx/se-agrava-el-covid-19-en-yucatan-tercer-muerto-y-75-contagiados-incluyendo-un-nino/>);
- (56) [...] *la economía va a caer a una eslabón muchísimo más bajo y más ahora cuando se agudiza el Covid-19* [...] (<https://elperiodicodemonagas.com.ve/economia/fedeagro-cosecha-de-cafe-2020-2021-retrocedio-mas-de-cien-anos/>).

Como vemos, hay una multitud de verbos que puedan actualizar *COVID-19*, pero en cada uno de los contextos hay que verificar si el nombre refiere realmente a la noción de enfermedad o se trata de un uso metonímico.

### 3.6. *COVID-19* en posición de argumento

Como admiten, entre otros, D. Le Pesant y M. Mathieu-Colas (1998: 17–18), P.-A. Buvet (1998: 99) y G. Gross (2013: 173, 350–353), los nombres predicativos pueden aparecer también en una posición de argumento de un predicado, generalmente de carácter verbal. En los casos de este tipo, el verbo actualiza el nombre, como el verbo soporte, pero, además, lo implica, por lo cual queda el núcleo predicativo de la oración. Se considera que se trata de una predicación de orden superior (o de segundo orden, según la nomenclatura de muchos lingüistas franceses), superordinado a otro predicado, el de primer orden. La diferencia entre los argumentos objetos (de tipo *coronavirus* en su sentido principal) y los predicados en posición de argumento (de tipo *COVID-19*) consiste en que los segundos refieren a un concepto abstracto. Los predicados de primer orden implican sus propios argumentos y la estructura está subordinada a la del predicado de orden superior.

La multitud de contextos posibles no nos permite analizar las estructuras de una manera exhaustiva. Sin embargo, queremos dar unos ejemplos para indicar las diferencias entre las estructuras con verbo soporte y las con predicados de orden superior.

Podemos notar que si *curarse* se considera como verbo soporte aspectual terminativo porque el predicado nominal *COVID-19* guarda su argumento sujeto, *curar* implica sus propios argumentos y uno es el esquema predicativo de un predicado de la clase de <enfermedades>:

- (57) [...] *la medicación lo ha curado de la COVID* [...] (<https://nuevodiario.es/noticia/13301/andalucia/entrevista-a-rogelio-mena-psoe-albox-no-se-me-rece-el-penoso-espectaculo-de-torrecillas-al-frente-de-la-corporacion.html>).

Pero en la mayoría de los casos, se trata de verbos que no tienen ninguna relación formal con los verbos soporte:

- (58) *Muere por covid el segundo teniente de alcalde de Utrera [...]* (<https://www.lavanguardia.com/local/sevilla/20210528/7489945/muere-covid-segundo-teniente-alcalde-utrera-jose-manuel-doblado-lara.html>);
- (59) *Guatemala regala una bolsa de arroz a quienes se vacunan del coronavirus* (<https://www.periodicodeibiza.es/etiqueta/campa%C3%B1a+de+vacunaci%C3%B3n.html>).

Generalmente, el predicado de orden superior pertenece a la clase de verbos, sin embargo, puede tratarse también de otro predicado nominal:

- (60) *Un fuerte dolor de garganta es el síntoma que más caracteriza a las nuevas infecciones de covid-diecinueve.* (<https://www.diariosigno.com/el-nuevo-sintoma-del-covid-diecinueve/>).

Claro está, el predicado de primer orden puede aparecer también en una estructura con verbo soporte:

- (61) *Lupita D'Alessio aclara que no tiene Covid-19 [...]* (<https://www.tvazteca.com/aztecauno/venga-la-alegria/notas-vla/notas/lupita-dalessio-no-tiene-covid-19-nuera-charito-esposa-ernesto-dalessio-contagio>).

## 4. Conclusiones

La vida de la terminología especializada en contextos no especializados no es fácil. Si la terminología está sujeta a la normalización, muy a menudo internacionalizada, y los especialistas aprecian la univocidad de sus elementos, la lengua general tiende a evolucionar continuamente y en cada uno de los actos de comunicación interactúan sujetos con diversos niveles de experiencia y conocimiento de dominios especializados. Y si la verificación científica requiere mucho tiempo y mucha paciencia, la mayoría de la gente, y sobre todo los periodistas, quiere y precisa intercambiar la información en seguida.

De un lado, la escasez y, de otro, la complejidad de la información especializada provocan que se cree un léxico popular, muy a menudo a base del especia-

lizado. Podemos hablar de una *sinonimitis*, el término propuesto por A. Kacprzak y J.-P. Goudaillier (2014) para indicar la sobreabundancia de sinónimos, esencialmente si tomamos en consideración los diversos niveles de la lengua.

Podemos destacar la aparición de la variación gráfico-fonética y de la metonimia. Sobre todo, el segundo fenómeno puede causar unas ciertas confusiones porque la causa puede referir a la consecuencia y al revés. En el texto hemos podido observar una metonimia recíproca de *COVID-19* y *coronavirus* en el discurso de los periodistas. Ambos significantes pueden referir tanto a la enfermedad, como al brote y, además, a los enfermos, a una o varias olas de infecciones o bien a la situación pandémica en general. ¿Cómo reconocer el sentido si el lenguaje general carece de precisión? Nos ayuda el contexto y sobre todo la actualización. Tanto el uso de la determinación como el de los verbos soporte dependen del contenido temporal y aspectual de la oración, que tienen que estar compatibles con el concepto expresado por el predicado.

En el presente estudio, nos hemos concentrado en los contextos en los que *COVID-19*, así como sus variantes y sinónimos, pertenece a la clase de predicados de <enfermedades infecciosas>. Hemos analizado su compatibilidad con los determinantes y modificadores, tomando en cuenta que la modificación es, en realidad, un fenómeno infinito y que algunos modificadores pueden juntarse al nombre y formar con él una unidad semántica formalmente compleja. Hemos presentado también los verbos soporte, neutros y aspectuales. Todo este conjunto de actualizadores puede confirmar la pertenencia de un nombre a la misma clase, por ello se debería comprobar, en los pasos siguientes, los contextos en los que aparecen otras enfermedades consideradas como infecciosas.

Queremos subrayar que, como ya lo hemos indicado en la introducción, no se trata de un análisis cuantitativo o cualitativo. Sin embargo, el corpus recogido nos ha permitido ejemplificar la combinatoria sintáctica del nombre predicativo *COVID-19* (y sus sinónimos y variantes).

Hay que tomar en cuenta que hemos analizado textos de un nivel de especialización relativamente bajo, por ello el estudio de textos científicos, pero también el de las redes sociales, podrían llevar a resultados diferentes, sobre todo en lo que se refiriera a las preferencias de uso de uno u otro verbo soporte.

## Referencias bibliográficas


Blanco, X., & Buvet, P.-A. (1999). À propos de la traduction automatique des déterminants de l'espagnol et du français. *Meta: Journal des traducteurs*, 44(4), 525–545.

- Buvet, P.-A. (1998). Détermination et classes d'objets. *Langages*, 131, 91–102.
- Comezana, G. P. (2022). *El verbo hacer del español en la interfaz léxico-sintaxis. Construcciones de hacer con complemento nominal* [Tesis de doctorado, Universidad de Buenos Aires]. Repositorio Institucional de la Facultad de Filosofía y Letras, UBA. [http://repositorio.filo.uba.ar/bitstream/handle/filodigital/16188/uba\\_ffyl\\_t\\_2022\\_se\\_Comezana%20C3%B1a.pdf?sequence=1&isAllowed=y](http://repositorio.filo.uba.ar/bitstream/handle/filodigital/16188/uba_ffyl_t_2022_se_Comezana%20C3%B1a.pdf?sequence=1&isAllowed=y).
- Gross, G. (1996). *Les expressions figées en français : Les noms composés et autres locutions*. Ophrys.
- Gross, G. (2008). Les classes d'objets. *Lalies*, 28, 111–165. <https://shs.hal.science/halshs-00410784>.
- Gross, G. (2013). *Manual de análisis lingüística: Aproximación syntactico-semántica al léxico* (X. Blanco Escoda, Trad.). Editorial UOC.
- Hallin, D. C., Briggs, Ch. L., Mantini-Briggs, C., Spinelli, H., & Sy, A. (2020). Mediatización de las epidemias: la cobertura sobre la pandemia de la gripe A (H1N1) de 2009 en Argentina, Estados Unidos y Venezuela. *Comunicación y Sociedad*, 17, 1–24. <https://doi.org/10.32870/cys.v2020.7207>.
- Kacprzak, A., & Goudaillier, J.-P. (2014). Dénominations des maladies en langue populaire et argotique (de la 'synonymite' des noms de maladies). *E-Scripta Romanica*, 1, 1–8. <https://doi.org/10.18778/2392-0718.01.01>.
- Le Pesant, D., & Mathieu-Colas, M. (1998). Introduction aux classes d'objets. *Langages*, 131, 6–33.
- Real Academia Española (2013-). *Diccionario histórico de la lengua española (DHLE)*. <https://www.rae.es/dhle/>.
- Real Academia Española (2022). *Diccionario de la lengua española (DLE)*. <https://dle.rae.es/>.



*Beata Śmigielska*

Université de Silésie à Katowice  
Pologne

 <https://orcid.org/0000-0002-3383-0030>

# Traduction automatique et désambiguïsation des sens des mots. Le cas du verbe français *louer*

**Automatic translation and disambiguation of word meanings.  
The case of the French verb *louer***

## **Abstract**

Machine translators such as those offered by Google, Deep, ChatGPT and others are already using NMT systems. They apply the techniques of artificial intelligence that are constantly being developed, so they are currently providing increasingly correct and suitable translations.

In the first part of this article, the author examines the case of the French verb *louer* by having it translated in different contexts by three automatic translators mentioned to see how effective they currently are as translation tools. Next, the verb *louer* is presented and analyzed as seen through the prism of the Antidote 8 dictionary. The last part of the article is devoted to the French-Polish description of the verb studied using the Object-Oriented Approach, with the aim of disambiguating it and developing the system of syntactico-semantic rules (object classes, frames, scripts) enabling the appropriate automatic translation of sentences in which it may appear.

## **Keywords**

Disambiguation, object-oriented approach, supervised automatic translation, object classes, frames, scripts

## I Introduction

Il est indéniable que nous vivons à une époque intéressante, marquée par un développement croissant de l'intelligence artificielle (IA) au cours des dernières années, dont, entre autres, la traduction automatique. Ses débuts remontent aux années 60 du XX<sup>e</sup> siècle où les systèmes de traduction s'appuyaient généralement sur les règles de correspondance entre des langues. Ce type de description contrastive fonctionnait déjà bien dans différents dictionnaires et glossaires en mettant en relation les mots et les expressions de la langue source et ceux de la langue cible. Il était donc électronisé et utilisé avec succès pendant de nombreuses années, même si l'introduction de ses données lexicales exigeait énormément de temps (toutes ses données étaient introduites une par une à la main) et de ressources financières importantes. Ensuite, dans les années 90, avec le temps d'Internet, des processeurs plus rapides et des disques d'ordinateur de plus grande capacité, vient la méthode des traductions automatiques statistique (TAS : *Statistical Machine Translation* [SMT]). Elle s'est avérée beaucoup plus opérante par rapport à la précédente quant à l'efficacité de la traduction, moins chère, et son application nécessitait moins d'effort et de temps. Dans le processus de la préparation de la traduction, le système opérait sur des textes parallèles bilingues ou multilingues et sur leurs relations mutuelles de traduction. Et enfin, dans les dernières années, avec un progrès important de l'IA est apparu le modèle de traduction neuronale (TAN ou NMT : *Neural Machine Translation*). Il tente d'imiter celui du cerveau humain et est aujourd'hui utilisé naturellement non seulement dans la traduction automatisée, mais aussi dans de nombreux domaines (p. ex. reconnaissance faciale, génération d'images à partir de descriptions textuelles, simulations médicales, découverte de nouveaux médicaments, analyse de grandes bases de données, conduite autonome de véhicules routiers, etc.). Pour pouvoir traduire d'une langue à une autre et pour que cette traduction soit de bonne qualité, les systèmes de TAN analysent des milliards d'exemples de traduction à base des corpus bilingues et ces analyses-là les conduisent à « apprendre » petit à petit une traduction appropriée en établissant des liens entre les mots, expressions et/ou phrases correspondants. Les réseaux neuronaux sont constamment entraînés à travers l'apprentissage en profondeur des masses de données dans le but d'améliorer continuellement leur performance. Finalement, l'écart entre le résultat obtenu par les réseaux neuronaux et le résultat attendu est évalué (soit il est accepté, soit, s'il ne l'est pas, les paramètres internes du réseau neuronal sont modifiés) (cf. p. ex. : Allen, 2003 ; Hutchins & Somers, 1992 ; Rubino, 2011 ; Trinth, 2019).



## II Traduction machine vs traduction humaine

La traduction automatique a fait et continue à faire l'objet d'un très grand nombre de différents travaux scientifiques. D'une part, il y a p. ex., en linguistique, des méthodes lexicographiques qui proposent des descriptions lexicales dans le but de perfectionner le traitement automatique des textes, dont, entre autres, pour ne citer que juste quelques-unes, le Sens-Texte de I. Melčuk (cf. p. ex. : 1997, 2012, 2016, etc. ; cf. aussi dans p. ex. : Marengo, 2021), le modèle du Lexique Génératif de J. Pustejovski (1995), l'Approche Orientée Objets de W. Banyś (cf. p. ex. : 2002a, 2002b, 2005, 2018 ; cf. aussi dans p. ex. : Czekaj 2018 ; Hrabia, 2016, 2018 ; Perz, 2013 ; Śmigielska, 2013, 2015, 2016, 2017, 2020 ; Żłobińska-Nowak, 2011a, 2011b, 2013, etc.), la méthode des Classes d'Objets de G. Gross (p. ex. 1994, 1997, 2008, 2012) et beaucoup d'autres. Les approches mentionnées ci-dessus proposent des règles de description sémantique et lexicale dans le but de résoudre tous les problèmes résultant du caractère polysémique de la langue ce qui pourrait conduire à éliminer des erreurs de traduction, surtout celles de nature sémantique. La structure de certaines descriptions linguistiques est élaborée pour être aisément intégrée aux systèmes de traduction automatique. Elles peuvent servir aussi, dans beaucoup de cas, à entraîner les réseaux neuronaux. Étant donné que les systèmes de TAN sont déjà en service des traducteurs automatiques, tels que p. ex. ceux qui sont proposés par Google, DeepL, ChatGPT et d'autres, et qu'ils basent sur les techniques de l'IA qui sont en évolution permanente, ils fournissent actuellement des traductions de plus en plus correctes.

D'autre part, certains auteurs notent qu'on a relativement peu mis l'accent sur l'intégration de prérequis linguistiques, surtout sémantico-lexicaux, dans la traduction automatique neuronale. De nombreux travaux antérieurs se sont surtout concentrés sur l'incorporation des connaissances linguistiques morphologiques et syntaxiques du côté de la langue source (cf. p. ex. Eriguchi *et al.*, 2017 ; Shi *et al.*, 2016 ; Tan *et al.*, 2020 ; Wu *et al.*, 2016)

L'un des problèmes clé de la traduction automatique est, mentionnée ci-dessus, la désambiguïsation des sens des mots à traduire (cf. p. ex. la présentation des derniers développements dans Bevilacqua *et al.* 2021 ; Tyagi et al 2022). C'est un défi très complexe pour le traitement automatique des langues. Les programmes de traduction automatique utilisant l'IA, basés sur les modèles GPT (*Generative-Pretrained-Transformer*), un type d'apprentissage profond qui se sert d'un réseau neuronal fondé sur les transformateurs, met en œuvre plusieurs techniques pour gérer la désambiguïsation lors de la traduction. C'est surtout la compréhension contextuelle des mots et des phrases par exploration contextuelle. Les modèles IA

prennent en compte le contexte pour déterminer le sens approprié d'un mot et ses relations avec d'autres mots dans le texte. Cette compréhension contextuelle est possible maintenant grâce à une propagation et une utilisation de plus en plus intense de l'apprentissage profond. Les réseaux neuronaux profonds utilisés dans les modèles IA ont la capacité de trouver des motifs et des relations complexes et de déterminer différentes significations des mots en fonction de leur utilisation dans différents contextes. L'apprentissage profond rend aussi possible un entraînement de la compréhension contextuelle sur de très vastes ensembles de données, des milliards d'exemples, contenant une grande variété d'utilisations de mots dans différents contextes. Cet entraînement conduit les systèmes à « apprendre » petit à petit une traduction appropriée en établissant des liens entre les mots, les expressions et/ou les phrases correspondants. Le succès de l'IA du type GPT était possible grâce à un changement de paradigme dans la création des réseaux neuronaux. Comme le dit le titre de l'article qui a commencé la révolution : « L'attention est tout ce dont nous avons besoin » (Vaswani *et al.*, 2017). Les mécanismes d'attention qui sont à la base des architectures de l'IA comme Transformer permettent au modèle de se concentrer sur des parties spécifiques du texte de départ lors de la génération de la traduction. Cela aide à déterminer les parties les plus pertinentes pour désambiguïser le sens d'un mot.

