

Aude Grezka

Lexiques, Dictionnaires, Informatique (LDI)
CNRS — Université Paris 13, UMR 7187

La base de données *Figement*

Quant au mot «figée», il peut étonner au premier abord, mais il est préférable, me semble-t-il, à «fossile» ou à «pétrifiée», parce que ceux-ci font penser à quelque chose de mort, tandis que les locutions en question restent vivantes, bien que d'une autre manière qu'à l'origine. De plus, «figée» peint mieux que «fixe» le changement qu'ont subi ces locutions.

Lars Lindberg, 1898

Abstract

At present fixity plays an important role in linguistic research because of its determining role in recognition systems. In this article we present ongoing research on fixity, in particular on creation of a database whose objective is to list all the frozen expressions, basing on formal criteria. Taking adverbial phrases as an example, we shall specify the criteria allowing us to point to absolute fixity before tackling the question of its relative aspect.

Keywords

Formal criteria, fixity, absolute fixity, adverbial phrases.

Introduction

L'objectif du laboratoire LDI est de fabriquer de la ressource automatisable qui va pouvoir être utilisée dans des applications informatiques telles que l'analyse automatique de textes. Cela implique un degré de couverture des faits linguistiques suffisant, que seul un dictionnaire électronique construit à cet effet peut apporter. Toutes les unités lexicales qui constituent un texte doivent bien entendu être recon-

nues morphologiquement, mais aussi décrites selon leurs propriétés combinatoires, c'est-à-dire syntaxiques, et leur signification en contexte. Pour que la machine identifie toutes ces formes, il lui faut des ressources linguistiques de deux ordres : un dictionnaire morpho-syntaxique et un dictionnaire syntactico-sémantique. Le dictionnaire morpho-syntaxique est capable d'identifier les formes linguistiques et les variations qu'elles peuvent subir par les règles morphologiques de la langue en question ; le dictionnaire syntactico-sémantique est capable d'associer des formes à des sens. Cet objectif exige également que soient prises en compte les suites figées, dont les éléments constitutifs n'ont pas d'autonomie.

Le figement, longtemps absent des études linguistiques, occupe actuellement une place importante dans les préoccupations de la recherche, et ce en raison de son rôle déterminant dans les systèmes linguistiques de reconnaissance. Les études récentes (J. François, S. Mejri, 2006 ; S. Mejri, 1997 ; F. Grossmann, A. Tutin, 2003 ; M.-H. Svensson, 2004) montrent qu'il s'agit d'une donnée de base essentielle dans la description des langues, d'un fait économique pour le fonctionnement du système¹, problématique pour les descriptions disponibles et les méthodologies en application. Toutes les théories linguistiques l'intègrent d'une manière ou d'une autre. Les constructions figées, dont les variations internes sont très amples, sont nombreuses et demandent à être recensées. On est en présence d'un fait massif qui, bien qu'il ait complètement échappé à la tradition grammaticale, doit être considéré comme une propriété définitionnelle des langues naturelles. Le figement pose des problèmes spécifiques en analyse automatique des textes en langues naturelles. Les systèmes existants ne le prennent pas en compte de manière satisfaisante car on manque de méthodes précises pour le reconnaître.

Nous présentons ici le travail en cours d'élaboration sur le figement, notamment la réalisation d'une base de données qui a pour objectif de répertorier l'ensemble des locutions totalement figées, en s'appuyant sur des critères formels². Nous préciserons les critères permettant d'aboutir à un figement absolu avant d'aborder la question de son aspect relatif. Nous discuterons les difficultés d'application de la théorie du point de vue de l'encodage de ces locutions et des solutions théoriques qui ont permis de les dépasser. Nous analyserons plus amplement le phénomène du figement par le biais des locutions adverbiales. C'est en partant d'un bilan général des études portant sur ce phénomène que nous expliciterons ces caractéristiques et préciserons les perspectives que l'étude du figement ouvre à la recherche linguistique en général.

