

**Krzysztof Bogacki**

*Université de Varsovie*

## **Le traitement de l'article dans la traduction automatique du polonais vers le français**

### **Abstract**

Krzysztof Bogacki discussed the problems connected with the translation of articles by automatic translation programs. He analyzed three possible situations: a translation from an article-based language into another article-based language, a translation from an article-based language into a language with no articles as well as a translation from a language with no articles into an article-based language. The author, who based his analysis on the available on the Internet programs like Systran, PROMT or Kompas, discovered that although the translations from an article-based language into another language were satisfactory, the translations from a language with no articles into an article-based language were characterized by lexical inaccuracy and misuse of articles. The author also surveyed publications devoted to the use of articles (written by: Gniadek, Wilczyńska, Karolak and Nowakowska, Blanco and Buvet) but did not find there any solutions for the improvement of automatic translation programs, which are unable to analyze pragmatic and semantic factors in order to use a correct article. Finally, the author attempted to introduce certain modifications to Systran so as to improve its translating effectiveness.

### **Keywords**

Article, usage, nominal syntagm, translation, automatic translation program, declension, semantic features, aspect

On s'accorde volontiers pour reconnaître qu'il y a peu de travaux théoriques sur le traitement des articles en traduction automatique. Pourtant dans ce domaine les lacunes et les imperfections dans le fonctionnement des logiciels sont nombreuses et variées. Dans cet article, nous allons en donner quelques exemples avant de nous interroger sur la façon de traiter ces mots dans le processus de traduction d'une langue source sans articles vers une langue cible à articles. Nous allons discuter ces problèmes en relation avant

tout avec le traducteur polonais-français de Systran dont nous nous sommes occupés dans le cadre d'un projet européen financé par la Commission Européenne<sup>1</sup>.

Il est aisément de découvrir des différences dans l'emploi des articles entre les langues telles que l'anglais, le français, l'italien ou l'espagnol. Ainsi l'article est omis en anglais dans les phrases génériques devant un nom en position sujet alors qu'en français dans les phrases équivalentes on trouve un article défini. On emploie l'article zéro, dans les mêmes phrases en position d'attribut en anglais tandis que le français recourt à l'article indéfini. Des différences sont visibles aussi dans d'autres positions syntaxiques et avec d'autres paires de langues :

*Ships travel four or five days from Southampton to New York.*

*Les bateaux font le voyage entre Southampton et New York en quatre ou cinq jours.*

*Horses are animals.*

*Les chevaux sont des animaux.*

*Juan juega (a + al) fútbol.*

*Jean joue (\*à + au) football.*

La situation est bien plus compliquée si l'on compare une langue sans articles (p.ex. le polonais ou le russe) et une langue à articles (p.ex. l'anglais ou le français). Or l'écart entre les systèmes linguistiques observables sous ce rapport fait que les erreurs de traductions commises par les traducteurs automatiques sont plus nombreuses que dans le cas des opérations faites sur deux langues à articles (anglais-français, français-espagnol, anglais-italien, espagnol-italien etc.). Nous allons examiner trois cas de figure (pour les textes continus aussi bien que pour les phrases isolées) :

- traduction entre deux langues à articles,
- traduction à partir d'une langue à articles vers une langue sans articles,
- traduction à partir d'une langue sans articles vers une langue à articles.

Considérons pour commencer les traductions d'un texte continu. Parmi différents traducteurs accessibles sur Internet<sup>2</sup> nous avons choisi celui de la

<sup>1</sup> Project MATCHPAD (MAchine Translation System for the TChek, Hungarian and Polish Administrations), CE-12256, III. 2000–VIII. 2002. Le logiciel n'a pas encore atteint le seuil de la commercialisation restant à l'étape de prototype.