D'autre part, les systèmes IA de traduction automatique ont recours aux règles heuristiques. Dans les cas où le modèle est incertain quant à la traduction, il devrait pouvoir recourir à des règles de ce type pour choisir une traduction plus générale et moins susceptible d'être incorrecte. Cela devrait garantir une traduction plus sûre, même si elle n'est pas la plus précise ou nuancée (dans notre approche appliquée ci-dessous, ces règles sont appelées préférentielles, mais elles sont naturellement fondées sur des heuristiques du type décrit ci-dessous). Mais, là encore, il est bien qu'il y ait une rétroaction de la part des utilisateurs. Certains systèmes se servent déjà des retours utilisateurs pour améliorer les traductions au fil du temps (p. ex. DeepL). Les corrections ou les évaluations des utilisateurs sont particulièrement importantes et servent à perfectionner la désambiguïstation et les choix de traduction. Il est important à noter que, malgré ces techniques et des progrès extraordinaires et fascinantes dans la qualité de la traduction automatique grâce à l'IA, la désambiguïstation reste toujours un vrai défi en traduction automatique, et il peut y avoir des cas où les modèles peuvent encore commettre des erreurs ou ne pas comprendre correctement le contexte.

Les analyses du type présenté ci-dessous sont un exemple de la contribution linguistique et de la réalisation des deux derniers points mentionnés ci-dessus :

recours aux règles heuristiques/préférentielles et retours des évaluateurs pour rendre meilleures les traductions automatiques proposées. D'autant plus qu'une plus grande intégration des données linguistiques sémantiques dans des systèmes de traduction neuronale améliorerait certainement encore plus la performance de ces systèmes, en particulier en ce qui concerne la désambiguïsation du sens des mots et des phrases entières, et par conséquent fournirait une traduction appropriée.

C'est à un cas particulier, mais pas du tout isolé, de ce type de situation où l'information sémantique précise permettrait de désambiguïser d'une manière efficace le sens des mots et de proposer une traduction appropriée que les systèmes de traduction automatique analysés n'effectuent pas toujours, que cet article est consacré. Dans sa première partie, nous examinerons le cas du verbe *louer* en le faisant traduire dans différents contextes par trois traducteurs automatiques : Google Traduction, DeepL, ChatGPT pour voir à quel point ils sont actuellement efficaces ou non comme outils de traduction. Ensuite, dans la partie suivante, nous analyserons comment le verbe en question est décrit par le dictionnaire d'Antidote 8, pour arriver à la fin, à la proposition des schémas de description de *louer* selon l'Approche Orientée Objets (AOO) en vue de la traduction automatique des textes (cf. Banyś 2002a, 2002b, 2005).

### III Verbe *louer* et ses traductions dans les traducteurs automatiques choisis

Selon p. ex. le dictionnaire français Le Robert Dico en Ligne, le verbe *louer* est défini comme suit (<https://dictionnaire.lerobert.com/definition/louer>, consulté en 08.2023) :

Définition de *louer* verbe transitif :

1. Déclarer (qqn ou qqch.) digne d'admiration ou de très grande estime. → exalter, louanger. *Louer qqn sans mesure.* → encenser, flatter.
2. *Louer qqn de ou pour qqch.* → féliciter.
3. *Louer Dieu, le Seigneur.* → bénir, glorifier. locution *Dieu soit loué!*, exclamation de joie, de soulagement.

### Définition de *louer* verbe transitif :

I Donner (qqch.) en location. *Louer une chambre meublée à un étudiant.*

→ aussi sous-louer. *Maison à louer.*

– pronominal (passif) *Cet appartement doit se louer cher.*

### II

1. Prendre en location, à bail. *Louer un appartement*, en être locataire.

2. Réserver, retenir en payant. *Louer sa place dans un train, un avion ; au spectacle.*

D'après la définition ci-dessous, on peut dégager en polonais deux sens principaux correspondant au verbe *louer* :

1. chwalić/pochwalić
2. wynajmować/wynająć

Vu les acceptions du verbe *louer*, il est bien visible qu'il appartient à la catégorie des homonymes (il y a deux colonnes séparées dans le dictionnaire cité ci-dessus). Le premier sens distingué vient de l'étymon latin « laudare », tandis que l'autre de « locare ». Les deux sens n'ont pas donc un invariant sémantique en commun, ce qui devrait encore plus faciliter la tâche aux traducteurs automatiques (la vérité est que plus les sens des mots sont sémantiquement proches, plus il est difficile de les bien traduire d'une langue à l'autre) (cf. p. ex. Śmigielska, 2015). Cependant, ce qui est le plus important dans nos recherches, ce n'est pas le statut linguistique du verbe analysé, mais l'efficacité des traductions automatiques des phrases où il apparaît. Le fait si un mot donné appartient aux polysèmes ou aux homonymes n'a rien à avoir pour la machine. Ce qui compte, c'est que la même forme possède plusieurs sens et il faut en choisir le meilleur équivalent dans la langue cible pendant le processus de traduction.

Voyons donc ci-dessous, pour commencer, si les traductions des phrases du français vers le polonais comprenant l'expression *louer la/sa femme* (*louer + complément d'objet animé humain*) proposées par Google Traduction, DeepL et ChatGPT sont correctes (les traductions avec un ou deux points d'interrogation, ne semblent pas, de notre point de vue, à des degrés différents, sémantiquement acceptables dans le contexte donné).

Comme on va le voir dans ce qui suit, les traductions obtenues sont souvent fonction du temps et de la version du traducteur. Celles qu'on présente ci-dessous ont été collectées en septembre 2023.

1. *Il a loué la femme.*

DeepL, Google

Pochwalił kobietę.

ChatGPT 3.5 et 4

Wynajął kobietę.

2. *Il a loué la femme pour faire le ménage.*

DeepL

⁉Zatrudnił kobietę do prac domowych.

Google

⁉Zatrudnił kobietę do sprzątanía.

ChatGPT 3.5

Wynajął kobietę do wykonania sprzątanía.

ChatGPT 4

Wynajął kobietę do sprzątanía.

3. *Il a loué la femme pour une nuit.*

DeepL, Google, ChatGPT 3.5 et 4

Wynajął kobietę na jedną noc.

4. *Le mari loue sa femme pour la nuit qu'ils ont passée ensemble.*

DeepL

⁉⁉Mąż wynajmuje żonę na noc, którą spędzają razem.

DeepL Pro (version 2023 for Mac)

Mąż chwali żonę za wspólnie spędzoną noc.

Google

Mąż chwali żonę za wspólnie spędzoną noc.

ChatGPT 3.5 et 4

Mąż chwali swoją żonę za noc, którą spędzili razem.

5. *Le mari loue sa femme pour une belle nuit passée ensemble.*

DeepL

⁉⁉Mąż wynajmuje swoją żonę na noc.

DeepL Pro (version 2023 for Mac)

Mąż chwali żonę za piękną noc spędzoną razem.

Google, ChatGPT 3.5 et 4

Mąż chwali swoją żonę za piękną spędzoną razem noc.

6. *Il a loué la femme parce qu'elle a préparé un bon gâteau.*

DeepL, DeepL Pro (version 2023 for Mac)

Pochwalił kobietę, ponieważ upiekła dobre ciasto.

Google

Pochwalił kobietę za to, że upiekła dobre ciasto.

ChatGPT 3.5

Wynajął kobietę, ponieważ przygotowała smaczny tort.

ChatGPT 4

Pochwalił kobietę, bo przygotowała dobry tort.\*

\*On laisse de côté ici la traduction différente de *un bon gâteau* dans les deux versions du ChatGPT par rapport aux traductions de DeepL et Google.

7. *Il a loué sa femme pour le repas/pour un bon repas.*

DeepL

Wynajął swoją żonę na posiłek. / Pochwalił swoją żonę za dobry posiłek.

DeepL Pro (version 2023 for Mac)

Wynajął swoją żonę na posiłek. / Pochwalił swoją żonę za dobry posiłek.

Google

Wynajął swoją żonę na posiłek. / Pochwalił swoją żonę za dobry posiłek.

ChatGPT 3.5

Wynajął swoją żonę na posiłek. / Wynajął swoją żonę na dobre jedzenie.

ChatGPT 4

Pochwalił swoją żonę za posiłek. / Wynajął swoją żonę za dobry posiłek.

8. *Il a loué sa femme pour le repas qu'elle avait préparé pour la famille.*

DeepL

Pochwalił swoją żonę za posiłek, który przygotowała dla rodziny.

Google

Pochwalił swoją żonę za posiłek, który przygotowała dla rodziny.

ChatGPT 3.5

Wynajął swoją żonę na posiłek, który przygotowała dla rodziny.

*ChatGPT 4*

Pochwalił swoją żonę za posiłek, który przygotowała dla rodziny.

9. *Il a loué sa femme pour le repas. Il en était très content parce que ce repas était le meilleur qu'il ait jamais mangé de sa vie.*

*DeepL, DeepL Pro (version 2023 for Mac)*

Pochwalił swoją żonę za posiłek. Był bardzo szczęśliwy, ponieważ był to najlepszy posiłek, jaki kiedykolwiek jadł w swoim życiu.

*Google*

⁹Zatrudnił żonę do posiłku. Bardzo się z tego ucieszył, bo ten posiłek był najlepszym, jaki kiedykolwiek jadł w życiu.

*ChatGPT 3.5*

⁹Wynajął swoją żonę na posiłek. Był bardzo zadowolony, ponieważ ten posiłek był najlepszy, jaki kiedykolwiek w swoim życiu zjadł.

*ChatGPT 4*

Pochwalił swoją żonę za posiłek. Był bardzo zadowolony, ponieważ ten posiłek był najlepszym, jaki kiedykolwiek jadł w swoim życiu.

10. *Le mari loue sa femme en secret.*

*DeepL*

Mąż potajemnie chwali swoją żonę.

*DeepL Pro (version 2023 for Mac)*

Mąż potajemnie wynajmuje swoją żonę.

*Google*

Mąż w tajemnicy chwali żonę.

*ChatGPT 3.5 et 4*

Mąż w tajemnicy chwali swoją żonę.

11. *Le mari loue sa femme en secret, car il veut cacher cette situation honteuse devant la famille.*

*DeepL*

Mąż chwali swoją żonę w tajemnicy, ponieważ chce ukryć tę haniebną sytuację przed rodziną.

*Google*

Mąż w tajemnicy chwali żonę, bo chce ukryć tę haniebną sytuację.

*ChatGPT 3.5 et 4*

Mąż wynajmuje swoją żonę w tajemnicy, ponieważ chce ukryć tę haniebną sytuację przed rodziną.

12. *Le mari loue sa femme en secret. Il veut qu'elle ne le dise à personne. Ce qui importe, c'est de l'argent qu'elle gagne comme ça.*

*DeepL*

Mąż chwali swoją żonę w tajemnicy. Nie chce, by komukolwiek o tym mówiła. Liczą się pieniądze, które w ten sposób zarabia.

*Google*

Mąż w tajemnicy chwali żonę. Nie chce, żeby komukolwiek mówiła. Ważne są pieniądze, które w ten sposób zarabia.

*ChatGPT 3.5*

Mąż wynajmuje swoją żonę w tajemnicy. Chce, żeby nikomu o tym nie mówiła. To, co ma znaczenie, to pieniądze, które w ten sposób zarabia.

*ChatGPT 4*

Mąż chwali swoją żonę w tajemnicy. Nie chce, by ktokolwiek o tym wiedział. Liczy się dla niego pieniądze, które w ten sposób zarabia.

Après avoir analysé les traductions automatiques ci-dessus, on peut tirer quelques conclusions intéressantes.

Il n'est pas surprenant que l'exemple 1, où l'expression *louer la femme* n'est accompagnée d'aucun contexte plus large, soit traduit de deux façons correctes.

La phrase 2, qui est plus riche en sens par rapport à la première, reçoit une traduction plus appropriée dans le cas des deux versions du ChatGPT, tandis que les deux autres traducteurs ont proposé l'équivalent polonais *zatrudnił* à la place de *wynajął*, et c'est ce dernier qui y est plus convenable. Même si ces deux verbes polonais sont sémantiquement proches, ils diffèrent par leur sens, et ils ne sont pas automatiquement interchangeables dans tous les contextes, même si c'est le cas ici, si l'on fait abstraction de différence de sens et on se concentre sur la situation référentielle décrite similaire : d'une manière ou d'une autre, la femme travaille pour lui. Il est évident aussi que le verbe *louer* – *wynajmować/wynająć* n'a pas le même sens que celui d'*embaucher* – *zatrudniać/zatrudnić*. Alors, la traduction de la phrase 2 *zatrudnił kobietę* est moins correcte de ce point de vue. Il faut ajouter aussi que si la suite après *pour* plaçait la situation au passé, c'est la traduction *pochwalił*, et non pas *wynajął* qui s'imposerait :

2a. *Il a loué la femme pour avoir fait le ménage.*

Et c'est ce que font le DeepL et le Google Traduction ci-dessous, mais non pas, bizarrement, le ChatGPT 4 :



*DeepL:*

Pochwalił kobietę za wykonywanie prac domowych.

*Google:*

Pochwalił kobietę za sprzątanie.

*Chat GPT 4:*

?Wynajął kobietę do sprzątania.

Cela nous invite aussi à analyser, naturellement pas ici, faute de place et de temps, la sensibilité du ChatGPT 4 aux relations temporelles lors de la traduction, d'autant plus que la traduction de la même phrase avec l'ajout de *hier* est correcte: *wynajął* et non pas *pochwalił*:

2b. *Il a loué la femme pour faire le ménage hier.*

Wynajął kobietę do sprzątania wczoraj.

La phrase 3 est bien traitée, préférentiellement dans ce contexte très limité (dans un contexte plus large, où certaines informations sont à trouver dans les phrases précédentes et/ou suivantes, la phrase étant aussi très elliptique, la traduction par *chwalić* serait aussi possible), par tous les trois traducteurs.

Le DeepL pour Windows, à la différence de la version pour Mac 2023, ne s'est pas bien débrouillé avec les phrases 4 et 5.

En revanche, le ChatGPT 3.5 n'a pas bien traduit la phrase 6, préférentiellement toujours, dans ce contexte très limité – dans un contexte plus large, où certaines informations sont à trouver dans les phrases précédentes et/ou suivantes, la phrase étant aussi très elliptique, la traduction par *wynając* serait pourtant possible, la femme étant louée pour faire *un bon gâteau*, vu qu'elle a déjà fait pour une autre famille un excellent gâteau d'anniversaire. La question de sensibilité aux relations temporelles apparaît de nouveau.

L'exemple 7 n'est bien traité que par le ChatGPT 4. Ce qui est intéressant dans cet exemple de phrase, c'est qu'au moment où un élément de plus a été ajouté à la phrase de départ (*pour le repas* a été remplacé par *pour un bon repas*), les propositions de traduction du DeepL et du Google se sont avérées correctes, pendant que celle du ChatGPT 3.5, à la différence du ChatGPT 4, ne l'était toujours pas.

La phrase 8 représente le même cas où c'est seulement le ChatGPT 3.5 n'a pas trouvé d'équivalent polonais approprié.

La traduction de l'exemple 9 est un peu étonnante, parce que cette fois-ci seules les traductions proposées par le DeepL et le ChatGPT 4 sont réussies, même si le contexte de phrases avec *louer sa femme* est plus large par rapport à l'exemple précédent.

La phrase 10, courte et ambiguë, est traduite de la même façon – *chwali* – par tous les traducteurs, sauf le DeepL pour Mac 2023, où *wynajmuje* a été choisi, et tous les deux équivalents polonais proposés sont correctes, vu le manque d'un contexte plus large.

On pourrait s'attendre à ce que les phrases 11 et 12, où l'expression *louer sa femme en secret* apparaît dans le contexte plus large par rapport à l'exemple précédent, ne créent pas trop de divergences dans les traductions, mais on voit bien que ce n'est pas le cas, le contexte n'étant visiblement pas encore suffisamment précis, et les deux équivalents polonais possibles entrent en jeu et il est intéressant de voir ici le choix différent des traductions entre les traducteurs, en particulier les deux versions du ChatGPT. Faute d'un contexte plus large, dans les deux cas le fait que *le mari loue* – dans les deux sens en polonais: *chwali* et *wynajmuje* – *sa femme en secret* pourrait être lié au fait que l'argent qu'elle gagne est gagné d'une manière particulière, peut-être non acceptée, de la perspective du mari, par tout le monde. On doit avoir un contexte, un cadre et/ou un script, encore plus large pour décider de la traduction appropriée.