¹ « Ainsi peut-on réaliser le caractère extrêmement économique du figement : il participe à la formation d'unités polylexicales touchant tout le spectre catégoriel, fournit à la langue son outillage syntaxique et y ajoute une sorte de troisième articulation qui fait de toutes les unités du lexique d'une langue des morphèmes d'un type particulier susceptibles de donner de nouvelles unités polylexicales » (S. Mejri, 2003 : 8).

² L'équipe de recherche est formée par : Aude Grezka, Salah Mejri, José-Luis Meneses Lerin et Lichao Zhu.

1. Objectifs

Nous avons deux objectifs. Le premier objectif est de réaliser un inventaire exhaustif des séquences complètement figées, ce qu'on appelle traditionnellement « figement absolu » ou « figement total ». Le figement absolu représente une exception, il ne touche en moyenne qu'une expression sur dix. G. Gross (1996 : 16) affirmait que « les variantes sont plus fréquentes que le figement total ». Ces séquences sont versées dans la description des prédicats, des arguments et des actualisateurs dont elles font partie, description qui ne tient compte que de leur combinatoire externe exactement comme ce qui se passe avec les unités monolexicales. De cet objectif résulte nécessairement le deuxième qui est d'évaluer le degré de figement des autres locutions (c'est-à-dire le figement relatif). Est dit figement relatif tout figement qui admet au moins une variation relevant de la combinatoire libre. Énoncée ainsi, cette définition gagne à être affinée. Cette notion ne peut en effet se concevoir indépendamment de celle de degré. Il ne faut pas limiter la notion de degré uniquement à la dimension syntaxique. Une corrélation étroite avec le contenu sémantique fait que l'un va rarement sans l'autre. Le degré de figement d'une séquence dépend du nombre de critères qui la caractérisent (S. Mejri, 2005).

L'élaboration d'une grande base de données comportant tous les types de constructions figées (adverbiales, adjectivales, nominales et verbales) va ainsi permettre l'observation détaillée du phénomène de figement avec une visualisation, d'une part, du degré de figement à l'aide de critères formels et sémantiques, classés en fonction de leur pertinence ; d'autre part, des possibilités de variantes (dans le cadre du figement relatif). L'indication des variantes facilitera, aux usagers, aux traducteurs ainsi qu'aux systèmes de traduction automatique, l'identification et la traduction de ces locutions et permettra de comprendre les mécanismes qui mènent au figement absolu. La description des séquences figées doit faire partie intégrante de la description des items lexicaux de la langue. L'ensemble s'inscrit dans un continuum qui va du moins contraint au plus contraint. Cette description relève du programme général de la description de la langue. Qu'il soit polylexical ou monolexical, le prédicat se soumet, dans tous les cas de figure, à un ensemble de contraintes qui marquent son emploi dans la langue.

Bien évidemment, la question générale posée par le figement n'est autre que celle de la composition lexicale. Sont pointées en particulier les difficultés de définition du figement, notamment dans le cadre du figement relatif. Une séquence figée est une unité phraséologique constituée de plusieurs mots, adjacents ou non, qui présentent un certain degré de figement sémantique, lexical et ayant une fixité morphosyntaxique. Sur quels critères doit-on se fonder pour donner au syntagme le statut d'unité ? Peut-on toujours parler d'unité polylexicale pour les moules structurels sous-jacents à plusieurs paradigmes locutionnels (à la + nom ethnique / nationalité, etc.) ? La syntaxe de ces unités polylexicales est régulière et libre même

lorsqu'elles sont régies par des restrictions lexicales. Quelles séquences doivent figurer dans la base de données, et lesquelles doivent en être exclues ?

Notre analyse part de la constatation qu'une structure est d'autant plus libre que le nombre de relations linguistiques entre les différents composants est élevé (G. Gross, 1988). Quand il n'y a aucune relation syntaxique entre les différents éléments, nous disons que la structure est dite totalement figée. Les restrictions au niveau des relations potentielles entre les éléments d'un composé montrent que les paradigmes sont variables. C'est cette variabilité qui permet de parler de degré de figement d'une suite donnée.