<sup>2</sup> The European Association for Machine Translation recense périodiquement traducteurs automatiques, dictionnaires et autres utilitaires d'aide à la traduction dont le nombre, déjà impressionnant, ne cesse de croître. Cf. W.J. Hutchins (2003).

société PROMT qui permet de tester sur son site<sup>3</sup> plusieurs paires de langues à articles. Le texte ci-dessous :

*A camera is a piece of equipment used for taking photographs. The camera lets in the light from an image in front of it and directs the light onto photographic film. The light has an effect on the chemicals which cover the film and forms a picture on it. When the film is developed it is washed in chemicals which make the picture permanent. It is then possible to print the picture onto photographic paper.*

a été traduit en espagnol de la façon suivante :

*Una cámara es un pedazo de equipo usado para tomar fotografías. La cámara deja entrar la luz de una imagen delante de ello y dirige la luz en la película fotográfica. La luz tiene un efecto sobre los productos químicos que cubren la película y forma un cuadro sobre ello. Cuando la película es desarrollada es lavado en productos químicos que hacen el cuadro permanente. Es entonces posible imprimir el cuadro en el papel fotográfico.*

À une exception près (maladresse dans la traduction de *a piece of equipment* par *un pedazo de equipo*), le résultat est irréprochable au niveau du traitement des articles. Le même degré d'adéquation est atteint avec d'autres paires de langues à articles. La tâche n'est pas difficile non plus lorsqu'on traduit d'une langue à articles vers une langue sans articles. Le même texte anglais traduit en russe<sup>4</sup> donne le résultat suivant :

*Камера – часть оборудования, используемого для того, чтобы брать фотографии. Камера позволяет на свету от изображения перед этим и направляет свет на фотографический фильм. Свет имеет эффект на химикалии, которые закрывают фильм, и формирует картину на этом. Когда фильм развит, это вымыто в химикалиях, которые делают картину постоянной. Тогда возможно напечатать картину на фотографическую бумагу.*

Lors du premier essai effectué en juin 2003, on a relevé des erreurs qui témoignent des lacunes au niveau lexical (les mots *developed*, *paper* ont été laissés sans traduction). Elles ont disparu quelques mois après, une fois les dictionnaires enrichis. Sur le plan des déterminants aucune imperfection ne peut être signalée. Il en va autrement lorsqu'on traduit du russe en anglais.

<sup>3</sup> Cf. [http://www.translate.ru/text.asp#tr\\_form](http://www.translate.ru/text.asp#tr_form).

<sup>4</sup> PROMT ne permet pas de travailler sur le polonais.

Le résultat obtenu, outre les maladresses lexicales, présente des défauts montrant que le passage d'une langue sans articles vers une langue à articles est une tâche ardue :

*The chamber – a part of the equipment, used to take photos. The chamber allows on light from the image before it and directs light on photographic film. Light has effect on himikalii which close film, and forms a picture on it. When film is advanced, it is washed up in himikalijah which do a picture of a constant. Then it is possible to print a picture on a photographic paper.*

La traduction du russe vers le français, faite avec le même logiciel, aboutit au texte que voici :

*La chambre – la partie de l'équipement utilisé pour que prendre de la photo. La chambre permet sur la lumière de la représentation devant cela et dirige la lumière pour le film photographique. La lumière a l'effet pour les agents chimiques, qui ferment le film, et forme le tableau sur cela. Quand le film est développé, cela est lavé dans les agents chimiques, qui font le tableau de la constante. Alors il est possible d'imprimer le tableau pour le papier photographique.*

Si l'on fait abstraction des erreurs et des lacunes lexicales – elles sont frappantes surtout dans le texte français mais ne manquent pas dans la traduction anglaise (p.ex. absence inattendue de traduction de *химикалии*) – on s'aperçoit que le problème du traitement des articles ne se fait pas sans heurts. Ainsi on trouve au début du texte *the chamber* au lieu de *a chamber*, l'article défini est omis devant *light* dans la deuxième phrase, devant *image* on trouve *the* à la place de *a* etc. La comparaison de la traduction avec le texte modèle élaboré par un humain fait voir que sur 19 cas d'emplois d'articles, 5 à peine sont conformes aux attentes d'un sujet natif.