Et encore, ci-dessous, les traductions des phrases comprenant l'expression *louer la maison, louer la voiture (louer + complément d'objet inanimé concret)* et *louer le style de la maison (louer + complément d'objet abstrait)* :

13. *Le mari loue le style de la maison que sa femme a décidé de louer.*

*DeepL (les deux versions), Google*

Mąż wynajmuje dom w stylu, na jaki zdecydowała się jego żona.

*ChatGPT 3.5 et 4*

Mąż chwali styl domu, który jego żona zdecydowała się wynająć

14. *Le mari a loué la maison que sa femme avait décidé de louer de ses voisins. Il l'a louée pour sa beauté et son style.*

*DeepL*

Mąż wynajął dom, który jego żona zdecydowała się wynająć od sąsiadów.

Wynajął go ze względu na jego piękno i styl.

*Google*

Mąż wynajął dom, który jego żona zdecydowała się wynająć od sąsiadów.

?Chwalił ją za urodę i styl.

*ChatGPT 3.5*

Mąż wynajął dom, który jego żona zdecydowała się wynająć od swoich sąsiadów. Wynajął go ze względu na jego piękno i styl.

*ChatGPT 4*

Mąż wynajął dom, który jego żona postanowiła wynająć od sąsiadów.  
Chwalił go za jego piękno i styl.

15. *Le mari a loué cette voiture pour son économie d'utilisation.*

*DeepL*

?Mąż wziął ten samochód w leasing ze względu na ekonomię użytkowania.

*DeepL pour Mac 2023*

Mąż wynajął ten samochód ze względu na jego ekonomiczną eksploatację

*Google*

Mąż wynajął ten samochód ze względu na oszczędność użytkownika.

*ChatGPT*

Mąż wynajął ten samochód ze względu na oszczędność użytkownika.

*ChatGPT 4*

Mąż wynajął ten samochód ze względu na jego ekonomiczność w użytkowaniu.

À la base des phrases ci-dessus, on peut observer que même l'introduction dans le contexte des éléments (*louer la maison, louer la voiture (louer + complément d'objet inanimé concret), louer le style de la maison (louer + complément d'objet abstrait)*) et l'ajout de *pour*, qui sembleraient suggérer que c'est plutôt *chwalić* qui serait le bon équivalent de *louer* dans les contextes cités, ce n'est pas le cas, parce que *wynajmować* y va aussi très bien.

Il est intéressant de remarquer que, toutes choses égales par ailleurs, les deux versions du ChatGPT ont différentes préférences quant à la traduction de la seconde phrase de l'exemple 14 : les deux propositions étant correctes, le ChatGPT 4 semble tendre à préférer à ne pas répéter le même mot (*wynajmować*) qui apparaît déjà dans la phrase précédente, ce qui peut suggérer que le maintien d'une certaine congruité du texte a été pris en considération.

Les analyses des exemples ci-dessus montrent que les traducteurs automatiques actuels sont déjà capables de donner généralement de très bonnes traductions, mais il arrive parfois, et la quantité d'oscillations de traductions dépend encore du traducteur choisi, que certaines traductions laissent encore à désirer.

Il nous donc paraît important et opportun de proposer une description d'unités linguistiques, p. ex. grâce à la Méthode Orientée Objets de W. Banyś (cf. 2002a,

2002b, 2005), pour fournir aux systèmes neuronaux des données linguistiques élaborées qui peuvent constituer un complément pour l'apprentissage profond qu'ils réalisent.

#### IV Verbe *louer* dans le dictionnaire

Un point de départ pour le faire à ne pas négliger est de voir comment les dictionnaires traitent le verbe *louer*. Dans ce but, nous avons consulté, entre autres, le dictionnaire d'Antidote 8. Sa façon de présenter les articles nous a paru particulièrement intéressante, surtout la partie consacrée aux cooccurrences des mots. Le dictionnaire les représente regroupées en catégories autour de chaque sens d'un mot donné.

Après avoir analysé les cooccurrences du verbe *louer*, il est possible de tirer les conclusions suivantes (tous les exemples cités ci-dessous sont tirés d'Antidote 8) :

1. *louer* suivi de complément d'objet direct du type *inanimé abstrait* = *approuver* (chwalić, pochwalić), p. ex. :
  - louer la qualité*
  - louer la vertu*
  - louer la beauté*
  - louer le courage*
  - louer les mérites*
  - louer la sagesse*
  - etc.
2. *louer* suivi de complément d'objet direct du type *animé humain* = *approuver* (chwalić, pochwalić), p. ex. :
  - louer le prince*
3. *louer* suivi de complément d'objet direct du type *animé* = *approuver* (chwalić), p. ex. :
  - louer le Seigneur*
  - louer le Créateur*
  - louer le Dieu*
4. *louer* suivi de complément d'objet direct du type *inanimé concret* = *donner en location* (wynajmować, wynająć), p. ex. :
  - louer une voiture*
  - louer un appartement*
  - louer une maison*

*louer une villa*  
*louer un local*  
*louer un studio*  
 etc.

5. *louer* suivi de complément d'objet direct du type *inanimé concret* = *réserver* (rezerwować, zarezerwować), p. ex. :

*louer une voiture*  
*louer un appartement*  
*louer une maison*  
*louer une villa*  
*louer un local*  
*louer un studio,*  
 etc.

6. *louer* suivi de complément d'objet direct du type *animé humain* = *embaucher* (zatrudniać, zatrudnić), p. ex. :

*louer les pleureuses*  
*louer les services.*

Ce qui est intéressant de voir à travers toutes les cooccurrences de *louer*, c'est que, d'une part, le dictionnaire ne prend pas en compte le fait que *les inanimés concrets* (p. ex. *maison, voiture, appartement, livre*, etc.) peuvent être aussi bien *loués* au sens d'*approuver, d'apprécier, complimenter* (chwalić, pochwalić) que de *donner en location* (wynajmować, wynająć), et d'autre part, que *les animés humains* (p. ex. *femme, homme, frère, prisonnier, boulanger*, etc.) peuvent être *loués* aussi bien au sens d'*approuver, complimenter* (chwalić, pochwalić) (cf. IV 2 – il y a un seul exemple de ce type d'emploi dans le dictionnaire, comme s'il était complètement marginal dans la langue, tandis que, en réalité, il ne l'est pas), qu'au sens d'*embaucher* (wynajmować, wynająć, zatrudniać, zatrudnić) (le dictionnaire ne donne que deux exemples de ce type d'emploi). On pourrait se demander aussi pourquoi, là où l'on parle de *louer* suivi de complément d'objet direct *inanimé concret* (cf. IV 4 et 5), on propose bizarrement deux catégories de sens différents avec les mêmes exemples ci-dessus, si toutes les deux concernent la même action de *louer* au sens de *donner qqch. en location* (wynajmować, wynająć) et non pas au sens de *réserver qqch.* (rezerwować, zarezerwować). Si l'on *loue* quelque chose, on peut le faire *sans réservation*, et cette négation prouve que ces deux verbes ne possèdent pas les mêmes sens ni les mêmes référents, donc le choix des équivalents polonais ne devrait pas être facultatif dans ce cas-ci (cf. p. ex. la notion du triangle sémantique de Ogden & Richards, 1946 et la différence entre le sens et la référence, l'intension et l'extension).

## V Approche Orientée Objets (AOO) et description du verbe *louer*

Étant donné que les expressions testées plus haut (*louer la/sa femme, louer la voiture, louer la maison*) sont ambigües et ce fait produit un réel malaise pour les humains et les programmes de traduction automatique face à la désambiguïsation et la traduction, cette tâche ne sera pas accomplie, sans référence à l'étude des contextes proches ou plus éloignés du mot ou de l'expression en question. Autant dire que, pour doter les dispositifs automatiques de données nécessaires, il faut, d'un côté, définir la totalité des emplois des mots, et, de l'autre côté, décrire les cadres et les scripts où tel ou tel sens du mot ou de l'expression peut se réaliser.

Dans ce qui suit, nous proposerons des schémas de description orientée objets de *louer*. Ils permettront de définir aussi bien les compléments du verbe en termes des classes d'objets auxquelles ils appartiennent, ainsi que les cadres et les scripts typiques dans lesquels les phrases avec le verbe *louer* apparaissent et le script particulier que *louer* réalise.

### LOUER :

#### A : WYNAJMOWAĆ / WYNAJAĆ

1. X – [ANM humain] – *louer* – Y – [CONCRET INANM <objets et espaces louables>] de – Z – [ANM humain] pour – V – [ABSTRAIT <montant d'argent déterminé>] – pour – S – [ABSTRAIT <temps>] – *wynajmować/ wynająć* coś od kogoś za coś (za jakąś sumę pieniędzy) na jakiś okres czasu  
p. ex. : *Pierre a loué la maison à Lyon de cette agence immobilière pour [le montant de] x euros par mois pour un an.*
2. X – [ANM humain] – *louer* – Y – [CONCRET INANM <objets et espaces louables>] de – Z – [ANM humain] contre – V – [ABSTRAIT <services>] – pour – S – [ABSTRAIT <temps>] – *wynajmować/wynająć* coś od kogoś w zamian za coś (za usługę) na jakiś okres czasu.  
p. ex. : *Il a loué sa maison contre des achats et des aides quotidiens pour six mois.*
3. X – [ANM humain] – *louer* – Y – [CONCRET INANM <objets et espaces louables>] à – Z – [ANM humain] pour – V – [ABSTRAIT <montant d'argent déterminé>] – pour – S – [ABSTRAIT <temps>] – *wynajmować/ wynająć* coś komuś za coś (za jakąś sumę pieniędzy) na jakiś okres czasu.

p. ex. : *Il a loué un appartement à son collègue de travail pour x euros par mois pour un an.*

Le schéma 3 est une variante de permutation du schéma 1, on le laisse, ici et dans les cas pareils qu'on verra par la suite, comme une variante autonome juste pour une plus grande clarté de la présentation des schémas.

4. X – [ANM humain] – louer – Y – [CONCRET INANM <objets et espaces louables>] à – Z – [ANM humain] contre – V – [ABSTRAIT <services>] – pour – S – [ABSTRAIT <temps>] – wynajmować/wynająć coś komuś w zamian za coś (za usługę) na jakiś okres czasu.

p. ex. : *Il veut louer une chambre de sa maison à un étudiant contre un service de garde le chien en vacances pour six mois.*

Le schéma 4 est une variante de permutation du schéma 2.

5. X – [ANM humain] – louer – Y – [CONCRET INANM <objets et espaces louables>] à – Z – [ANM humain] – [ABSTRAIT <montant d'argent déterminé>] – pour – V – [ABSTRAIT <temps>] – wynajmować/wynająć coś komuś na jakiś okres czasu.

p. ex. : *Cette agence de location de voitures a loué deux Peugeot aux clients étrangers pour x euros pour 15 jours.*

Le schéma 5 est une variante de permutation du schéma 4.

6. X – [ANM humain] – louer – Y – [ANM humain] de – Z – [ANM humain] pour – V – [ABSTRAIT <montant d'argent déterminé>] – pour – S – [ABSTRAIT <temps>] – wynajmować/wynająć kogoś komuś za coś (za jakąś sumę pieniędzy) na jakiś okres czasu.

p. ex. : *Cette agence de travail loue les femmes de ménage pour x euros par semaine pour six mois.*

7. X – [ANM humain] – louer – Y – [ANM] à – Z – [ANM] pour – [ABSTRAIT <montant d'argent déterminé>] – V – [TEMPS] – wynajmować/wynająć kogoś komuś na jakiś okres czasu.

p. ex. : *Cette agence loue les femmes de ménage aux clients pour x euros pour une journée.*

*Cette agence loue des chevaux pour x euros par semaine pour six mois.*

**B: CHWALIĆ / POCHWALIĆ**

1. X – [ANM humain] – louer – Y – [ANM] pour (de) – Z – [ABSTRAIT <traits physiques, psychologiques, comportementaux>; action] – devant – V – [ANM humain] – chwalić/pochwalić kogoś za coś (za posiadanie jakiejś cechy) przed kimś  
p. ex. : *Il a loué sa femme pour son honnêteté devant tout le monde.*
  
2. X – [ANM humain] – louer – Y – [ANM] pour – Z – [ACTION] – devant – V – [ANM humain] – chwalić/pochwalić kogoś za coś (za robienie lub nierobienie czegoś) przed kimś.  
p. ex. : *Il a loué sa femme pour son travail devant sa mère.*  
*Il a loué ce cheval pour son travail devant son chef.*  
Le schéma 2 est une variante de permutation du schéma 1.
  
3. X – [ANM humain] – louer – Y – [CONCRET INANIMÉ] pour (de) – Z – [ABSTRAIT <traits physiques, psychologiques, comportementaux>] – devant – V – [ANM humain] – chwalić/pochwalić coś za coś (za posiadanie jakiejś cechy) przed kimś.  
p. ex. : *Il a loué sa voiture pour sa solidité devant son ami.*  
*Il a loué sa voiture pour pouvoir émettre des sons extraordinaires devant son ami.*  
Le schéma 3 est une variante de permutation du schéma 1 et 2.
  
4. X – [ANM humain] – louer – Y – [ABSTRAIT <traits physiques, psychologiques, comportementaux>] – de – Z – [ANM] – devant – Z – [ANM humain] – devant – V – [ANM humain] – chwalić/pochwalić coś u kogoś przed kimś.  
p. ex. : *Il a loué la façon de s'habiller de sa femme devant ce créateur de mode.*  
*Il a loué la persévérance de ce cheval devant son amie.*
  
5. X – [ANM humain] – louer – Y – [ABSTRAIT <traits physiques, psychologiques, comportementaux; action>] – de – Z – [INANM] – devant – V – [ANM humain] – chwalić/pochwalić coś (u) czegoś przed kimś.  
p. ex. : *Il a loué le style de sa maison devant sa femme.*  
*Il a loué la solidité de sa voiture devant son voisin.*  
*Il a loué la vitesse de sa voiture devant son copain.*



Le schéma 5 est une variante de permutation du schéma 3 et 4, et par conséquent du schéma 1 et 2.

6. X – [ANM humain] – *louer* – Y – [ANM] – pour (de) – Z – [ABSTRAIT <traits physiques, psychologiques, comportementaux>; action] – devant – V – [ANM humain] – *chwalić/pochwalić kogoś za coś przed kimś*.

p. ex. : *Il a loué sa femme pour son honnêteté devant son ex-mari.*

*Il a loué la persévérance de son cheval devant les acheteurs.*

Le schéma 6 est une variante de permutation du schéma 1 et 2.

7. X – [ANM humain] – *louer* – Y – [ABSTRAIT <traits physiques, psychologiques, comportementaux; action>] – devant – Z – [ANM humain] – *chwalić/pochwalić coś przed kimś*.

p. ex. : *Il a loué son esprit d'organisation devant ses chefs.*

Il est à souligner que les classes d'objets proposées, qui définissent et collectent des unités de langue selon le critère des mêmes propriétés sémantiques, sont encore à expliciter davantage en ajoutant autant d'éléments que possible, en particulier des éléments spécifiques pour le cadre et le script en question. Et ainsi, p. ex. la classe d'objets de <traits physiques, psychologiques, comportementaux> contiendrait les mots qui satisfont cette caractéristique, p. ex. *beauté, sagesse, intelligence, assiduité, style, persévérance, modestie, piété, vitesse, etc.*

S'il s'agit des classes d'objets générales, telles que p. ex. la classe [ANM humain], il ne faut pas oublier que, à part les humains du type, p. ex. *policier, agent, homme, femme, enfant, boulanger, professeur, prêtre, chauffeur, chanteuse, etc.*, peuvent y apparaître aussi des unités employées au sens métonymique. Dans le cas de *louer*, on peut trouver souvent sur la position du sujet ou du complément des mots ou des expressions reflétant une relation métonymique fondée sur un lien logique entre p. ex. : *agences, institutions, organisations, sociétés, etc.*, et *leurs employés* (cf. p. ex. Czekaj, 2011, 2018).

Quant aux cadres et scripts, ils rendent possible une bonne traduction des phrases ambiguës, non seulement lorsque les classes d'objets ne fournissent pas suffisamment d'informations pour le faire ou lorsque toutes les positions présentées dans les schémas ne sont pas réalisées dans les phrases. Le cadre est un fonds statique de la situation générale dont on parle dans un fragment de texte. Si la situation est dynamique et elle représente une séquence d'actions visant la réalisation d'un but, nous avons affaire à un script qui se réalise toujours dans un cadre donné.

Le script est composé de plusieurs phrases qui situent les classes d'objets des phrases du texte dans des unités notionnelles plus grandes qui orientent leur rôle, leur interprétation et leur spécification. Il peut être représenté par un seul verbe/prédicat constituant une phrase, ou plusieurs phrases constituant un fragment de texte. Le script peut être donc représenté, d'une part, grâce et par une sorte de décomposition sémantique amenant à un ensemble de phrases avec plusieurs verbes/prédicats ou, d'autre part, par une sorte de cumulation de ces éléments sémantiques dans un seul verbe constituant une phrase.