2. Complexité syntactico-sémantique des locutions adverbiales

Nous avons fait le choix de commencer ce projet par la description des locutions adverbiales (à long terme, il sera bien évidemment question des locutions adjectivales, verbales et nominales). Ce choix est lié au fait que les adverbes, d'une manière générale, sont des éléments importants dans le traitement de la langue (C. Guimier, 1996 ; C. Molinier, F. Levrier, 2000) : ce sont des actualisateurs de prédicats et ils constituent un des critères fondamentaux dans la description des prédicats (ils permettent de caractériser et de distinguer les prédicats entre eux)³. Nous convenons de désigner sous l'appellation de *locutions adverbiales*, toutes les formes adverbiales non soudées présentant un certain «degré» de figement, quelle qu'en soit la présentation (G. Gross, 1988).

Ces locutions sont extrêmement difficiles à traiter. Elles constituent de loin la catégorie la plus hétérogène. Leur nombre est très élevé et leur diversité particulièrement grande. En effet, le nombre des adverbes composés est beaucoup plus important que celui des adverbes simples. Le recensement le plus complet est celui effectué par M. Gross. Dans son ouvrage *La syntaxe de l'adverbe* (1986), l'auteur met en valeur la conformité entre la formation interne des séquences figées et la syntaxe libre, et il préconise l'intégration de la description des séquences figées dans la description de la syntaxe générale. Il est nécessaire selon lui d'élaborer des

³ Les adverbiaux font partie des critères essentiels pour traiter, par exemple, la polysémie verbale. Il existe en effet pour chaque classe de verbes, des adverbiaux appropriés (A. Grezka, 2009 ; A. Grezka, F. Martin-Berthet, eds, 2007). Les verbes de *parole* admettent ainsi de nombreuses expressions qualifiant le niveau sonore : à voix haute/basse, à mi-voix, etc., à moins que cette caractérisation ne soit comprise dans le verbe (*crier, gueuler, murmurer*, etc.). Les verbes d'*ingurgitation* (du type *boire*) peuvent être modulés selon la quantité et/ou la rapidité (*à petites gorgées, à petits coups, cul sec*, etc.). De la même façon, les verbes de *coups* offrent des formes spécifiques pour exprimer des valeurs comme l'intensité (*violemment, brutalement*, etc.) ou l'itérativité (*à coups redoublés*, etc.).

tests spécifiques aux unités polylexicales décrites. Les grammaires traditionnelles masquent bien souvent leurs véritables régularités syntaxiques, en se fondant sur des critères hétérogènes qui ne distinguent pas les différents niveaux d'analyse.

Leur syntaxe mériterait une analyse beaucoup plus approfondie que celle que nous proposons ici mais notre classification n'a d'autre but que de présenter un canevas pour illustrer la diversité des locutions adverbiales. Nous retiendrons deux grandes propriétés syntactico-sémantiques. On constate premièrement, que ces locutions peuvent être prédicatives ou non prédicatives (G. Gross, 1988 : 108). On trouve ainsi :

- des locutions adverbiales de liaison, employées hors de la phrase simple ; elles ne font donc pas partie du schéma argumental qui les caractérise ; les informations qu'elles véhiculent prennent en charge diverses indications :
 - le domaine : *du point de vue philosophique*,
 - l'énonciation : *de toute évidence*,
 - la reformulation : *autrement dit, en d'autres termes*,
 - les marques de ponctuation du discours/du récit : *en premier lieu, d'entrée de jeu*,
 - la connexion : *en conséquence, au demeurant*;
- des locutions adverbiales qui sont des prédicats du second ordre, c'est-à-dire qu'elles opèrent sur d'autres prédicats ; ces adverbes sont appelés habituellement « adverbes de manière » ; le fait de les considérer comme des prédicats permet de ne pas rentrer dans des distinctions sémantiques fines entre moyen et manière, qu'il est difficile de justifier :

Luc marche (rapidement, de façon rapide).

Ces locutions ne sont pas actualisées par elles-mêmes. L'information est prise en charge par le prédicat de la principale.

Deuxièmement, parmi les locutions adverbiales, certaines ont un champ d'action très large, puisqu'elles s'appliquent à un grand nombre de prédicats. Ainsi, la locution *de manière correcte* peut caractériser tous les prédicats d'action :

Luc a écrit ce mot de manière correcte.