Parmi les langues sans articles, nous avons aussi testé le polonais. Voici la traduction fournie avec un programme de la société polonaise Kompas. Le texte anglais a été traduit par :

*Aparat fotograficzny jest kawałkiem. Aparat fotograficzny wpuszcza {pozwala w} światło od obrazu przed tym i kieruje światło na film. Światło ma skutek na chemikaliach, które pokrywają film i tworzą obraz {zdjęcie} na tym. Kiedy film jest rozwinięty, to jest umyte w chemikaliach, które robią obraz {zdjęcie} trwałym. To jest wtedy możliwe, by wydrukować obraz {zdjęcie} na papier fotograficzny wyposażenia używanego dla brania fotografii.*

Après les corrections qui s'imposaient<sup>5</sup>, le texte traduit du polonais en anglais a pris la forme suivante :

*The camera is an element of the equipment used to photographing. The camera lets in the reflected light from the picture {image} being found before him and directs the light on the film. The light affects chemicals which cover the film and create on him the removal {photo}. When the film is called out, he is bathed in chemicals which fix the removal {photo}. Possible is then printing of taking off paper photographic.*

La traduction du polonais vers le français faite avec un autre logiciel de la même société donne un résultat à peine lisible :

*Appareil photographique est élément de l'équipement usité à photographier. Appareil photographique **wpuszcza** la lumière tirée du tableau se trouvant avant lui et dirige la lumière sur le film. La lumière agit sur **chemikalia** qui couvrent le film et créent sur lui la photographie. Quand le film est provoqué, est il baigné dans **chemikaliach**, qui perpétuent la photographie. Possible est alors **wydrukowanie** les photographies sur le papier photographique.*

Les résultats sont tout aussi décevants lorsqu'on soumet à l'essai des phrases isolées. Comme pour les textes continus, ils sont meilleurs lorsque la langue source et la langue cible disposent des articles. Les traducteurs testés se tirent plus ou moins bien d'affaire même là où on note des différences dans ce domaine d'une langue à l'autre. La situation change lorsqu'on envisage une langue sans articles et une langue à articles. On chercherait en vain une uniformité et une cohérence parfaite dans le traitement des articles. Ainsi le traducteur Systran anglais-russe<sup>6</sup> dans certaines phrases fait correspondre à *some* la forme *некоторый* alors que PROMT hésite entre *немного* et *некоторое*. Le traducteur humain l'omettrait l'une et l'autre forme bien évidemment, cf. :

*I ate some bread and drank some wine.*

- *Выпивший я съел некоторый хлеб и некоторому вину.* (Systran)
- *Я съел немного хлеба и выпил некоторое вино.* (PROMT)

<sup>5</sup> *Aparat fotograficzny jest elementem wyposażenia używanym do fotografowania. Aparat fotograficzny wpuszcza światło odbite od obrazu znajdującego się przed nim i kieruje światło na film. Światło działa na chemikalia, które pokrywają film i tworzą na nim zdjęcie. Kiedy film jest wywoływany, jest on kąpany w chemikaliach, które utrwalają zdjęcie. Możliwe jest wówczas wydrukowanie zdjęcia na papierze fotograficznym.*

<sup>6</sup> <http://www.systransoft.com>.