Ce type de décomposition sémantique s'effectue par la recherche non seulement des éléments sémantiquement nécessaires, comme c'est le cas d'une décomposition sémantique d'un prédicat dans les approches logiques (cf. p. ex. Karolak, 2007), mais aussi des éléments typiques de l'action, si besoin est. Elle ne s'effectue pas non plus à l'aide des éléments sémantiquement plus simples ou des primitifs sémantiques, mais grâce à l'identification de la séquence ordonnée des actions constituant la situation en question.

La première situation globale de *LOUER* dans le sens de *location* a lieu quand on a affaire aux éléments du type ci-dessous (*LOUER* est écrit en majuscule, parce qu'on parle de la situation qui peut être représentée de différentes façons, non seulement par le verbe *louer*):

*LOUER (location):*

*Déterminer ce qu'on veut avoir, pour combien de temps et à quel prix.*

*Chercher qui ou ce qu'on veut.*

*Comparer différentes possibilités et choisir la plus appropriée.*

*Examiner si ce qu'on a choisi correspond à ce qu'on veut.*

*Analyser le contrat de location.*

*Négocier le contrat si besoin est.*

*Signer le contrat et en prendre une copie pour soi-même.*

*Payer la location et, si besoin est, un dépôt de garantie.*

La deuxième situation est représentée par les emplois de la forme verbale *louer*. Ces emplois reprennent en général les éléments qui sont constitutifs pour les séquences particulières d'actions constituant la situation de *LOUER* (p. ex. *objet ou service désiré, prix, temps, contrat, paiement*, cf. les exemples ci-dessus).

Quant au deuxième sens de *louer* : *apprécier, complimenter*, la présentation par les scripts de cette situation aurait la forme p. ex. :

*LOUER (apprécier, complimenter) :*

*Evaluer qch ou qqn.*

*L'apprécier beaucoup.*

*Communiquer cette évaluation, généralement publiquement et pendant des occasions spéciales.*

La deuxième situation descriptive est représentée par les emplois verbales de *louer*, qui reprennent en général les éléments qui sont constitutifs pour les séquences particulières d'actions constituant cette situation de *LOUER* (p. ex. objet ou qualité considérés comme ayant une valeur ou une qualité remarquable, allant de pair avec des manifestations des émotions positives, cf. les exemples ci-dessus).

On remarque que dans le cas de *louer* dans le sens de *complimenter, apprécier*, à la différence de l'autre sens de *louer*, dans sa représentation par les scripts, donc dans la séquence des actions constitutives, il n'y a aucune référence aux éléments (*argent, contrat, durée, etc.*) d'une transaction financière. Cela ne veut pourtant pas dire que la transaction financière ne puisse être l'objet d'une louange, comme p. ex. dans le cas des phrases du type :

*Le président du groupe a loué publiquement son chef du département pour avoir mené à bien une transaction financière réussie.*

Les phrases représentant un cadre ou un script donné contiennent les mots-activateurs nécessaires pour une désambiguïsation efficace des fragments douteux. Ainsi, dans les cas analysés, pour que les traducteurs automatiques puissent choisir entre *wynajmować/wynając* et *chwalić/pochwalić* dans le type des phrases mentionnées, il suffirait de faire référence au script « *transaction financière* » dont le script *location* est une partie, représenté par des mots spécifiques à ce type de situation contenus dans l'ensemble des phrases entourant la phrase ou les phrases analysées, et qui activerait le premier des deux sens du verbe *louer* – *wynajmować/wynając*.

Il en résulte que les mots-activateurs qui y sont propres et qui sont reliés, chacun, à tout un réseau sémantico-lexical représentant le champ notionnel construisant le script et les classes d'objets en question, ce sont p. ex. (faute de place, on se limite ici à mentionner seulement ceux qui sont les plus représentatifs): *argent (et tous les mots du réseau sémantico-lexical d' « argent » et d'autres mots mentionnés ci-dessous), fonds, liquidités, fortune, capital, richesse, avoir, épargne, trésor, investissement, portefeuille, produit, gain, propriété, numéraire, valeur, cassette, patrimoine, placement, fruit, masse, possession, succession, fric, sous, pognon,*

*finances, billet (de banque), liquide, espèces, argent comptant, numéraire, coupure, papier-monnaie, comptant, argent liquide, appartements, maisons, voitures, vélos, bateaux, agence immobilière, agence de travail, société, entreprise, transaction, contrat, accord écrit, accord oral, paiement, rémunération, frais, location, service, dépôt, garantie, termes, conditions, assurance, conditions de retour, carte bancaire, prélèvement bancaire, gagner, perdre, perte, profit, marge, réduction, bénéfice, bénéfique, net, brut, contrat, signer, location, locataire, propriétaire, bail, bailleur, propriétaire, agence, immobilier, occupant, habitant, hôte, preneur, sous-locataire, sous-location, rendre, caution, payer, paiement, réserver, coûter, coût, loueur, montant, taxe, client, monnaie, prix, virement, somme, offre, filiale, commerce, commerçant, commercial, fonds, bien, chèque, propriété, assurance, immobilier, immobilière, négociier, négociation, finaliser, annuler, marché, marchand, affaires, finance, financier, financière, réservation, business, heure, jour, mois, an, légal, etc.*

Et, dans le cas p. ex. de la phrase (cf. III 12) :

*Le mari **loue** sa femme en secret. Il veut qu'elle ne le dise à personne. Ce qui importe, c'est de l'argent qu'elle gagne comme ça.*

puisque l'expression *louer sa femme en secret* est ambiguë et il est difficile de la désambiguïser grâce aux schémas avec des classes d'objets seuls ci-dessus, les mots-activateurs du script « *location – transaction financière* » : *argent* et *gagner*, qui apparaissent dans son contexte lointain, pourraient faciliter la désambiguïstation et, en principe, activer la bonne traduction. Mais, comme on l'a vu ci-dessus, la situation dans ce cas-ci est encore plus complexe, parce que les deux traductions concurrentes entrent toujours en jeu si l'on se limitait seulement au contexte de ces trois phrases et il faut puiser encore plus dans le contexte précédant et suivant pour relever les mots-activateurs du script dont il est question.

Cf. à cet égard le premier contexte ci-dessous :

*Le mari **loue** sa femme en secret. Il veut qu'elle ne le dise à personne. Ce qui importe, c'est de l'argent qu'elle gagne comme ça. Mais, il se rend bien compte que le fait que sa femme gagne de l'argent grâce au fait qu'il lui a permis de travailler chez leur voisin en effectuant des travaux sans payer d'impôts, même s'ils se sont mis d'accord que ce sera seulement pour une semaine, n'est pas une situation légale.*

*Gagner, travailler chez leur voisin, travaux, payer, impôts, pour une semaine, légale* sont les mots-activateurs du script « *location – transaction financière* » et *louer* serait dans ce contexte-ci à traduire en polonais comme *wynajmować/wynająć*.

Cf. le deuxième contexte :

*Le mari loue sa femme en secret. Il veut qu'elle ne le dise à personne. Ce qui importe, c'est de l'argent qu'elle gagne comme ça. Elle est très courageuse et se débrouille très bien ! Elle a décidé d'être gardienne de nuit dans un casino, mais c'est pourtant une honte que ce soit une femme et non pas un homme qui fasse ce type de choses !*

*Très courageuse, se débrouiller très bien*, sont les mots-activateurs du script « *louanges – compliments* » et louer serait dans ce contexte-ci à traduire en polonais comme *chwalić/pochwalić*.

## VI Conclusion

Le Modèle Orienté Objets (cf. Banyś 2002a, 2002b, 2005) propose une description détaillée des unités linguistiques apparaissant dans le contexte large d'une phrase dans le but de résoudre les problèmes de la traduction, surtout automatique, des mots et des expressions ayant plusieurs sens en y appliquant quelques outils linguistiques appropriés. Dans cette conception, chaque unité polysémique est décrite à l'aide des classes d'objets (cf. aussi Buvet, 1998 ; Buvet & Grezka, 2009 ; Gross, 1994, 1997, 2008, 2012 ; Mathieu-Colas, 1998 ; Le Pesant & Mathieu-Colas, 1998 ; Prandi, 1998 ; Vivès, 1998, etc.) créées par un ensemble d'opérateurs et d'attributs qui les définit. Cependant, souvent, la répartition des unités en classes d'objets n'est pas une opération suffisante pour en dégager les sens. Dans ce type des situations, les expressions ou phrases ambiguës ne se laissent pas bien traduire sans faire recours aux cadres et/ou scripts où ils apparaissent. Il n'est possible de les désambigüiser que grâce aux mots-activateurs propres à un cadre ou un script donné. Ainsi, dans le cas analysé, le script appelé ici « *location – transaction financière* » comprenant les mots-activateurs qui sont représentatifs pour la situation en question permet de traduire le verbe *louer* comme *wynajmować/wynająć* et non pas *chwalić/pochwalić*. D'autres mots-activateurs apparaissent si l'on a affaire au script représenté par le verbe *louer* dans le sens de *complimenter*.

Une description détaillée des schémas descriptifs orientés objets et des cadres et/ou scripts, dont un échantillon limité vient d'être présenté ci-dessus, constituent une bonne source d'information et une base possible d'apprentissage

pour les réseaux neuronaux des traducteurs automatiques actuels afin qu'ils puissent traduire les textes d'une manière encore plus efficace et encore plus rapidement.

## Références citées

- Allen, J. (2003). Post-editing. *Computers and Translation: a Translator's Guide*, 297–317.
- Banyś, W. (2002a). Bases de données lexicales électroniques – une approche orientée objets. Partie I : Questions de modularité. *Neophilologica* 15, 7–28.
- Banyś, W. (2002b). Bases de données lexicales électroniques – une approche orientée objets. *Neophilologica* 15, 206–249.
- Banyś, W. (2005). Désambiguïsation des sens des mots et représentation lexicale du monde. *Neophilologica* 17, 57–76.
- Banyś, W. (2018). Nouveaux anciens paradigmes : Approche orientée objets, Classes d'objets, Psychologie écologique et Linguistique. *Neophilologica* 30, 25–41.
- Bevilacqua, M., Pasini, T., Raganato, A. & Navigli, R. (2021). Recent Trends in Word Sense Disambiguation: A Survey. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21) Survey Track*, 4330–4338.
- Buvet, P-A. (1998). Détermination et classes d'objets. *Langages* 131, 91–102.
- Buvet, P-A. & Grezka, A. (2009). Les dictionnaires électroniques du modèle des classes d'objets. *Langages* 176, 63–79.
- Czekaj, A. (2011). Question de métonymie dans la traduction automatique. *Neophilologica* 23, 136–149.
- Czekaj, A. (2018). Perception et métonymie – problèmes de traduction automatique. *Neophilologica* 30, 76–88.
- Eriguchi, A., Tsuruoka, Y. & Cho, K. (2017). Learning to Parse and Translate Improves Neural Machine Translation. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (72–78). Association for Computational Linguistics.
- Gross, G. (1994). Classes d'objets et description des verbes. *Langages* 115, 15–30.
- Gross, G. (1997). Synonymie, polysémie et classes d'objets. *Meta* 42, 147–154.
- Gross, G. (2008). Les classes d'objets. *Lalie* 28, 113–165.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Presses universitaires du Septentrion.
- Hrabia, M. (2016). *Itérativité dans la traduction automatique*. Wydawnictwo Uniwersytetu Śląskiego.

- Hrabia, M. (2018). Verbe remettre et ses équivalents polonais. Désambiguïsation et analyse sémantico-préfixale pour les besoins de la traduction automatique. *Neophilologica* 30, 109–120.
- Hutchins, W. & Somers, H. (1992). *An introduction to machine translation*. Academic Press New York.
- Karolak, S. (2007). *Składnia francuska o podstawach semantycznych*. Collegium Columbinum.
- Le Pesant, D. & Mathieu-Colas, M. (1998). Introduction aux classes d'objets. *Langages* 131, 6–33.
- Marengo, S. (2021). *La théorie sens-texte : concepts-clés et applications*. L'Harmattan.
- Mathieu-Colas, M. (1998). Illustration d'une classe d'objets : les voies de communication. *Langages* 131, 77–90.
- Melčuk, I. (1997). *Vers une linguistique Sens-Texte. Leçon inaugurale*. Collège de France.
- Melčuk, I. (2012). *Semantics: From Meaning to Text*. John Benjamins.
- Melčuk, I. (2016). *Language: From Meaning to Text*. Academic Studies Press.
- Ogden, Ch. K. & Richards, I. A. (1946[1923]). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism. Eighth edition*. Harcourt, Brace & World, Inc.
- Perz, M. (2013). La polysémie adjectivale – un défi pour le traitement automatique des langues. *Roczniki Humanistyczne* 61, 61–74.
- Prandi, M. (1998). Contraintes conceptuelles sur la distribution : réflexions sur la notion de classes d'objets. *Langages* 131, 34–44.
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.
- Rubino, R. (2011). *Traduction automatique statistique et adaptation à un domaine spécialisé*. Thèse de doctorat, Université d'Avignon et des Pays de Vaucluse. <https://theses.hal.science/tel-00879945/document>, consulté en 07.2023.
- Shi, X., Padhi, I. & Knight, K. (2016). Does string-based neural mt learn source syntax? *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1526–1534.
- Śmigielska, B. (2007). Remarques sur la traduction automatique et le contexte. *Neophilologica* 19, 253–267.
- Śmigielska, B. (2013). Description des cadres dans l'approche orientée objets en vue de la traduction assistée par ordinateur. *Roczniki Humanistyczne* 61(8), 49–60.
- Śmigielska, B. (2015). Traduction automatique des mots sémantiquement proches dans l'approche orientée objets. *Neophilologica* 27, 241–252.
- Śmigielska, B. (2016). Inclusion et traduction automatique. *Language* 42(2), 139–152.
- Śmigielska, B. (2017). De la désambiguïsation et de la traduction de « éducation » ou de l'éducation de la désambiguïsation et de la traduction. *Roczniki Humanistyczne* 65(8), 157–169.

- Śmigielska, B. (2020). Cadres, scripts, registres – complexité de traduction des mots polysémiques. *Neophilologica* 32, 348–368.
- Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M. & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open* 1, 5–21.
- Trinth, S. (2019). L'intelligence artificielle et la traduction évoluent ensemble. *Le Magazine d'Information des Langagiers Circuit* 141, <https://www.circuitmagazine.org/dossier-141/l-intelligence-artificielle-et-la-traduction-evoluent-ensemble>, consulté en 07.2023.
- Tyagi, N., Chakraborty, S., Kumar, A. & Romeo, N. K. (2022). Word Sense Disambiguation Models Emerging Trends: A Comparative Analysis. *Journal of Physics: Conference Series* 2161.
- Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*, 5998–6008.
- Vivès, R. (1998). « Les mots pour le dire » : vers la constitution d'une classe de prédicats. *Langages* 131, 64–76.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J. R., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G. S., Hughes, M., & Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *ArXiv, abs/1609.08144*.
- Żłobińska-Nowak, A. (2011a). Sur la notion de classe d'objets en linguistique et son utilité dans la désambiguïsation des sens des mots. *Linguistica Silesiana* 32, 203–212.
- Żłobińska-Nowak, A. (2011b). Validité de la modélisation objet dans la langue et dans la traduction. *Neophilologica* 23, 249–258.
- Żłobińska-Nowak, A. (2013). Traduction automatique des langues – espérances et enjeux des outils appliqués. *Neophilologica* 25, 56–66.






### **Zygmunt Vetulani**

Adam Mickiewicz University, Poznań  
Poland

 <https://orcid.org/0000-0003-4833-8601>

### **Grażyna Vetulani**

Adam Mickiewicz University, Poznań  
Poland

 <https://orcid.org/0000-0002-2138-3704>

# Towards Lexicon-Grammar Verbnet Through Lexical Ontologies

#### **Abstract**

In this article, we present research directly inspired by the Princeton WordNet lexical *ontology* project (Miller, Fellbaum), which was a response to the real need for ontologies corresponding to the *natural conceptualization* common to all language users, within a given natural language, or within a specific *sublanguage*. Lexical ontologies for a given language or language subsystem determined by the scope of communication needs turn out to be useful and even necessary for constructing formal models of linguistic competence and, consequently, for designing and implementing AI systems with linguistic communicative competence, both passive and active. An important milestone of the research program presented in this work is the acquisition of tools in the form of extensive lexical ontologies of a new type, referred to in this work as *Lexicon-Grammar Verbnet*. In the article, we refer to the works of authors such as: Alain Colmerauer, Charles Fillmore, Christiane Fellbaum, Gaston Gross, Maurice Gross, Thomas R. Gruber, Richard Kittredge, George A. Miller, Martha Palmer, Kazimierz Polański, and Piek Vossen.