Luc s'est comporté de manière correcte.

À l'inverse, d'autres locutions ne sont possibles qu'avec un nombre restreint de prédicats et parfois même un seul :

Luc crie/chante à tue-tête.

Léa est tombée sur son ennemi à bras raccourcis.

La désignation très générale d'adverbes de manière a en fait détourné les chercheurs de la description des restrictions distributionnelles parfois très fines qui

caractérisent la relation de certains adverbes avec le verbe. Ces adverbes appropriés permettent le plus souvent de distinguer les modalités d'action (*parler à voix haute/basse, à mi-voix...*) des modalités d'état (*se porter comme un charme...*). Ces restrictions de cooccurrence ont été décrites de façon détaillée par M. Gross (1986). Le figement relève d'une description syntaxique et délimite des ensembles très diversifiés.

3. Présentation de la base de données

Actuellement, nous avons recensé plus de 5 000 entrées dont 1 811 entrées mots-vedettes différents (c'est-à-dire l'unité constituant le « cœur » de la locution, ayant un rôle privilégié du point de vue taxinomique)⁴. Dans cette base de données, l'utilisateur peut faire deux types de recherche : par « mot-vedette » ou par « construction » (fig. 1).



Figure 1. Page d'accueil de la base de données

En entrant un mot dans le moteur de recherche, l'utilisateur peut entre autre consulter :

- toutes les locutions adverbiales qu'il est possible de construire avec celui-ci :

⁴ Il est intéressant d'observer que les domaines les plus récurrents pour ces locutions sont ceux liés au temps et à la durée.

*d'un **coup** d'aile*
*d'un **coup** de baguette magique*
*par un **coup** du sort*
*à **coups** de pied au cul*
*sans **coup** f  rir*
 etc.

- toutes les compositions lexicales possibles de ce mot    l'int  rieur m  me de la locution adverbiale, comme les noms compos  s construits sur celui-ci ou les adjectifs qui lui sont appropri  s :

N compos  s : **coup** + de + (*aile, pied, vent, t  te, baguette magique...*)

- toutes les constructions dans lesquelles le mot va appara  tre :

Pr  p D  t N Pr  p N (*sur un coup de t  te*)
 Pr  p D  t N Pr  p N Adj (*d'un coup de baguette magique*)
 Pr  p D  t N (*sur le coup*)
 N Pr  p N (*coup sur coup*)
 Pr  p D  t Adj N (*pour un malheureux coup*)
 etc.

- les variations possibles : suppression, d  placement, etc.

L'utilisateur peut ainsi, en fonction de ses besoins, cibler sa recherche sur l'un ou l'autre des   l  ments. La recherche peut   tre orient  e aussi bien sur la structure interne de la locution et de sa repr  sentation (c'est-  -dire l'  tude des combinaisons lexicales qui forment la locution) que sur sa structure externe.

3.1. Les variations

Comme pour les autres cat  gories grammaticales, on constate un continuum entre les locutions adverbiales libres et celles qui sont totalement fig  es. Le figement n'est pas un ph  nom  ne absolu. L'int  r  t de son   tude consiste    en retracer le continuum,    en d  crire les degr  s et en donner une description qui va se focaliser sur les ph  nom  nes pr  cis impliqu  s par le figement dans le fonctionnement linguistique des locutions adverbiales. Il est n  cessaire de mettre au point des crit  res pr  cis permettant de d  cider si une suite donn  e doit entrer ou non dans la base et quel est son degr   de figement. L'  tablissement de ces crit  res a fait l'objet de nombreuses recherches (G. Gougenheim, 1971 ; M. Gross, 1982 ; S. Mejri, 1997). Ces crit  res n'ont certes pas tous la m  me port  e, mais ils vont tous dans le m  me sens : une construction est d'autant plus fig  e qu'elle a moins de propri  t  s de restructuration, de variation.