Lors de la traduction en sens inverse, avec Systran on ne trouve aucun article devant le mot *bread* (le mot *vin*, pourtant très fréquent, est ignoré !) tandis que PROMT tantôt omet l'article, tantôt insère un article défini à la place de l'article partitif attendu :

*Я выпил вино и съел хлеб.*

→ *I sawed vino and ate bread.* (Systran)

→ *I have drunk ???o and have eaten bread.* (PROMT)

*Я ем хлеб и я пью воду.*

→ *Je mange le pain et je bois l'eau.* (PROMT)

Au temps passé, la même phrase traduite en anglais fait voir que l'insertion d'article défini n'est pas un fait constant. On trouve en effet :

*Я ел хлеб и пил вино.*

→ *I eat bread and saws a wine.* (PROMT)

*Я съел хлеб и выпил молоко.*

→ *I have eaten bread and have drunk milk.* (PROMT)

Le traducteur polonais-anglais développé par l'équipe de K. Jassem<sup>7</sup> laisse supposer que l'emploi d'articles est lié aux substantifs et apparaît ainsi comme une affaire de lexique plutôt que de syntaxe. Ainsi dans les phrases testées, le mot *wino* 'vin' est accompagné de l'article défini quel que soit le contexte (phrase affirmative ou phrase négative) alors que les noms d'autres boissons (*eau*, *thé*, *jus de fruit*, *lait*) sont employés sans article :

*Zjadłem chleb i wypilem wino.*

→ *I ate bread and I drank the wine up.*

*Jadłem chleb i piliem wino.*

→ *I was eating bread and I was drinking the wine.*

*Lubię (mleko + wino + masło).*

→ *I like (milk + the wine + butter).*

*Pil (wino + mleko + woda + herbatę + sok owocowy).*

→ *He was drinking (the wine + milk + water + tea + fruit juice).*

*Nie pil wina.*

→ *He wasn't drinking the wine.*

*(Wino + woda) jest napojem.*

→ *(The wine + water) is the drink.*

Le dernier concurrent en lice, le traducteur de la société polonaise Kompas, préfère insérer un article défini dans le cas de toutes les boissons, y compris le vin :

<sup>7</sup> <http://www.ceti.pl/poleng/zasoby/tlumaczenie/english.html>.

*Zjadłem chleb i wypiłem (mleko + wodę + herbatę + sok owocowy + wino).*  
 → *I ate the bread and I drank the (milk + water + tea + fruit juice + wine).*

Il découle des observations ci-dessus que les difficultés éventuelles dans le traitement des articles sont facilement surmontées si la traduction se fait entre deux langues à articles. Dans le cas contraire, seul le passage d'une langue à articles vers une langue sans articles se fait pratiquement sans problèmes. La situation devient dramatique si la langue source ne dispose pas d'articles alors que la langue cible, au contraire, est une langues à articles. Le verdict est sévère : aucun des traducteurs décrits n'est pleinement satisfaisant, certains n'ont pas atteint, à notre avis, le niveau de la commercialisation avant tout à cause de la mauvaise gestion des articles et des lacunes lexicales.

Dans la suite de cette communication nous allons nous interroger sur les mesures à prendre pour améliorer le traitement de l'article<sup>8</sup>. À première vue le problème consiste à sélectionner la bonne forme dans un jeu de 19 d'unités (*un, une, à un, à une, d'un, d'une, le, la, l', de la, du, les, des, à des, de l', à la, à l', aux, au*). Cependant la liste des choix possibles se réduit vite si l'on retranche, premièrement, les formes combinées avec une préposition (contractées ou non – leur traitement, pour la partie prépositionnelle, serait géré par la composante syntaxique), deuxièmement, si l'on fait abstraction du facteur morphologique (opposition forme élidée vs forme non-élidée) qui serait pris en charge par une banale règle d'élation et, troisièmement, si l'on confie la sélection de la forme féminine / masculine de l'article au module lexical. En fin de parcours on reste avec une opposition entre article défini, article indéfini et article partitif ce qui n'est simple qu'en apparence. Les difficultés liées au traitement de ces formes viennent sans doute du fait que la spécificité de la traduction automatique consiste en ce que l'on ne peut pas recourir à des facteurs sémantiques et pragmatiques sans contrepartie formelle qui pourtant sont souvent mis en avant lors de la description et de l'interprétation des différents emplois des articles en linguistique théorique. En effet, on parle souvent d'intention du locuteur qui dans son texte peut viser l'ensemble dénoté par le substantif (article défini) ou un seul objet qui, à son tour, peut être considéré comme identifié (article défini) ou non-identifié (article indéfini), on invoque le savoir partagé par les participants du dialogue ou, enfin, on fait appel à différents facteurs pragmatiques qu'on peut difficilement relier à la présence d'indices formels dans le texte source.