#### **Keywords**

lexical ontology, synonymy, valency structure, wordnet, PolNet, Lexicon-Grammar Verbnet, IT systems with language competence

The basis of all human mental activity is the formation of abstract concepts. At first, this enabled humans to understand a situation and make decisions. Later, the need to organize themselves in communities forced *communication* with other individuals of the group. As individuals began to identify themselves and distinguish themselves from other members of the community, there was a need to *plan and implement* collective activities within the group based on a *conceptual model of the surrounding world*. The model of the environment in which a human being functions is necessary to undertake rational actions consistent with the adopted goals. *Conceptualization* as a process of creating abstract concepts was of interest to the philosophers of the late nineteenth and early twentieth centuries. In the field of computer science, the term “conceptualization”<sup>1</sup> was popularized by Tomasz R. Gruber (1993) by aptly linking it with the understanding of the term “ontology,” introduced by him to computer science. Gruber characterized the term “ontology” in a compact form as follows: “An ontology is an explicit specification of a conceptualization. [...] Ontology is a systematic account of Existence. For knowledge-based systems, what *exists* is exactly that which can be represented.” (1993, p.199).

Information technologies, including AI technologies, which are the subject of this article, refer to the issues of modeling *human language competence* in order to use the models obtained for the design and implementation of *systems with linguistic communicative competence*.

The research presented below was directly inspired by the Princeton WordNet lexical ontology pioneering project implemented by G. A. Miller and C. D. Fellbaum (see e.g., Miller et al., 1990), which was a response to the real need for ontologies corresponding to the *natural conceptualization* common to all users of a given language, or a sublanguage used in a strictly defined field. The Princeton WordNet has been an inspiration for many lexical ontologies for various languages (including PolNet).

Lexical ontologies for a given language system or subsystem (determined by specific and well-defined communication needs) turned out to be useful, and even necessary for constructing formal models of linguistic competence, and consequently for designing and implementing *AI systems* with language communication competence, both passive and active.

---

<sup>1</sup> We use the term “conceptualization” in the sense given by Thomas R. Gruber, who wrote: “A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly” (Gruber, 1993, p. 199). In nineteenth-century philosophy, the term was used to refer to the formation of abstract concepts.

Knowledge processing and the ability to build a model of knowledge about the environment in which language users participating in the speech act function (people, devices, systems) are two key components necessary to achieve the goal of creating new generation AI systems at a level significantly exceeding current *chatGPT* systems. We will show, among other things, a number of results, including our own, which make up a methodologically coherent whole, and which bring us – step by step – closer to the above-defined goal. An important stage of the research program outlined here is obtaining tools in the form of complex lexical ontologies of a new type, referred to as *Lexicon-Grammar Verbnets*.

The title of this paper, “Towards Lexicon-Grammar Verbnets Through Lexical Ontologies,” illustrates the course (of a part) of our work in the field of Human Language Technologies from the 1980s until now. We consider obtaining a lexicon-grammar of the verbnets type (*Lexicon-Grammar Verbnets for Polish*) with a rich conceptual coverage as a solid basis for further R&D and implementation works in the field of IT.

In this article, we present the results scattered across a number of our publications containing essential elements and ideas. They form the backbone of a long-term, ongoing research program.<sup>2</sup>

The review of the results begins with early works, conducted in the conditions of scarcity of digital language resources, both lexical and grammatical. These are prototypes of systems constituting the BPII (Basic Polish for Information Interchange) family, as well as results in the field of digital lexical data and digital grammatical data formats obtained as part of national and European projects (POLEX and the EU projects PECO-COPERNICUS CEGLEX and PECO-COPERNICUS GRAMLEX); see section Early Works. Section WordNet Like Lexical Ontologies focuses on the development of a wordnet lexical ontology. The part concerning basic research mainly deals with the problems of synonymy, while the practical part presents the implementation of a lexical ontology of the WordNet type (for the Polish language) PolNet v1. Section From PolNet 1.0 to Lexicon-Grammar VerbNets is the main part of the work and concerns the transformation of the PolNet v1 lexical database into a lexical ontology, which is a VerbNet type Lexicon-Grammar. The most important challenge at the current stage of development of WordNet systems with Lexicon-Grammar features turned out to be the extension of synonymy relations and homonymy/hyperonymy relations to predicative synsets. This section discusses, among other things, the currently performed tasks.

---

<sup>2</sup> Theoretical and practical studies reported here relate directly to Polish, but are largely localizable to other language systems, primarily from the Indo-Aryan language family (Vetulani et al., 2021).

## Early Works

The research referred to in the title of the article is the direct result of previous projects that made us aware of the shortages of basic resources and IT tools for processing the Polish language. The beginnings of our work on systems with linguistic competence in the 1980s and 1990s, and partly their continuation, were characterized by the lack of access to digital linguistic resources (dictionaries, grammars) in a form that would enable their direct use in IT applications. Nevertheless, the Polish language belongs to a small elite group of languages with a long tradition of linguistic work, which turned out to be a solid theoretical base for our research.

The successful development of systems with language competence became possible thanks to the work we started on a grammatical description of the Polish language, suitable for IT use in parsing algorithms, that is, in algorithms that perform syntactic analysis, which is a preparatory stage in the process of calculating the meaning of a text.<sup>3</sup> The result of this work were the POLINT grammars developed since 1980s. Our source of inspiration was the question-answering system ORBIS implemented in PROLOG for English and French by A. Colmerauer and R. Kittredge (using DCG) (Colmerauer & Kittredge, 1982), later extended by Vetulani with a Polish module (Vetulani, 1988).

The first POLINT programs (see Vetulani, 1988) focused on modeling question-answer dialogues, were created in order to demonstrate the application potential in terms of language coverage in BPII systems<sup>4</sup> and to obtain practical knowledge of linguistic resources necessary to meet the needs of application systems. This potential was positively tested in the confrontation with the empirical material in the form of a corpus of empirically generated dialogues (Vetulani, 1990), and finally confirmed in the POLINT-112-SMS system (Vetulani & Osiński, 2017).

The first successful attempts to parse sentences of the Polish language already allowed us, in accordance with our expectations and with postulates of Antonio

---

<sup>3</sup> It should be noted here that this research was carried out under simplifying assumptions, namely the compositionability and computability of the language. These assumptions have been discussed among philosophers of language and linguists since at least the Enlightenment period, but they seem essential for creating IT-useful, precise models of language, pushing the boundaries of deterministic language modeling.

<sup>4</sup> The subset of Polish corresponding to the grammatical coverage of POLINT prototypes from the late 1980s is referred to in this period as BPII (Basic Polish for Information Interchange).

Zampolli,<sup>5</sup> creator and promotor of the concept of *Language Industry* to identify priorities in terms of technological needs of Language Engineering. One of the first of our ventures was the POLEX dictionary project (1994–1996).<sup>6</sup>

Polish is a language with a complex inflection system and has a relatively free word order. Therefore, simple adaptation of processing algorithms efficient for languages like English or French appeared hard to apply, as in Polish the basic information concerning the function of a word in the sentence is typically being encoded in the word form, independently of its linear position in the sentence. Dictionaries are a suitable place to store this information. At the time we started our research, good-quality grammatical descriptions of Polish existed only in the form of traditional dictionaries and grammars addressed to traditional customers. However, these resources, typically addressed to human readers, appeared to be of low usefulness for automatic processing because of lack of precision.

Our solution, the POLEX Polish Lexicon, is an electronic morphological dictionary which includes the core Polish vocabulary of general interest acquired from the traditional paper dictionary (Szymczak, 1983–1985).<sup>7</sup> POLEX is based on a precise machine-interpretable format (coding system), the same for all grammatical categories (Vetulani et al., 1998a).

The POLEX format we propose is uniform for all grammatical categories (parts of speech) and does not apply exceptions to the rules, which makes creating algorithms for generating and lemmatizing text much easier than when using traditional language descriptions, which place high demands on programmers due to the excessive complexity of the description. The POLEX dictionary entries take the following form:

BASIC\_FORM+LST\_OF\_STEMS+PARADIGMATIC\_CODE+DISTRIBUTION\_OF\_STEMS

---

<sup>5</sup> Antonio Zampolli considered the lack of resources in the form of IT-processable corpora, dictionaries, and digital grammars necessary to build language models to be a critical obstacle in the development of utility IT systems. See *Language Resources. Overview* by J. J. Godfrey and Antonio Zampolli (1996) and (Zampolli, 2006).

<sup>6</sup> The first public release of the resource contained over 42,000 nouns, 12,000 verb, 15,000 adjectives, 25,000 participles, and about 200 pronouns.

<sup>7</sup> Supplemented by the basic swear words not found in this dictionary and the most frequently used elements of jargon, regional and colloquial vocabulary. Several paper editions of *Słownik Języka Polskiego PWN* [Polish Language Dictionary PWN] by Mieczysław Szymczak were edited between 1978 and mid 1990s. For our purposes we used the three volume version published from 1983 to 1985; see (Szymczak, 1983–1895).

For example, dictionary entries for two inflected variants of the word *sucker*<sup>8</sup> look as follows:

frajer; frajer, frajerz; N110; 1;1-5,9-13;2:6-8,14

frajer; frajer, frajerz; N110; 1;1-5,8-14;2:6-7

The *paradigmatic inflection code* contains full paradigmatic information about *inflection*, that is, the way of associating endings with stems in order to obtain the desired word form. The inflection code (here *N110*) includes full information on morphology and inflection, in particular a list of endings appropriate for all paradigmatic positions. The distribution parameter (*distribution\_of\_stems*) relates stems (here *frajer*, *frajerz*) to paradigmatic positions. The information stored in a dictionary entry is complete and unambiguous, and inflection classes are constructed in such a way that there is no need to consider exceptions.

The other two projects discussed in this section were of a different nature.

The main goal of the CEGLEX consortium (Vetulani et al., 1998b) was to test the EU EUREKA project GENELEX that proposed a reusable generic model for lexicons assumed to respond to IT needs. GENELEX was implemented (between 1990 and 1994) for a number of Western European languages, such as French, English, German, Italian. Within CEGLEX three Central-European languages, Polish and Czech (Slavic) and Hungarian (Finno-Ugric) were used as testbeds to verify genericity of the GENELEX model.

It is worth noting that the final Polish module developed in CEGLEX/GENELEX went further than original GENELEX, which focused on morphological and syntactic layers while the semantic layer was addressed only marginally.

The three layers of the CEGLEX/GENELEX model were confronted with linguistic data of the languages under consideration with generally positive results. For the Polish module of the project this confrontation consisted in adapting the model GENELEX to Polish language data.

The CEGLEX project resulted in a successful attempt to test (on representative linguistic data) the feasibility of an IT-oriented lexicon-grammar covering all three basic layers (morphological, syntactic, and semantic) of grammatical description.

---

<sup>8</sup> The word *frajer* (en. *sucker*) is a masculine-personal noun (pol. *rzeczownik męskoosobowy*), inflected for number and case (two numbers /singular and plural/ and seven cases. The themes (*frajer* and *frajerz*) are the same in both entries. Code *N110* represents a 14-position string of endings, the same in both entries for the lexeme *sucker*: (, a, owi, a, em, e, e; y, ów, om, ów, ami, ach, y). The fourth parameter, topic distribution, describes the assignment of each paradigmatic ending to the appropriate topic. In this example, 1:1-5,9-13 means that the first stem (*sucker*) is combined with the endings from paradigm positions 1 to 5 and 9 to 13. Similarly, the expression 2:6-8,14 indicates that the endings from positions 6, 7, 8 and 14 are connected to the second stem (*sucker*).

The CEGLEX/GENELEX methodology together with the results of the POLEX project were the starting point for our work within the PECO-COPERNICUS GLAMLEX project, carried out from 1995 to 1998. The main goal of this project was to build, in accordance with the GENELEX methodology, morphological digital dictionaries and related IT-oriented tools. The intention of the tasks of the GRAMLEX project (Vetulani et al., 1998b) regarding the Polish language was to contribute to the improvement of the situation in the field of language engineering tools and resources. Among the main achievements of GRAMLEX was the creation of a corpus-based morphological dictionary for the Polish language encoded in SGML (in the proprietary GRAMCODE format).<sup>9</sup>

The GRAMLEX project turned out to be the first step towards implementing a lexicon-grammar for the Polish language.

## WordNet Like Lexical Ontologies

### Synonymy, Hyperonymy and Inheritance<sup>10</sup>

The concept of *synonymy* refers to the concept of *meaning*, which is commonly used in informal discourse and usually does not raise controversy. Consequently, it is generally used as a *primary concept*, not requiring analytical definition referring to other concepts treated as known. If the reference to the obviousness of a concept turns out to be inappropriate, then an *axiomatic definition* can be used. Definitions of this type do not enter into the ontological nature of the defined concept, but are operational in nature, specifying the way of using the concept by referring to another, assumed to be already known. A classic example is the Peano arithmetic (around 1889),<sup>11</sup> where primitive concepts such as additions, multiplications, natural numbers, etc., are explained by axioms that, by reference to other concepts, determine how to use these concepts. In traditional linguistics, similar methods are sometimes used to determine the meaning of a word or phrase by giving usage examples considered representative. When defining

---

<sup>9</sup> The GRAMCODE dictionary included over 22,500 dictionary entries along with related tools and applications (lemmatizer, inflectional generator, concordance generator and others).

<sup>10</sup> This section is based on our paper “Synonimie et granularité dans les bases lexicales du type WordNet” (Vetulani, Z. and Vetulani, G., 2015) and “EuroWordNet General Document” (Vossen, 2002). In particular, we follow Vossen in using the term “hyperonymy” for nouns and verbs.

<sup>11</sup> See Giuseppe Peano (1889) at [https://en.wikipedia.org/wiki/Giuseppe\\_Peano](https://en.wikipedia.org/wiki/Giuseppe_Peano).

synonymy, reference to meaning may be appropriate if well-defined procedures are used to compare word meanings, for example, by applying context-of-use analysis (see Vossen, 2002).

Our initial work on lexical ontologies was motivated by the desire to obtain a basic ontology for the Polish language inspired on the one hand by Linnaeus' systematics, and on the other hand by the pioneering work of cognitive scientists and linguists from Princeton (Miller, Fellbaum and others) on the WordNet system. The Princeton WordNet was an ontology directly linked to lexical material in the form of abstract nouns grouped into classes of synonyms called *synsets*. The inspiration turned out to be accurate and led to the creation of the linguistic ontology, PolNet – Polish Wordnet (version 1.0), which satisfactorily corresponds to the conceptualization reflected in the nouns of the Polish language.

### ***The Problem of Defining the Concept of Synonymy***

In natural languages, concepts (understood as mental equivalents of complex or simple entities) are represented by words. *Synonymy* is commonly understood as a binary relation holding between two words (terms, expressions) if and only if its arguments have *the same or similar meaning*. The need to define the relation of synonymy more precisely leads us to distinguish the three cases where the term synonymy will be used.

Case 1. If the concepts represented by their names (being simple or compound nouns) are *extensional* (that is, when they can be fully described by specifying which of the entities fall under the given concept and which do not), then by *synonymy* of some two names we understand that both names refer to *the same set of entities*. In a similar way, we can construct an extensional definition of *synonymy* of verbs: two verbs are said to be synonymous when they both refer to the same set of states and/or events.

Case 2. In turn, when the meaning of each of the two words compared to each other can be unambiguously determined by *a specific* set of features and their values, then their *synonymy* means that both can be uniquely described by *the same* set of features (attributes) taking the same values.

Case 3. If neither of the above two cases occurs, then it remains to refer to definitions of the nature of procedures referring to the circumstances of the use of each of the compared terms.

Let us compare three of the frequently discussed solutions:

- 1) Leibnitz's proposal,
- 2) Princeton WordNet proposal (Miller-Fellbaum),
- 3) EuroWordNet proposal (Vossen).



Ex. 1) We quote, after Vossen (2002, p.18), a very strong definition of synonymy given by Leibnitz:

“two expressions are synonyms if the substitution of one for the other never change the truth value of a sentence in which the substitution is made.”

Note that when using this definition, *synsets* are generally very small, or even composed of one element only. This means a significant flattening of the hierarchy based on the hypernymy relationship, which in turn reduces the potential benefits of the inheritance mechanism of attributes associated with synsets and the values of these attributes. The advantage of the Leibnitz’s proposal is that synonymy is an equivalence relation and thus marks a partition in a set of words.

Ex. 2) George A. Miller and Christiane Fellbaum (see Vossen, 2002, p. 18) proposed a less restrictive approach to synonymy, encapsulated in the formula:

“two expressions are synonymous in a linguistic context C, if the substitution of one for the other in C does not alter the truth value.”

In the literal sense, it means that to conclude that these expressions are *not synonymous*, it is enough to refer to just one selected context C in which the replacement of one expression with another will change the logical value of the whole sentence.<sup>12</sup> *De facto*, this procedure (correctly) indicates as synonyms only those words for which the fixed context C can be considered *representative* of a particular meaning. In dictionary practice, the condition of representativeness of examples (containing the context of use) for illustrating a typical meaning is not strictly observed (see e.g., Polański, 1980–1992), which in practice may significantly hinder the creation of WordNet-type systems based on the above definition of synonymy.