Pour mesurer la gradation dans le figement que connaissent les séquences polylexicales, il faut effectuer toutes les manipulations générales partagées par toutes les séquences et passer ensuite aux manipulations spécifiques à la partie du discours à laquelle appartient la séquence. C'est par le croisement des deux manipulations que nous pourrions mesurer le degré de figement de chaque séquence. Pour ce qui est des locutions adverbiales, un grand nombre d'entre elles acceptent des variations sur l'un ou l'autre des éléments. Dans l'état actuel du travail, on peut constater notamment :

- des changements de prépositions :

Nous avons rejoint Paris (de, en) une seule traite
À coups de pied (dans, à) le cul

- des changements de détermination ; la distribution est assez ouverte (article zéro, article défini générique, indéfini, partitif, possessif, numéral, etc.) :

En (deux, trois, cinq) coups de cuiller à pot

- des variantes lexicales :

À onze heures (sonnantes, tapantes)
À coups de pied dans le (cul, derrière)⁵

- des coréférences :

De (ma, ta, sa) plus belle plume

- des suppressions :

Dans l'intimité (la plus stricte)
D'un (seul) coup

- des changements de position :

À voix haute et intelligible/à haute et intelligible voix

- des variations morphologiques :

Par intermittence(s)

⁵ Ces variations sont souvent liées au changement de registre.

- des variations graphiques :

À la clé/clef
À fond/donf

- des traits d'union :

À la croque(-)au(-)sel

- des abréviations :

Au max/maximum

Nous partons du principe que le degré de figement peut être calculé à partir des résultats des tests mis en œuvre ci-dessus. Plus la séquence répondra négativement aux critères de variation, plus la séquence sera figée. On parle de figement total d'une suite lorsqu'aucun degré de liberté n'existe, c'est-à-dire lorsqu'aucun des éléments composant la chaîne de mots ne relève d'un choix et, par conséquent, ne peut faire l'objet d'un paradigme. Par exemple, une locution comme *à fond la caisse* n'accepte aucune substitution :

**À fond la (la voiture, la bagnole)*
 **À fond (ta, notre) caisse*

Ces suites fonctionnent en bloc comme des catégories simples et c'est à tort qu'on les fait figurer dans les dictionnaires sous l'un ou l'autre des termes. Elles devraient en effet constituer des entrées indépendantes.

Mais cette situation n'est pas la plus fréquente. Souvent, dans une séquence donnée, seul un sous-ensemble fait l'objet de figement. Une description linguistique fine doit rendre compte, pour une séquence donnée, de ce qui est figé et de ce qui ne l'est pas, surtout dans une perspective de traitement automatique. Dans une position donnée, les possibilités de commutation sont plus ou moins importantes : l'absence de paradigme est un cas rare⁶.

⁶ Les constructions libres sont caractérisées par des paradigmes permettant des substitutions définies par les contraintes d'arguments et par des modifications et des restructurations qui dépendent de la nature sémantique et syntaxique de la relation existant entre le prédicat et ses arguments. On peut ainsi calculer le nombre de variations potentielles pour une construction donnée. En revanche, par leur nature même, les séquences contraintes n'ont pas cette possibilité. Cependant, le figement peut être mis en évidence par le jeu du défigement. Cela consiste à ouvrir des paradigmes là où, par définition, il n'y en a pas. Le défigement s'observe de plus en plus dans la presse, qui s'en sert en vue de certains effets particuliers, destinés à attirer l'attention du lecteur. L'effet de surprise révèle la présence d'un figement.

Syntaxiquement, ce travail permet de reconsidérer la conception absolue qu'on a toujours eue de la syntaxe libre et de la syntaxe figée. En introduisant la notion de degré de figement, il est possible de corriger certaines conceptions qui ne tenaient pas compte de l'extrême complexité des questions de syntaxe. Partant de ce principe, il n'est plus envisageable de sceller les contraintes du côté du figement et la liberté combinatoire du côté de la syntaxe libre. La syntaxe figée est désormais perçue comme un continuum dans lequel s'inscrivent toutes les contraintes propres aux constructions libres et au figement. Ainsi, les séquences figées se distinguent des combinaisons libres par une plus grande solidarité entre les constituants, solidarité qui reste à mesurer pour tous les types de séquences.