<sup>8</sup> Nous laissons de côté les autres déterminants : adjectifs démonstratifs, possessifs, quantificateurs de toute sorte (*beaucoup de, plusieurs, trois, un tas de, ...*), comparatifs (*autant de ... que, plus de ... que*), intensifs (*très, ...*) parce qu'ils ont des équivalents lexicaux dans la langue cible et de ce fait sont inclus dans le dictionnaire de transfert. Cf. X. Blanco, P.-A. Buvet (1999).

Examinons trois exemples. Le plus décevant est sans aucun doute S. Gniadek (1979) qui se contente de relever purement et simplement qu'il existe des articles en français et qu'il n'y en a pas en polonais ce qui est une source de difficultés et une cause d'erreurs fréquentes. Il essaie d'associer à chacun des articles français les valeurs sémantiques appropriées : il parle de «l'étendue générale du nom, de la sélection déterminée et de marque d'une portion quantitativement indéterminée», de la «sélection déterminée / indéterminée» ou d'«une portion quantitativement indéterminée» (ibidem : 74). On chercherait en vain chez lui une liste d'équivalents polonais des articles français. Or le problème auquel on est confronté consiste à s'aider d'un tel répertoire pour découvrir l'article approprié qui doit être inséré devant un lexème substantival.

W. Wilczyńska (1989) reste, elle aussi, au niveau des valeurs sémantiques et pragmatiques que l'on peut attribuer à des phrases françaises contenant des substantifs (ou syntagmes nominaux) avec article. Elle arrive à la conclusion que «le critère fondamental de l'emploi de l'article défini et de l'article indéfini correspond à la façon de considérer l'objet sous le rapport de sa familiarité pour les deux interlocuteurs, le choix de l'article ou sa lecture dépendant essentiellement de leur décision à cet égard» (ibidem: 18). Les observations concernent en fait différentes interprétations que l'on peut associer aux groupes nominaux déjà comportant un article. Or lors de la traduction du polonais vers une langue à articles, on doit, au contraire, parcourir le chemin inverse pour sélectionner la bonne forme dans un ensemble d'unités. Le problème consiste donc à délimiter et à décrire en termes formels les situations où l'on doit recourir à telle forme polonaise plutôt qu'à telle autre.

S. Karolak et M. Nowakowska (1999), prenant le contrepied de Wilczyńska, partent de l'idée que le choix de l'article est déterminé par deux types de facteurs contextuels (le lexème substantival lui-même et le contexte syntaxique dans le cadre de la phrase). Force est de reconnaître cependant que les règles qu'ils formulent contiennent parfois des informations qui relèvent de l'interprétation du texte par le sujet parlant. Ainsi il est question des noms «employés avec une intention identificatrice» («nazwa [...] użyta zgodnie z intencją identyfikacyjną», p. 170). La règle 11 vise «les noms propres généraux au pluriel employés avec une intention particularisante» («nazwy własne ogólnie w liczbie mnogiej użyte z intencją szczególną») etc. Or ces informations pourraient être utilisées pour l'élaboration des règles exploitables par le traducteur automatique à condition de trouver dans le texte des marques formelles qui correspondent à des valeurs sémantiques et pragmatiques ci-dessus.