Ex. 3) Piek Vossen proposes the synonymy tests for various parts-of-speech (including noun-noun, verb-verb, noun-verb, etc.) implemented in the Euro-WordNet project (Vossen, 2002).

---

<sup>12</sup> “The weak point of Miller’s approach is the synonymy criterion (above) which – alone – is not sufficient to form synsets because it does not guarantee transitivity when the C context changes. To remedy this defect, the initial criterion of synonymy must be reinforced by imposing the reference to the same context C.” (Vetulani, Z. and Vetulani, G., 2015, p. 117). [Translation from French by Z. Vetulani: “Le point faible de l’approche de Miller est le critère de synonymie (ci-dessus) qui – seul – ne suffit pas pour former les synsets car il ne garantit pas la transitivité quand le contexte C change. Pour remédier à ce défaut il faut renforcer le critère initial de la synonymie en imposant la référence à un même contexte C.”]

What follows is an example of an EuroWordNet context-based tests (applied to English) for noun-noun synonymy (Test 1) (Vossen, 2002, p. 19).<sup>13</sup>

#### Test 1. Synonymy between nouns

yes	a	if it is (a/an) X then it is also (a/an) Y
yes	b	if it is (a/an) Y then it is also (a/an) X
Conditions:		X and Y are singular or plural nouns
Example:	a	if it is a fiddle then it is a violin
	b	if it is a violin then it is a fiddle
Effect:		synset variants {fiddle, violin}

#### ***Hierarchical Organization of Concepts in PolNet***

The classic wordnet organization for nouns is based on a hierarchy of concepts referring to the relation of hyponymy/hyperonymy *for nouns*. This hierarchy has a *tree structure*. More general concepts are higher in this hierarchy and those more specific are lower down. Tree organization is intended to allow *inheritance of properties*, essential for knowledge representation and inference (see *Linnaean systematics*<sup>14</sup>). The extension of the PolNet lexical ontology to *predicative synsets*<sup>15</sup> introduces relations between predicative synsets and other ontology entities (synsets or not). Of particular importance is the introduction of relations that connect the predicative synset with arguments that are assigned attributes called *semantic roles* (which are synsets or other objects of the PolNet ontology). Assigning semantic roles to the argument positions opened in predicative expressions serves to determine links or connectivity constraints between these expressions and arguments.<sup>16</sup> Expanding PolNet with predicative synsets requires special caution when extending the hyponymy/hyperonymy relationship to predicative synsets.

#### ***PolNet Development Incremental Algorithm (Nouns)***

In this section we present an algorithm of creating synsets and hierarchical relations based on hyponymy/hyperonymy relations between nouns (simple and compound). This algorithm was directly used by lexicographers in the first phase

<sup>13</sup> Notice however, that context-based tests ignore differences due to the pragmatic factors.

<sup>14</sup> Carl von Linné (1707–1778), *Systema Naturae* (1770); see [https://en.wikipedia.org/wiki/Systema\\_Naturae](https://en.wikipedia.org/wiki/Systema_Naturae).

<sup>15</sup> By *predicative synsets* we mean synsets whose typical elements are predicative nouns or verb-noun collocations.

<sup>16</sup> The set of these assignments constitutes the *valency structure* of the synset.

of building the PolNet v.1 database. The DebVisDic platform developed at the Masaryk University Brno was used (Pala et al., 2007).

Application of the algorithm requires:

- the Visdic or DEBVisDic platform (or any functionally equivalent tool),
- on-line access to Princeton WordNet,
- a *good* monolingual lexicon<sup>17</sup> (called *reference dictionary* in the algorithm description), preferably accessible on-line (we used *Uniwersalny słownik języka polskiego PWN*<sup>18</sup> (Dubisz, 2006) as the basic reference lexicon and *Słownik języka polskiego PWN* (Szymczak, 1995) as a complementary one),
- a team with both language engineering and lexicographical skills.

The algorithm input consists of a list of words (lexemes). The output is a WordNet code segment for: a) synsets, b) the ISA relation between synsets (determined by the hyponymy/hyperonymy relation).

The general procedure for expanding PolNet consists in performing a sequence of operations, step by step:

1. Looking through the reference dictionary, we search for *word-meanings*<sup>19</sup> that are synonyms.
2. We create synonymy classes using appropriate definition criteria. These classes are called *synsets*.
3. For created or modified synsets, we search for candidates for hyponyms and hyperonyms using our own language competence, dictionaries, LSR list and knowledge of the Princeton WordNet structure.
4. For pairs of synsets selected in step 3, we perform hyponym and hyperonym definition tests.

Short example:

Let us take the Polish word *zamek* as an example. The list of the word-meanings identified at in step 1 will be:

- *zamek-1* (*zamek* I-1 in the dictionary): *a lock*
- *zamek-2* (separated from the *zamek-1* meaning, where the phrase *zamek błyskawiczny* is mentioned): *a zip fastener*
- *zamek-3* (*zamek* I-2): *a machine blocking lock, e.g. a valve lock*
- *zamek-4* (*zamek* I-3): *a gun lock*
- *zamek-5* (*zamek* II-1): *a castle*

*Zamek-2*, *zamek-3* and *zamek-4* will all be hyponyms of *zamek-1*.

---

<sup>17</sup> By a *good dictionary* we mean one where different word-meanings are explicitly distinguished.

<sup>18</sup> PWN is the name of a Polish publishing house.

<sup>19</sup> By *word-meaning* we mean a meaning of the *literal* together with its reference-dictionary-meaning number.

### ***Language Resources Used: Dictionaries and Tools***

The research, the main results of which are summarized in this article, make up a description of the research path leading to a coherent methodology enabling the design and implementation of large AI systems<sup>20</sup> with language competence. One of the most important milestones of the long-term research program discussed here is the implementation of a prototype of a large AI system used for practical verification of decisions regarding the selection and/or development of appropriate tools and methods for natural language engineering.<sup>21</sup>

In addition to standard tools and methods commonly recognized as elements of the canon of IT and linguistic knowledge, during our research (until the implementation of the testing system /POLINT-112-SMS/), we considered it appropriate to use two classes of resources:

- A) specialized resources and publicly available tools – necessary or useful in the project,
- B) own resources and tools developed in the project, which turned out to be needed to implement the milestones of our work.

Class (A) includes:

- IPI PAN National Corpus of Polish Language (on a limited scale) (Przepiórkowski, 2004),
- *PWN Polish Language Dictionary* (version edited by M. Szymczak, 1995),
- *Universal Dictionary of the Polish Language* (edited by S. Dubisz, 2006),
- *Syntactic-Generative Dictionary of Polish Verbs* (Polański, 1980–1992),
- Internet dictionary SJP.PL, more on this topic in (Vetulani et al., 2010, p. 158–159),
- Tools for generating WordNet lexical networks – VisDic and DebVisDic (Masaryk University Brno) (Pala et al., 2007).

Group (B) includes:

- formats and vocabularies created in the POLEX, GRAMLEX and CEGLEX projects (Vetulani et al., 2010),
- a corpus of private SMS records – collected and made available by Justyna Walkowska,
- a corpus of experimental SMSs (collected and described by Justyna Walkowska in her PhD dissertation<sup>22</sup> (see, e.g., Vetulani et al. 2010),
- a corpus of legal texts (compiled from open sources),

---

<sup>20</sup> By “large AI system” we mean a utility application at the stage of at least pre-commercial testing.

<sup>21</sup> The appropriate system called POLINT-112-SMS has been described in a collective monograph (Vetulani et al., 2010) and its brief characteristics are in the annex to this work.

<sup>22</sup> See (Walkowska, 2012).

- a corpus of recordings from the emergency telephone 997/112 (confidential recordings, not intended for sharing),
- verb-noun collocations for the Polish language: methodology, data formats, predicative-nouns /basic resource/ (Vetulani, G., 2000), basic noun-synsets-creation algorithm (Vetulani et al., 2007), algorithms for expanding the collocation dictionary (Vetulani, G., Vetulani, Z., Obrębski, T., 2008), a digital dictionary of verbal-nominal collocations (Vetulani, G., 2012),
- coding algorithms for valency dictionaries,
- various algorithms for expanding the PolNet database (as of 2010).

### Inspirations. Princeton WordNet

Creating advanced systems with language competence, such as AI systems, requires knowledge processing, and thus referring to abstract concepts. For this purpose, ontologies (as defined by T. R. Gruber) are used (see the opening paragraph of the article). Ontologies, which on the one hand correspond to the natural conceptualization of the world – real or fictitious, and on the other hand are formal entities subject to IT processing, are *WordNet-type systems*.<sup>23</sup> The WordNet lexical ontology (also known as Princeton WordNet /PWN/) is an implementation, in the 1980s by G. A. Miller and colleagues at Princeton University's Cognitive Sciences Laboratory, of a new method for describing semantic vocabulary that has proven particularly useful for searching information on the Internet. The key idea of this method is to present the lexicon described by referring to the concepts of synonymy and hyperonymy. PWN is composed of classes of synonyms called *synsets* and is organized hierarchically by the relation of hyponymy/hyperonymy between synsets. Some other semantic relations between synsets (as meronymy, antonymy, etc.) are implemented as well. WordNet-like systems have an advantage over traditional ontologies because they explicitly account for the relationships between the words of the language and the concepts of the ontology /represented by synsets/.

### Lexical Ontology PolNet v1

Our research initiated in the early 2000s was inspired by the work of George A. Miller and his team on the WordNet lexical ontology, as well as by the work

---

<sup>23</sup> This term is used to describe ontological systems modeled on PWN.

led by Piek Vossen in the EuroWordNet project. At later stages of work on the PolNet system, we also relied on the pioneering research of Maurice Gross on the concept of Lexicon-Grammar, initially implemented for the French language (Gross, M., 1975; 1994; 1981) and independently conducted work (in the same period) by Kazimierz Polański, and crowned with the implementation in 1980–1992 of the *Syntactic-Generative Dictionary of Polish Verbs*, as well as on the results of the FrameNet (Fillmore et al., 2002) and VerbNet (Palmer, 2009) projects, close to the assumptions of Lexicon-Grammar.

The launch in 2006 of the construction of PolNet (a lexical ontology intended to be a wordnet-type lexical database) was a response to the need for a language processing module for implementation of an stand-alone, large-scale IT system with language competence (POLINT-112-SMS) (Vetulani, Z., 2014). While the concept and structure of the PolNet database was modeled on the solutions adopted for the Princeton WordNet system (Miller and Fellbaum, 2007), the methodology for creating the PolNet database was developed from scratch by a team of Polish computer scientists and lexicographers.<sup>24</sup> The adopted methodology assumed the use of existing dictionaries of Polish in order to maintain the conceptualization appropriate for users of the Polish language.

The PolNet database is a structure built from synonym classes and relations between these classes. Synonym classes (*synsets*) represent concepts identifiable in natural language, thanks to which PolNet can be used as a *lexical ontology* corresponding to the conceptualization reflected in the Polish language. PolNet v1 was built on the basis of high-quality traditional dictionaries of the Polish language and the study of available language corpora (such as IPI PAN Corpus (Przepiórkowski, 2004) and small domain corpora). Resource creation is done incrementally, starting with high-frequency vocabulary<sup>25</sup> and words that are (for various reasons) considered important.

While the initial work on PolNet was conducted towards a system with a structure similar to the Princeton WordNet and intended to serve as an ontology naturally associated with the language, over time, the PolNet project, influenced by theoretical work carried out independently by Maurice Gross and Kazimierz Polański and implementation-oriented works (in particular by Alain Colmerauer and Charles Fillemore), evolved into a Lexicon-Grammar by gradually incorpo-

---

<sup>24</sup> Mainly from the Department of Computer Linguistics and Artificial Intelligence of the Adam Mickiewicz University and the Faculty of Modern Languages and Literatures of the Adam Mickiewicz University.

<sup>25</sup> A departure from this principle, made for methodological reasons in order to enable early testing of the developed resource in applications for which the condition of lexical completeness must be met, was the inclusion of terminology specific to these applications.

rating simple and compound verbs. This evolution coincided with the progress of theoretical work on the development of a formalized dictionary of verbal-nominal collocations initiated in the 1990s by Grażyna Vetulani (see Vetulani G. 2000; 2012), and with Gaston Gross's independent research on the category of object classes (fr. *classes d'objets*) (see e.g., Gross, G., 1994).

The first versions of the PolNet database, made available to a limited extent before 2012, included mainly nouns and the most important verbs. It was also during this period that *verb-noun collocations* began to be included in the PolNet database. The addition of simple and complex verbs (verb-noun collocations) along with syntactic information was the first step towards giving the PolNet lexical database the character of Lexicon-Grammar (as understood by Maurice Gross and Kazimierz Polański).

What follows is a (simplified) example of a noun synset (code).<sup>26</sup>

```
<SYNSET>
<ID>PL_PK-28557</ID>
<POS>n</POS>
<DEF>drobniutki, sproszkowane ziarenka ziemi, piasku, różnego rodzaju
rozkruszonych lub bardzo rozdrobnionych ciał, unoszące się w powietrzu
i osiadające na powierzchni przedmiotów; kurz</DEF>
<SYNONYM>
<LITERAL lnote="U1" sense="1">pył</LITERAL>
<LITERAL lnote="U3" sense="3">proch</LITERAL>
<LITERAL lnote="U1a" sense="1">pyłek</LITERAL>
</SYNONYM>
<USAGE>Cząsteczki pyłu wirują w powietrzu.</USAGE>
<USAGE>Po wyburzeniu kamienicy wszystko spowijał pył.</USAGE>
<USAGE>Pył cementowy, wapienny, krzemowy, węglowy, azbestowy.
</USAGE>
<USAGE>Pył śnieżny, wodny, pustynny.</USAGE>
<USAGE>Tumany, kłęby pyłu.</USAGE>
<USAGE>Pył opada, osiada, wznosi się, wciska się w usta.</USAGE>
<USAGE>Zetrzeć z czegoś pył.</USAGE>
<USAGE>Otrześć, otrząsnąć, omieść coś z pyłu.</USAGE>
<ILR type="hypernym">POL-2141601944</ILR>
<SNOTE>--kurz</SNOTE>
<SNOTE>--próchno</SNOTE>
<BCS></BCS>
```

<sup>26</sup> An example of a noun synset from: (Vetulani Z. et al., 2010), p. 192.

```

<NL>>false</NL>
<STAMP>przemekr 2007-06-14 18:34:21</STAMP>
<CREATED>przemekr 2007-06-14 18:34:21</CREATED>
</SYNSESET>

```

Synset description (simplified):

Synset (set of synonyms)	{pył1,pyłek1a,proch3}
Synset ID	PL_PK-28557
	% <i>pył1</i> oznacza słowo 'pył' w pierwszym znaczeniu słownikowym,
	% <i>pył1a</i> oznacza zdrobnienie dla <i>pył1</i>
	% <i>proch3</i> oznacza słowo 'proch' w jego trzecim znaczeniu słownikowym
Definition	Drobniutki, sproszkowane ziarenka ziemi, piasku, różnego rodzaju rozkruszonych lub bardzo rozdrobnionych ciał, unoszące się w powietrzu i osiadające na powierzchni przedmiotów; kurz.
Use example	Cząsteczki pyłu wirują w powietrzu.
Use example	Z daleka widać było tumany, kłęby pyłu.
Use example	Po wyburzeniu kamienicy wszystko spowijał cementowy pył.
Hypernym ID	POL-2141601944

### Usefulness of WordNet Lexical Networks for IT Application Development

The usability of the PolNet network as a lexical ontology in specific applications (e.g., in AI systems with language competence) is primarily determined by the properties of the concepts of *synonymy* and *hyperonymy*, as well as the features of *lexical coverage* (more on the prospects for the development of lexical ontologies of the PolNet/Lexicon-Grammar VerbNet type later in the article).



## From PolNet 1.0 to Lexicon-Grammar VerbNets

The extent to which WordNet lexical networks will be useful in IT applications is determined by the properties of the concepts used in defining these networks. These concepts include, above all, the notion of *synonymy*, as well as *relations* defined on synsets. Of the latter, the relation of *hyperonymy* between the synsets representing particular concepts is the most important. Hyperonymy plays the role of the backbone for organizing the structure of the synset network. In the network, synsets can also enter into relationships with entities other than synsets, e.g., with attribute values or metadata.

Already the first attempts to extend the PolNet system with simple and complex verbs prompted us to in-depth reflection on synonymy and homonymy. The aim was to propose definitions that would correspond to the intuitive understanding of these concepts by linguists and at the same time be of a procedural nature, facilitating the writing of algorithms for creating synsets and extending the homonymy/hyperonymy relationship for the purposes of knowledge management using mechanisms of inheritance of the features of objects represented by synsets.