3.2. Les constructions

Il est illusoire de rendre compte du statut des locutions adverbiales sans passer par la première étape de toute classification rigoureuse, à savoir : la description de toutes les constructions syntaxiques dans lesquelles figurent ces locutions adverbiales. Indépendamment de leur structure interne, les adverbes ont une syntaxe très hétérogène, que l'on ne peut qu'évoquer rapidement ici. Si l'on veut décrire le figement de façon intégrée et reproductible, il est impossible de se contenter d'observations générales sur le figement. En effet, pour reconnaître automatiquement les locutions adverbiales, il faut être en mesure de prédire leur morphologie et d'établir au préalable une typologie de la composition, afin de dégager les problèmes spécifiques à chaque type. En fonction de la structure interne des locutions adverbiales, les éléments de la description varient. La reconnaissance automatique des unités figées ayant une structure hétérogène dans un texte est évidente, ce qui n'est pas le cas d'une structure canonique.

La plupart des grammaires françaises limitent la description des locutions adverbiales à un petit nombre de classes morphologiques, telles que : PRÉPOSITION + NOM (*par intermittence*), PRÉPOSITION + DET + NOM (*contre toute attente*), etc. La nécessité de cette typologie a été reconnue depuis longtemps par divers auteurs. Des analyses plus précises ont permis de mettre en évidence la diversité des locutions adverbiales dans la langue générale. La typologie, de loin la plus explicite, est celle établie par M. Gross. Elle comprend 16 types de constructions⁷ (dans lesquelles sont réparties les 6 400 locutions verbales) et rend compte de l'ampleur du phénomène.

⁷ Les constructions sont les suivantes : Adv (*soudain*) ; Prép C (*en bref*) ; Prép Dét C (*contre toute attente*) ; Prép Adj C (*de sa belle mort*) ; Prép C Adj (*à gorge déployée*) ; Prép C de C (*en désespoir de cause*) ; Prép C Prép C (*des pieds à la tête*) ; Prép C Conj C (*en tout et pour tout*) ; Prép C de N (*au moyen de N*) ; Prép C Prép N (*par rapport à N*) ; Prép V W (*à dire vrai*) ; P (phrase figée) (*Dieu seul le sait*) ; (Adj) comme C (*comme ses pieds*) ; (V) comme C (*comme un cheveu sur la soupe*) ; (V) comme Prép C (*comme dans du beurre*) ; Conj C (*et tout le tremblement*).

Notre recherche sur les locutions adverbiales se situe dans le prolongement de ces travaux. Le recensement systématique et informatisé auquel nous procédons nous conduit en même temps à prendre en considération des formes plus originales et/ou plus complexes. Le nombre important de formes nous a amenée à proposer une nouvelle typologie qui peut être appliquée à la fois à la langue générale et aux lexiques spécialisés. Leur multiplicité (plus de 400 types de locutions recensées actuellement) a pour effet de rendre compte, dans le détail, de la structure morphologique des locutions, et cela dans un double but : d'une part, faciliter l'analyse des procédés compositionnels, d'autre part, permettre la saisie et le traitement informatique de classes formellement homogènes. Nous avons répertorié près de 250 moules de formation d'adverbiaux. On voit que le phénomène est productif et de grande ampleur. Nous donnons ici un petit extrait de cette typologie (fig. 2).

Prép Abrév	Prép Adv Card	Prép Dét Adj Adj N
Prép Abrév Abrév N	Prép Adv Conj Adv	Prép Dét Adj Adv
Prép Adj	Prép Adv N	Prép Dét Adj Adv Prép N
Prép Adj Adj N	Prép Adv Ponct	Prép Dét Adj N
Prép Adj Adv Adj N	Prép Adv Prép Adv	Prép Dét Adj N Adj
Prép Adj Conj Adj N	Prép Adv Prép Card	Prép Dét Adj N Dét N
Prép Adj Conj Prép Adj	Prép Adv Prép Dét N	Prép Dét Adj N P
Prép Adj Dét N	Prép Adv Prép N	Prép Dét Adj N Prép Dét N
Prép Adj N	Prép Adv Prép N Adj	Prép Dét Adj N Prép N
Prép Adj N P	Prép Adv Prép N x	Prép Dét Adj N Prép N N
Prép Adj N Prép N	Prép Adv Vinf	Prép Dét Adj Prép Dét N
Prép Adj N x	Prép Adv/N	Prép Dét Adj Prép N
Prép Adj Onamat	Prép Card	Prép Dét Adj V
Prép Adj Ord N	Prép Card Prép N	Prép Dét Adj x
Prép Adj Prép Adj	Prép Card N	Prép Dét Adv
Prép Adj Prép Dét N	Prép Card N Adj	Prép Dét Adv Adj
Prép Adj V	Prép Card N Dét N	Prép Dét Adv Adj N
Prép Adj Vinf	Prép Card N Pprés	Prép Dét Adv Adj N Prép N
Prép Adj/N	Prép Card N Prép Adv	Prép Dét Adv Adj P
Prép Adv	Prép Card N Prép Dét N	Prép Dét Adv Adj Prép Dét N
Prép Adv Adj	Prép Card N Prép N	Prép Dét Adv N
Prép Adv Adv	Prép Card Prép Card	Etc.
Prép Adv Adv N Vinf	Prép Dét Adj	

Figure 2. Extrait de la typologie des constructions adverbiales

Nous obtenons ainsi à partir du mot *titre*, les constructions qui sont présentées dans la fig. 3.

L'identification morphologique pose parfois problème. Le rattachement des locutions à un type déterminé est compliqué. Les catégories traditionnelles sont définies trop sommairement pour rendre compte des finesses de la langue, telles qu'elles se manifestent dans la composition. Par exemple, quel est le statut du mot *haut* dans *du haut de Poss personne* ou de *italienne* dans *à l'italienne* ? Nous avons fait le choix de tenir compte de la classe d'origine des éléments et de leur

fonction dans la locution. Ainsi, dans ces locutions, les mots *haut* et *italienne* seront codés *Adj=N*. Bien évidemment, cette typologie a un caractère exploratoire et expérimental. Il est important d’identifier toutes les structures réellement attestées, sans en dissimuler leur complexité. Cette phase est loin d’être achevée, ne serait-ce qu’en raison de l’explosion combinatoire des locutions adverbiales complexes. On pourra reconsidérer, ultérieurement, la grille de classement. Certaines redistributions seront à faire, pour consolider théoriquement la typologie et pour en améliorer l’efficacité pratique.

■ Accueil	Entrée	Mot vedette	Mot vedette 1	Construction
■ Equipe	à Dét autre titre	titre	autre titre	Prép Dét Adj N
■ Recherche dans la base par Mot vedette	à juste titre	titre	juste titre	Prép Adj N
	à n'importe quel titre	titre	n'importe quel titre	Prép LocAdj N
	à ce simple titre	titre	simple titre	Prép Dét Adj N
■ Recherche dans la base par Construction	à quel titre ?	titre	quel titre	Prép Dét N Ponct
	à un tel titre	titre		Prép Dét Adj N
	à ce titre	titre	[simple] titre	Prép Dét N
■ CONTACT	à titre de document	titre	titre de document	Prép N Prép N
	à titre documentaire	titre	titre documentaire	Prép N Adj
	à titre honorifique	titre	titre honorifique	Prép N Adj
	à titre onéreux	titre	titre onéreux	Prép N Adj
	sur titres	titre		Prép N
	à titre amical	titre	titre amical	Prép N Adj
	à titre bénévole	titre	titre bénévole	Prép N Adj
	à titre comparatif	titre	titre comparatif	Prép N Adj
	à titre consultatif	titre	titre consultatif	Prép N Adj
	à titre définitif	titre	titre définitif	Prép N Adj
	à titre divers	titre	titre divers	Prép N Adj
	à titre exceptionnel	titre	titre exceptionnel	Prép N Adj
	à titre expérimental	titre	titre expérimental	Prép N Adj

Figure 3. Les construction à partir du mot *titre*

Conclusion

À la lumière des premiers résultats obtenus par la base de données, il est clair que les constructions complètement figées sont en nombre relativement limité. Les variantes lexicales sont sans aucun doute les plus nombreuses. Les systèmes de détection automatique ne sont pas encore en mesure de trouver toutes les possibilités de variantes. Ce travail doit donc se faire manuellement au cours d’un processus qui permettra de disposer d’une grande base de données (qui pourra alors être automatisée). La consultation d’autres dictionnaires ainsi qu’une analyse en profondeur des nombreux documents qui reflètent la langue actuelle (corpus, web, etc.) pourrait augmenter le nombre de variations possibles.