In order for the search for appropriate solutions to be tested on the basis of language material in applications of a practical nature (on a real scale), it was first necessary to supplement those language resources that were used to complete the first stage described in section WordNet Like Lexical Ontologies, as well as to acquire or create new resources. In this respect, a pioneering task was the development of dictionaries of predicative nouns and verbal-nominal collocations (Vetulani, G., 2000; 2012), as well as the proposal of a model for encoding and implementing grammatical information assigned to verb synsets for collocations. The most important of these tasks are listed in section Synonymy, Hyperonymy and Inheritance (see also Vetulani, Z. et al., 2010).

### New Inspirations

Kazimierz Polański (1929–2009), parallel to Maurice Gross (1934–2001), was a precursor of the idea of Lexicon-Grammar. In his formalized dictionary of Polish verbs Polański includes entries with morphological, syntactic, and semantic information related to the chosen word form, which is also the ID of the entry (Polański, 1976; 1980–1992). The dictionary was developed and published in the years 1980–1992 and included 7,000 entries for the most important Polish verbs.

At the same time and independently of Polański, Maurice Gross was working on the formal description of the French verbs. Gross's concept is similar to Polański's in that the word form of the verb is directly related to the relevant lexical and semantic information. Gross held the opinion that the determinants of the meaning of words are elementary sentences characterizing their typical uses. Both of the above approaches are also found in the idea of the WordNet lexical network implemented under the direction of Charles Miller at Princeton University and organized around the concept of *synonymy*, which makes WordNet legitimately considered a lexical ontology.

All three approaches in the initial phase were implemented independently for significantly different languages: English, French and Polish (in alphabetical order). These languages are characterized by a different grammatical, and, in particular, dictionary tradition, which (probably) explains the fact that the initial research was not mutually quoted.

An important reason for the wide-spread adoption of these ideas is their significant application potential, supported by insightful theoretical work aimed at strengthening the lexical and grammatical coverage of important data repositories and tools needed for the development of language engineering (including multilingual aspects).

The EU-funded EuroWordNet project led by Piek Vossen (Vossen, 2002) went in this direction. The excellent theoretical documentation of EuroWordNet, was an important source of inspiration for the Lexicon-Grammar Verbnet for Polish.

## **Creating Lexical and Grammatical Resources and Their Digitization: Predicative Nouns and Verb-Noun Collocations**

### ***The Need for Lexical and Grammatical Resources***

Initial work on the PolNet system was motivated by the desire to obtain ontologies sufficient to meet the basic needs<sup>27</sup> in the field of knowledge representation. At this stage, an ontology that well reflects the conceptualization typical of the language that people use every day seemed to be sufficient. Thus, in the initial period, limiting our work to the noun category was justified. However, this state turned out to be insufficient when there was a need to represent knowledge about situations, states, and events in AI systems, typically expressed in language

---

<sup>27</sup> The need for large lexical resources was not significant in the initial period of work discussed in Section Early Works, because until the end of the 1980s in Poland there were no favorable conditions for practical work in the field of natural language technology.

by predicative-argument structures, inspired by computer logic and knowledge engineering. Hence the need to extend PolNet with language constructions used to express relational content.

The basic lexical categories for this role are simple (one-word in Polish) or complex verbs of various types (see Vetulani, G., 2000). Among the latter, in Polish and a number of other languages, the most important are verb-noun collocations, composed of a supporting verb and a predicative noun, belonging to the category of abstract nouns. The support verb (Vsup) primarily plays a syntactic role, but sometimes also a different one (e.g., pragmatic or socio-linguistic), while the predicative noun (Npred) is associated with syntactic and semantic attributes. The latter are organized in valency structures expressing (through the attribute values) requirements or constraints of connectivity with arguments in a sentence structure. (More on the predicate-argument model in Vetulani, Z., 1998 and 2004 and Karolak, 1984).

Since the class of verb-nominal constructions is much more flexible and evolutionarily open than the class of simple verbs, we considered it reasonable to treat this class as a priority in the development of the PolNet database. The first step was to develop a careful methodology for recognizing the use of a compound structure as a verb-noun collocation acting as the center of a sentence. We will devote the rest of section Valency Structures in PolNet Lexical Ontologies to the acquisition of verbal-nominal collocations.

At the beginning of our research in this field, we focused on the description of the noun capable of playing the role of predicate in the verb-noun construction (Vetulani, G., 2000). In the 1970s and 1980s, the first important work on the predicative noun in the verb-noun construction appeared in the French literature; see (Giry-Schneider, 1978), (Danlos, 1980), (Vivès, 1983), (Gross, G., 1987). It is customary for a predicative noun (Npred) to appear in an analytic construction, partly fixed (frozen), forming with its accompanying verb (Vsup) a verb-noun collocation (Vsup + Npred) which plays the role of the sentence center (in simple sentences). The first tangible result of implementing the assumptions described above was the development of a digital dictionary covering over 14,600 Polish verb-noun collocations that can act as a predicate in a sentence (Vetulani, G., 2012).

### ***Valency Structures in PolNet Lexical Ontologies***

When the main goal of the first phase of the PolNet project, which ended with the implementation of PolNet v1, was to develop noun synsets and the relations between them (induced by semantic relations between the elements of synsets), the extension of PolNet to *verb synsets*, and more generally *predicative synsets*,

posed a significant challenge that forced redefinition of the concept of synonymy. The modification required the introduction of relations aimed at enabling the formulation of conditions of connectivity between verb and noun synsets (representing predicate and arguments, respectively). This function is played by valency structures.

The key to making the right decisions regarding the development of linguistic ontologies for grammatical categories other than abstract nouns is to follow the idea already successfully validated for nouns at the stage of PolNet v1 implementation. What we mean here is that the structure of a formal ontology is consistent with the natural categorization of knowledge, so that Gruber's<sup>28</sup> maxim, which has worked for noun ontologies, does not lose its validity for other grammatical categories. This was the direction of research by a number of linguists working on the formal description of the semantics of natural languages.

In this field the most active among Polish linguists was Kazimierz Polański, while for other languages, pioneering research was conducted by Charles Fillmore, Martha Palmer (English), Maurice Gross and Gaston Gross (French), Piek Vossen (Dutch) and others. It is important to take account of the work of mathematicians and logicians such as Gottlob Frege, Alfred Tarski, Richard Montague and Kazimierz Ajdukiewicz, who had an essential impact on the formation of the model of thinking about natural language in the pre-informatics period.

Grouping together verb synsets and noun synsets according to the semantic-syntactic connectivity constraints imposed by the argument positions opened in a sentence by the predicate gives the PolNet system the status of a lexicon-grammar. The key to extending synonymy to predicative phrases (simple predicative verbs, predicative nouns, verbal-nominal predicative collocations and other grammatical categories) in such a way as to respect compatibility with the idea of Lexicon-Grammar is the concept of valency structure (Vetulani Z. and Vetulani, G., 2014).

By *valency structure* we mean here information about all argument positions opened by a predicative word, taking into account both semantic constraints on arguments, as well as, morpho-syntactic constraints on text elements filling these positions (case, number, gender, etc.) (Vetulani, Z. and Vetulani, G., 2015). In particular, we require synonyms to have the same valency structure and the same assignments of *semantic role values*. Thus, the valency structure is one of the formal exponents of *meaning* and, *ipso facto*, imposes strong granularity constraints on the synonymy of predicative expressions.

---

<sup>28</sup> "An ontology is an explicit specification of a conceptualization" (Gruber, 1993, p. 199).

### Extending the Synset Definition.

Traditional descriptions of the vocabulary of natural languages generally distinguish words in terms of the *meaning* that is assigned to them. By *meaning* we understand the reference of a word<sup>29</sup> by the user to reality (real or fictional). This reference may associate a word with an object, a class of objects, or a bundle of relevant semantic features.

From the point of view of knowledge representation, particular importance is attached to the noun and verb categories. Both of these categories are composed of simple and complex forms. Typical *meanings of nouns* are entities (physical or abstract) or their descriptions. Typical *meanings of verbs* (simple or compound) are relationships between entities (physical or abstract), as well as states and events relating to entities (as well as other states, events, etc.).

Extension of the dictionary with synsets containing predicative expressions (predicative verbs and nouns, predicative collocations, etc.) is done with the use of predicative uses obtained from text corpora. Analysis of usage contexts provided the necessary syntactic and semantic information used for further work.

As in the lexicographical tradition (traditional dictionaries), during the development of the current versions of the PolNet system, examples were grouped to illustrate related uses, but showing differences in surface implementation (see, for example, Polański, 1980–1992). For verbs and other predicative expressions, the key property defining their meaning (in the above sense) is the way they function in the structure of the sentence, in which they play a central (organizing) role (according to the widespread opinion of many linguists).

The description of the function of the entry in the structure of the sentence specifies the conditions of connectivity between the predicative expression (verb) and noun groups (as arguments). Connectivity conditions are described in the *valency structure*, which consists of appropriately selected *syntactic frames* obtained from the analysis of empirical material (corpus).

### Implementation of Valency Structures.

The (simplified) example below is supposed to give a rough idea of the implementation of simple valency structures for predicative words. This is a code generated for the synset that includes four synonyms with a common meaning that may be translated to English as *to help*. This synset is composed of two Polish predicative simple verbs (Vpred) (*pomóc* and *pomagać*) and two (predicative) verb-noun collocations (Vsup+Npred) (*udzielić pomocy* and *udzielać pomocy*). (Each of the two

---

<sup>29</sup> Or other linguistic signs.

pairs contains the perfective and the imperfective forms, *pomóc-udzielić pomocy* and *pomagać-udzielać pomocy*, respectively.)

#### VALENCY STRUCTURE IMPLEMENTATION EXAMPLE (simplified)

The code below is a simplified representation of a verb synset (in the DEBVisDic system notation) containing a simple verb (*to help* in perfective and imperfective forms) and its synonym in the form of a collocation (*help* in perfective and imperfective forms).

POS: v ID: 3441

Synonyms: {pomóc:1, pomagać:1, **udzielić pomocy:1**, **udzielać pomocy:1**} (*to help*)

Definition: “wziąć lub brać udział w pracy jakiejś osoby (zwykle razem z nią), aby ułatwić jej tę pracę” (“*to participate in sb’s work in order to help him/her*”)

Frame: Agent(N)\_Benef(D)

Frame: Agent(N)\_Benef(D) Action(‘w’+NA(L))

Frame: Agent(N)\_Benef(D) Manner

Frame: Agent(N)\_Benef(D) Action(‘w’+NA(L)) Manner

Usage: Agent(N)\_Benef(D); “Pomogłem jej.” (*I helped her*)

Usage: Agent(N)\_Benef(D) Action(‘w’+NA(L)); “Pomogłem jej w robieniu lekcji.” (*I helped her in doing homework*)

Usage: Agent(N)\_Benef(D) Manner Action(‘w’+NA(L)); “Chętnie udzieliłem jej pomocy w lekcjach.” (*I helped her willingly doing her homework*)

Usage: Agent(N)\_Benef(D) Manner; “Ja chętnie jej pomagałem.” (*I used to help her willingly*)

Semantic\_role: [Agent] {człek:1, człowiek:1, homo sapiens:1, istota ludzka:1, ... } (*{man, human, ...}*)

Semantic\_role: [Benef] {człek:1, człowiek:1, homo sapiens:1, istota ludzka:1, ... } (*{man, human...}*)

Semantic\_role: [Action] {czynność:1} (*{activity}*)

Semantic\_role: [Manner] {ADVERB\_FEATURE\_QUALITY} (*an upper-ontology concept*)

#### Legend

The row “Synonyms” – members of the synset identified as “POS: v ID: 344” (POS stands for “Part Of Speech”)

Rows “Frames” – rows representing *semantic-syntactic frames* assigning attributes to argument positions / (slots) (in the basic /canonical/ surface order of a sentence)

The row “Definition” – (non-formalized) definition of the term corresponding to the synset /in Polish and English/

Rows “Usage” – typical examples corresponding to the intended meaning of the synset

Rows “Semantic\_role” – semantic role values (synsets or upper ontology concepts)

In the example above, the rows containing predicate-argument schemes of sentences (tagged as ‘frames’) are *compatible* with each other. We talk about the compatibility of predicative-argument schemes in a valence structure when any two rows of this structure tagged as frames have a common extension that is a frame in this structure. (For example, for ‘Agent(N)\_Benef(D) Manner’ and ‘Agent(N)\_Benef(D) Action(‘w’+NA(L))’ such common extension is ‘Agent(N)\_Benef(D) Action(‘w’+NA(L)) Manner’.) In sentences that match this extension, such as ‘Ja chętnie pomogłem jej w pracy’ /‘I gladly helped her in work’/, one can distinguish sub-sentences, as for example, ‘Ja chętnie pomogłem jej’ /‘I gladly helped her’/ and ‘Ja pomogłem jej w pracy’ /‘I helped her in work’/, corresponding to the respective sub-frames. In subsequent work, we will (primarily) use such valency structures in which all predicative-argument schemes are mutually compatible.

### **Intra-Synset Variations.**

Adoption of the concept of *meaning* as the starting point for the construction of lexical ontologies has important consequences for the utility aspects of the use of ontologies in engineering practice. The critical point is the *scope* of the commonly used, and thus vague, term *meaning*.

The carriers of intra-synset distinctions are often, but not exclusively, support verbs ( $V_{sup}$ ) in predicative compound constructions, such as verb-noun collocations ( $V_{sup}+N_{pred}$ ).

$V_{sup}$  plays an important role in the interpretation of complex predicative structures because:

- 1) in many cases it allows one to abolish the polysemy of the predicative form (here: a predicative noun) which is important from the point of view of applications, and also
- 2) brings information about register and aspect.

Grażyna Vetulani in her recent work on meaning-related aspects of support verbs ( $V_{sup}$ ), exhaustively analyzes the role of  $V_{sup}$  in complex predicative

expressions. She observes that despite its apparently subordinate non-predicative role, the support verb brings important semantic and grammatical information to the meaning of the whole predicative expression. In particular support verbs often serve to determine *register* and *aspect* of a collocation and, *ipso facto*, to concretize the *meaning* of the predicative noun (Npred) of the collocation (Vetulani, G., 2022).

### **Valency Structures in Lexicon Grammar VerbNets.**

Storing the valency structure together with synsets – as part of their description – brings a number of benefits from the point of view of use in NLP applications (parsing, computer understanding, text processing) and is consistent with the idea and practice of Lexicon-Grammar as a tool dedicated to the broadly understood area of Language Technology for real utility applications (Gross, M., 1979).

An example of using a lexicon-grammar for the Polish language (built on the basis of PolNet v3) is the prototype of the POLINT-112-SMS system intended to support information management and decision making in emergency situations (see e.g., Vetulani et al., 2010; Vetulani & Marciniak, 2011; Vetulani & Osiński, 2017). The system is able to interpret SMS texts messages, as well as understand and process information provided by the human user.

POLINT-112-SMS has also proven itself to be an environment for testing the usefulness of grammar lexicons in the creation of utility applications. In particular, easy access to valency information facilitated the creation of simple heuristics allowing for effective (smart) search space reduction in syntactic and semantic analysis (parsing). This feature enables computationally cheap creation and testing of prototypes of utility systems or their replaceable modules, as well as the development of systems with multilingual competence.

### **Verb-Noun Collocation Gathering.**

The family of verb-noun collocations is (in Polish and many Indo-European languages) an important group of compound verbs typically built of 1) an abstract predicative noun (Nsup) with a semantic and semantic-syntactic function, and 2) a support verb (Vsup), the main role of which is to introduce the predicative component (e.g., Nsup) and (often) to convey the pragmatic aspects (Vetulani, G., 2022). In some cases, the support verb is omitted from the surface structure (ellipsis). The semantic-syntactic function is primarily realized by the valency structure which fixes the conditions for the connection of the predicate with the arguments. In contrast to simple words, both nouns and (even more so) verbs, compound words are less well described than single-word forms for most languages. This is largely due to the scarcity of empirical research



based on representative corpora of texts, whether in written or spoken form. This circumstance is particularly important in the design and implementation of IT applications.

### **Digital Dictionary of Verb-Noun Collocations (Algorithm and Examples of Collocation Dictionary Entries).**

The dictionary developed by Grażyna Vetulani (2012) is the result of the IPI PAN corpus exploration (Przepiórkowski, 2004). The following five step algorithm served the lexicographers to set up the last public release<sup>30</sup> of the version of the lexicon.<sup>31</sup>

#### ***Algorithm.***

The 5-steps algorithm of corpus exploration and description includes the following steps:

- Step 1. Extraction from the corpus of the contexts with a high probability of containing verb-noun collocations, as well as detection of verb-candidates to be qualified as support verbs (automatically);
- Step 2. Manual analysis by lexicographers of the list of verb candidates obtained in the Step 1 in order to eliminate the evidently bad choices;
- Step 3. Automatic extraction of contexts in the form of concordances containing *verb-noun* pairs (selected through steps 1–2) as concordance centers;
- Step 4. Reading of the concordances by lexicographers, qualification of verb-noun pairs as collocations and providing their morpho-syntactic descriptions (manual);
- Step 5. Verification and final formatting.

The method used permitted the reduction (approx. 100 times) of the processing cost (estimation on the 5% sample).