Il y a donc lieu, d’une part, d’établir les descriptions systématiques de toutes les séquences figées, afin de les intégrer dans les dictionnaires électroniques élaborés pour le traitement automatique des langues ; d’autre part, de projeter ces diction-

naires sur des corpus textuels afin d'en déterminer le taux de couverture (S. Mejri, 2009) et d'exploiter ces résultats dans les différentes applications auxquelles le TAL peut donner lieu : extraction de données, recherche d'information, classement automatique de documents, aide à la rédaction, traduction automatique, etc.

L'élaboration de dictionnaires électroniques pour le traitement automatique des langues est un contexte applicatif qui permet de rendre valide les options théoriques d'une modélisation linguistique. La langue doit être représentée de façon formalisée pour la structuration des bases de données : les représentations doivent être explicites et normalisées. Cette contrainte de formalisation implique une théorie linguistique qui satisfait les critères scientifiques de cohésion, de reproductibilité et de vérifiabilité. L'état actuel du traitement automatique des langues naturelles n'est pas tel qu'on puisse parler de systèmes de compréhension de textes qui seraient à même de réaliser des tâches complexes. Le traitement automatique dépend du degré de couverture dont le système dispose. La reconnaissance des unités lexicales implique donc qu'on ait fait le recensement de toutes les suites figées.

Références

- François J., Mejri S., eds., 2006 : *Composition syntaxique et figement lexical*. Presses Universitaires de Caen, Coll. Bibliothèque de Syntaxe & Sémantique, 280 p.
- Gougenheim G., 1971 : *Étude sur les périphrases verbales de la langue française*. Paris, Nizet, 383 p.
- Greška A., 2009 : *La polysémie des verbes de perception visuelle*. Paris, L'Harmattan, Coll. Sémantiques, 292 p.
- Greška A., Martin-Berthet F., eds, 2007 : « Verbes et classes sémantiques ». *Verbum*, 1, [Presses Universitaires de Nancy], 172 p.
- Gross M., 1982 : « Une classification des phrases figées du français ». *Revue Québécoise de Linguistique*, 11(2), 151—185.
- Gross M., 1986 : *Grammaire transformationnelle du français*. 3. *Syntaxe de l'adverbe*. Paris, ASSTRIL.
- Gross G., 1988 : « Degré de figement des noms composés ». *Langages*, 90, 57—70.
- Gross G., 1996 : *Les expressions figées en français. Noms composés et autres locutions*. Paris, Ophrys, Coll. L'essentiel français, 161 p.
- Grossmann F., Tutin A., 2003 : *Les collocations : analyse et traitement*. Amsterdam, De Werelt, 144 p.
- Guimier C., 1996 : *Les adverbes du français. Le cas des adverbes en -ment*. Paris, Ophrys, Coll. L'essentiel français, 170 p.
- Mejri S., 1997 : *Le figement lexical : descriptions linguistiques et structuration sémantique*. Tunisie, Publications de la Faculté de lettres de la Manouba, 632 p.
- Mejri S., 2003 : « Le figement lexical ». *Cahiers de Lexicologie*, 82, 23—39.

- Mejri S., 2005 : « Figement absolu ou relatif : la notion de degré de figement ». *Linx*, **53**, [Université Paris X Nanterre], 183—196.
- Mejri S., 2009 : « Le mot, problématique théorique ». *Le Français Moderne*, **77** (1), 68—82.
- Molinier C., Levrier F., 2000 : *Grammaire des adverbes. Description des formes en -ment*. Droz, Genève—Paris.
- Svensson M.-H., 2004 : *Critères de figement. L'identification des expressions figées en français*. Umeå, Umeå universitet.