We decided to put in the collocation lexicon all verb-noun collocations found in the corpus, because the inclusion in an electronic dictionary of a huge number of items is not a problem, as it would be for a traditional one. This also means that the lexicon contains, together with the well-known and currently commonly used collocations, a large number of less frequently used ones. The verb-noun lexicon demonstrates the dynamic and open character of the domain of nominal predication in Polish.

---

<sup>30</sup> The 2012 edition includes, after their examination and supplementation based on the IPI PAN Corpus, collocations collected until 2000.

<sup>31</sup> The carriers of intra-synset distinctions are often, but not exclusively, auxiliary verbs (Vsup) in complex predicative constructions such as verb-noun collocations (Vsup+Npred).

Various stages of work on the lexicon of verb-noun collocations were described in (Vetulani, G., 2000) and (Vetulani, G., 2012) and with the 2012 publication the lexicon was made available in digital form. The dictionary resource obtained was used in the implementation of valency structures in PolNet v3. For this reason, PolNet v3 may be considered the *first mature version* of Lexicon-Grammar Verbnet (Vetulani, Z. & Vetulani, G., 2016).

### ***Collocation Dictionary Entries – Examples.***

Verb-noun dictionary entries. Extract from the dictionary (Vetulani, G., 2012).

=>agresja, ż

czuć agresję/ czuć(B)/N1\_do(D);wobec(D);w stosunku do(D),

**dokonać agresji/ dokonać(D)/N1\_na(Ms),**

dokonać agresji/ dokonać aktu(D)/N1\_na(Ms),

dokonywać agresji/ dokonywać(D)/N1\_na(Ms),

**dopuszczać się agresji/ dopuścić się(D)/N1\_na(Ms),**

dopuszczać się agresji/ dopuszcząć się(D)/N1\_na(Ms),

**doświadczać agresji/ doświadcząć(D)/N1\_ze strony(D),**

doświadczyc agresji/ doświadczyc(D)/N1\_ze strony(D),

kierować agresję/ kierować(B)/N1\_przeciw(C),

odczuwać agresję/ odczuwać(B)/N1\_do(D);wobec(D);w stosunku do(D),

popęlnić agresję/ popęlnić(B)/N1\_wobec(D),

przejawiać agresję/ przejawiać(B)/N1\_wobec(D),

przejawić agresję/ przejawić(B)/N1\_wobec(D),

**reagować agresją/ reagować(N)/N1\_wobec(D),**

zareagować agresją/ zareagować(N)/N1\_wobec(D),

**skierować agresję/ skierować(B)/N1\_przeciw(C),**

**wybuchać agresją/ wybuchać(N)/N1\_wobec(C),**

wybuchnąć agresją/ wybuchnąć(N)/N1\_wobec(C),

**wykazać agresję/ wykazać(B)/N1\_wobec(D),**

**wykazywać agresję/ wykazywać(B)/N1\_wobec(D),**

zareagować agresją/ zareagować(N)/N1\_wobec(D).

## Examples of selected contexts confirming the use of collocations in the text corpus

<b>*** dokonać</b>	
po tym jak	[* dokonało_ono_agresji *] na Kuwejt,
podobnie jak swego	
ostrzegając: jeżeli ktoś	[* dokonałby_agresji *] na Polskę w czasie,
gdy	
Wprowadzając stan wojenny,	[* dokonano_agresji *] w brutalny, bo
siłowy sposób	
<b>*** dopuścić</b>	
dzieckiem ojcu, który	[* dopuścił_się_agresji *] i stosował
przemoc wobec matki	
chcieli przecież nie	[* dopuścić_do_takiej_agresji *]
<b>*** doświadczać</b>	
w swej historii	[* doświadczała_obcej_agresji *].
<b>*** reagować</b>	
czynności fizjologicznych,	[* reagowało_agresją *] i krzykiem na próby
nawiązania	
<b>*** skierować</b>	
w Hucie Jedność	[* skierowali_swoją_agresję *] przeciwko
prezydentowi miasta	
<b>*** wybuchać</b>	
niezadowolona z siebie,	[* wybuchała_agresją *].
<b>*** wykazać</b>	
To nie policja	[* wykazał_agresję *], to związkowcy zasto-
sowali bezprawne	
<b>*** wykazywać</b>	
obserwowany, to znaczy	[* wykazuje_dużo_agresji *], brutalności
wobec osoby słabszej, a	
zgromadzenia, którzy	[* wykazywali_szczególną_agresję *].
przez okno albo	[* wykazuje_agresję *] wobec innego dziecka.
na inne dziecko,	[* wykazuje_agresję *] albo chce wyskoczyć
z okna	

## Final Comments

The transition from the PolNet v1 phase to PolNet v2 was a significant step towards the Lexicon-Grammar, when the concept of the *valency structure* was launched. Starting from PolNet v2, *valency structures* were used in PolNet systems as the basic exponent of the meaning of *collections of predicative expressions* (simple or complex) organized in synsets. The requirement of *mutual compatibility of syntactic patterns for all elements of the synset* adopted for PolNet

determines that the valency structure is *ipso facto* a determinant of the *meaning* of the predicative synset. The work currently being carried out aims to significantly increase the lexical and linguistic coverage of the class of complex predicative expressions, as well as to expand the scope of research covering the pragmatic layer of the Polish language.

Positive results in terms of practical usefulness of the lexical ontology model<sup>32</sup> indicate directions of natural continuation of previous work. These will be:

- at the grammatical description level: work covering the syntactic and semantic levels corresponding primarily to the needs generated by emerging application perspectives; this work will require further acquisition of empirical data from representative corpora certifying the use of units,
- at the pragmatic level (currently in the initial phase): extension of the model with new factors that may allow the internal structure of synsets (intra-synset relations) to be taken into account,
- at the tool level: development and implementation (or adaptation of existing ones) of the most effective systems collecting the necessary empirical data.

## References

- Colmerauer A., & Kittredge, R. (1982). ORBIS, *9th International Conference on Computational Linguistics, COLING*.
- Danlos, L. (1980). *Représentation d'informations linguistiques : constructions N être Prép X*. Thèse de 3 cycle. Paris, L.A.D.L., Université Paris VII.
- Dubisz, St. (Ed.). (2006). *Uniwersalny słownik języka polskiego PWN* [Universal Dictionary of Polish; PWN], 2nd edition. Wydawnictwo Naukowe PWN.
- Fillmore, Ch., Baker, C. F., & Sato, H. (2002). Seeing arguments through transparent structures. *Third International Conference on Language Resources and Evaluation, Proceedings*, Vol. III (pp. 787–791). ELRA.
- Giry-Schneider, J. (1978). *Les nominalisations en français : l'opérateur « faire » dans le lexique*. Librairie DROZ.
- Godfrey, J. J., & Zampolli, A. (1997). Language resources. Overview. In R. Cole (Ed.), *Survey of the state of the art in human language technology* (pp. 381–408). Cambridge University Press.

---

<sup>32</sup> The POLINT-112-SMS system was used as a platform for testing the practical usability of wordnet-type lexical ontologies for implementing IT applications.

- Gross, G. (1987). *Les constructions converses en français*. Librairie DROZ.
- Gross, G. (1994). Classes d'objets et description des verbes. *Langages*, 15, 15–30.
- Gross, M. (1975). *Méthodes en syntaxe*. Hermann.
- Gross, M. (1979). On the failure of generative grammar. *Language*, 55(4) (Dec. 1979), 859–885.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63, 7–52.
- Gruber, Th. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Karolak, S. (1984). Składnia wyrażen predykatywnych. In Z. Topolińska (Ed.), *Gramatyka współczesnego języka polskiego. Składnia* (pp. 11–30). Wydawnictwo Naukowe PWN.
- Miller, G. A., Beckwith, R., Fellbaum, Ch., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *Int. J. Lexicograph.*, 3(4), 235–244.
- Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38 (11), 39–41.
- Miller, G. A., & Fellbaum, Ch. (2007). WordNet then and now. *Lang. Resour. Evaluation*, 41(2), 209–214.
- Pala, K., Horák, A., Rambousek, A., Vetulani, Z., Konieczka, P., Marciniak, J., Obrębski, T., Rzepecki, P., & Walkowska, J. (2007). DEB Platform tools for effective development of WordNets in application to PolNet. In Z. Vetulani (Ed.), *Proceedings of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5–7, 2007, Poznań, Poland* (pp. 514–518). Wyd. Poznańskie.
- Palmer, M. (2009). Semlink: Linking PropBank, VerbNet and FrameNet. *Proceedings of the Generative Lexicon Conference*. Sept. 2009, Pisa, Italy. GenLex-09.
- Polański, K. (1976). Słownik syntaktyczno-generatywny czasowników polskich, zeszyt próbny. *Prace Naukowe Uniwersytetu Śląskiego*, 124. Wydawnictwo Uniwersytetu Śląskiego.
- Polański, K. (Ed.) (1980–1992). *Słownik syntaktyczno-generatywny czasowników polskich*, vol. I–IV, 1980–1990. Ossolineum, vol. V, 1992. Instytut Języka Polskiego PAN, Kraków.
- Przepiórkowski, A. (2004) *Korpus IPI PAN. Wersja wstępna*. Instytut Podstaw Informatyki PAN.
- Szymczak, M. (1983–1985). *Słownik Języka Polskiego PWN* [Dictionary of Polish Language, in Polish]. Państwowe Wydawnictwo Naukowe.
- Vetulani, G. (2000). *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*. Wydawnictwo Naukowe UAM.
- Vetulani, G. (2004). Le rôle du verbe dans le réseau dérivationnel des prédicats nominaux. *Studia Romanica Posnaniensia*, XXXI, 459–467.

- Vetulani, G., Vetulani, Z., & Obrębski, T. (2008). Verb-Noun Collocation SyntLex Dictionary – Corpus-based approach. *Proc. of 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco* (pp. 1561–1564). ELRA.
- Vetulani, G. (2012). *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I*. Wydawnictwo Naukowe UAM.
- Vetulani, G. (2022). L'apport du Vsup au tour prédicatif verbo-nominal en polonais. *Neophilologica*, 34 (2022), 1–18. Wydawnictwo Uniwersytetu Śląskiego. <https://doi.org/10.31261/NEO.2022.34.10>.
- Vetulani, Z. (1988). PROLOG Implementation of an access in Polish to a data base. *Studia z Automatyki, XII*. Państwowe Wydawnictwo Naukowe (PWN), 5–23.
- Vetulani, Z. (1989). *Linguistic problems in the theory of man-machine communication in natural language. A study of consultative question answering dialogues. Empirical approach*. Brockmeyer.
- Vetulani, Z. (1990). *Corpus of consultative dialogues. Experimentally collected source data for AI applications*. Wydawnictwo Naukowe UAM.
- Vetulani, Z., Walczak, B., Obrębski, T., & Vetulani, G. (1998a). *Unambiguous coding of the inflection of Polish nouns and its application in the electronic dictionaries – Format POLEX*. Wydawnictwo Naukowe UAM.
- Vetulani, Z., Martinek, J., Obrębski, T., & Vetulani, G. (1998b). *Dictionary based methods and tools for language engineering*. Wydawnictwo Naukowe UAM.
- Vetulani, Z., Obrębski, T., & Vetulani, G. (2007). Towards a lexicon-grammar of Polish: Extraction of verbo-nominal collocations from corpora (pp. 267–268). In *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference (FLAIRS-07)*, AAAI Press.
- Vetulani, Z. (2004). *Komunikacja człowieka z maszyną. Komputerowe modelowanie kompetencji językowej*. Akademicka Oficyna Wydawnicza EXIT.
- Vetulani, Z., Marcinak, J., Obrębski, J., Vetulani, G., Dabrowski, A., Kubis, M., Osiński, J., Walkowska, J., Kubacki, P., & Witalewski, K. (2010). *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego* [Language resources and text processing technologies. POLINT-112-SMS as example of homeland security oriented application]. Wydawnictwo Naukowe UAM.
- Vetulani, Z., & Marciniak, J. (2011). Natural language based communication between human users and the emergency center: POLINT-112-SMS (pp. 303–314). In Z. Vetulani (Ed.), *Human language technology. Challenges for computer science and linguistics. LTC 2009. Revised Selected Papers*. LNAI 6562. Springer-Verlag.

- Vetulani, Z. (2014). (PolNt-Polish WordNet (pp. 408–416). In Z. Vetulani & J. Mariani (Eds.), *Human language technology. Challenges for computer science and linguistics. LTC 2011 Revised Selected Papers. LNAI 8387*. Springer-Verlag.
- Vetulani, Z., & Vetulani, G. (2014). Through Wordnet to Lexicon Grammar. In F. K. Doa (Ed.), *Penser le lexique grammairre: perspectives actuelles* (pp. 531–543). Editions Honoré Champion.
- Vetulani, Z., & Vetulani, G. (2015). Synonymie et granularité dans les bases lexicales du type WordNet. *Studia Romanica Posnaniensia, XLII* (1), 113–127.
- Vetulani, Z., Vetulani, G., & Kochanowski, B. (2016). Recent advances in development of a lexicon-grammar of Polish: PolNet 3.0 (pp. 2851–2854). In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), Paris, France.
- Vetulani, Z., & Osiński, J. (2017). Intelligent information bypass for more efficient emergency management. *Computational Methods in Science and Technology, 23*(2), 105–123. <https://doi.org/10.12921/cmst.2017.0000019>.
- Vetulani, Z., Vetulani, G., & Mohanty, P. (2021). Development of real size IT systems with language competence as a challenge for a Less-Resourced Language: a methodological proposal for Indo-Aryan languages. *Journal of Information and Telecommunication, 5*(4), 514–535. <https://doi.org/10.1080/24751839.2021.1966236>.
- Vivès, R. (1983). *Avoir, prendre, perdre : constructions à verbe support et extension aspectuelle*. Thèse de 3 cycle. Paris, L.A.D.L. et Université Paris-VIII.
- Vossen, P. (Ed.) (2002). *Euro WordNet. General Document. Version 3*. University of Amsterdam.
- Walkowska, J. (2012). *Modelowanie kompetencji dialogowej człowieka na potrzeby jej emulacji w zarządzających wiedzą systemach informatycznych współpracujących z wieloma użytkownikami*. PhD Dissertation at IPI PAN Warszawa (supervised by Z. Vetulani).
- Zampolli, A. (1996). Współpraca międzynarodowa w dziedzinie LR [International cooperation in the field of Language Resources, in Polish]. *Informatyka, 3*, 34–37.

## Web pages

- <http://alain.colmerauer.free.fr/alcol/ArchivesPublications/PrologHistory/19november92.pdf> (Alain Colmerauer and Philippe Roussel (1992). The birth of Prolog. GIA, Marseille). Last access: 12.09.2023.
- [https://pl.wikipedia.org/wiki/Słownik\\_języka\\_polskiego\\_\(Mieczysław\\_Szymczak\)](https://pl.wikipedia.org/wiki/Słownik_języka_polskiego_(Mieczysław_Szymczak)). Last access: 15.12.2023.
- <https://www.jstor.org/stable/412748> (Maurice Gross (1979). On the Failure of Generative Grammar). Last access: 12.09.2023.

[https://en.wikipedia.org/wiki/Giuseppe\\_Peano](https://en.wikipedia.org/wiki/Giuseppe_Peano) (Giuseppe Peano, 1889). Last access: 12.09.2023.

<http://www.illc.uva.nl/EuroWordNet/docs/GeneralDocDOC.zip> (Piek Vossen. (2002). EuroWordNet. Final Document). Last access: 12.09.2023.



Rédaction / Copy editing

PAWEŁ GOLDA (textes français / French texts)

EWA ŚMILEK (textes espagnols / Spanish texts)

TOMASZ KALAGA (textes anglais / English texts)

Projet de la couverture et de la page de titre / Cover and front page design

TOMASZ JURA

Préparation de la couverture à l'impression / Cover preparation for printing

PAULINA DUBIEL

Composition / Typesetting

PAULINA DUBIEL

**ISSN 2353-088X**

Attribution – Partage dans les mêmes conditions 4.0 International (CC BY-SA 4.0) /

Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)



Auparavant, la revue paraissait sous forme imprimée suivie de l'identificateur ISSN 0208-5550 /

This journal was formerly published in print with the following identifier ISSN 0208-5550

La version primaire référentielle de la revue est sa version électronique (en ligne) /

The primary referential version of the journal is its electronic (online) version

La revue est distribuée gratuitement /

The journal is distributed free of charge

Maison d'édition / Publisher

**Wydawnictwo Uniwersytetu Śląskiego / University of Silesia Press**

**ul. Bankowa 12B, 40-007 Katowice**

<https://wydawnictwo.us.edu.pl>

e-mail: [wydawnictwo@us.edu.pl](mailto:wydawnictwo@us.edu.pl)

Printed sheets: 32,0. Publishing sheets: 33,0.

Free copy

ISSN 2353-088X

9 772353 088301 3 5

About this book