La même idée de lien entre lexème substantival et type d'article utilisé se retrouve dans la conception du traitement des déterminants de X. Blanco et P.-A. Buvet (1999). Le point de départ pour leur proposition est l'idée que

les substantifs appartenant à une des deux classes syntaxiques disjointes : celle des arguments élémentaires et celle des prédictats. Or l'observation de ces mots fait voir l'existence de nombreuses contraintes relatives à la détermination :

*Luc a eu (la + une) grippe.*  
*Luc a eu (le + \*un) tétanos.*  
*Luc a eu (\*la + une) variole.*

Il est donc clair que l'information sur le type de déterminant susceptible de figurer avec le substantif décrit lorsque celui-ci est employé dans les phrases doit être contenue dans le dictionnaire. Avec le formalisme adopté par X. Blanco et P.-A. Buvet, l'entrée *bronchiolite* comportera l'information que le mot est compatible avec DDEF\_CE\_LE\_Modif, DGEN\_LE1, DIND\_UN\_UN\_Modif, DQUA\_DQUA1, DPAR\_DUI, *hépatite* se caractérisera par le profil suivant: DDEF\_CE\_LE\_Modif, DGEN\_LE1, DIND\_UN+UN\_Modif, DQUA\_DQUA1 alors que *asthme* par DDIV\_une crise d', DPAR\_DUI1. Au niveau de la phrase on retrouvera une combinaison d'un prédictat avec ses arguments :

$P \rightarrow \text{Préd (Arg)}$

ce qui donne pour la phrase *Un enfant a mangé plusieurs pommes* le schéma :

$P \rightarrow \text{manger (enfant, pomme).}$

Ayant spécifié les configurations des arguments (en tant que sujet et complément d'objet du verbe *boire*) au point de vue de leurs déterminants, on peut générer des phrases par l'opération de linéarisation et d'actualisation.

On s'aperçoit que dans ce système il est nécessaire de disposer d'une typologie des noms prédictifs couplée aux différents verbes supports acceptés par ceux-ci et fonctionnant en accord avec d'autres modules incluant des informations supplémentaires indispensables pour la linéarisation. Il faut dire que suivre cette proposition reviendrait à remanier profondément l'architecture du traducteur. Nous chercherons donc dans celle qui a été adoptée par Systran des solutions utilisées dans les cas où la langue source ne dispose pas d'équivalent direct nécessaire pour figurer dans la langue cible.

Il y en a plusieurs. Examinons en premier lieu le problème des cas. On sait bien qu'à l'absence des oppositions casuelles en français moderne correspond en polonais une déclinaison. Dans le processus de la traduction, le choix de la forme flexionnelle est conditionné par un module morphologique dont le rôle est de générer les formes appropriées en fonction des informations syntaxiques contenues dans un module syntaxico-lexical polonais qui impose

le cas grammatical à chacun des arguments fléchis présents dans la phrase. Ainsi avec les verbes, si l'argument sujet est toujours au cas nominatif, celui des compléments est indiqué de façon explicite. On choisit entre le génitif, le datif, l'accusatif, l'instrumental, le locatif ou le vocatif prenant soin de spécifier aussi la conjonction et la préposition utilisées pour introduire compléments indirects et circonstants de toute sorte.

D'un autre côté les arguments obligatoires du verbe sont caractérisés par les traits sémantiques (ils sont plus d'une vingtaine, entre autres [+Humain], [+Animé], [+Concret], [+Liquide], [+Abstrait] etc.) et doivent donc se retrouver dans un des dictionnaires dont dispose le programme. Cette information est indispensable aussi parce qu'il existe des mots qui se déclinent de façon différente selon qu'ils sont [+Humains] ou [+Concrets] :

*kontroler* : Acc. *kontroler* [−Humain, +Concret] vs. *kontrolera* [+Humain]

On voit donc que dans ce genre de situation la solution est relativement facile à trouver et consiste, en gros, à donner une liste de formes fléchies (elle est contenue dans le module morphologique regroupant les substantifs, les adjectifs et autres éléments nominaux) et à définir, dans le module verbal, la forme flexionnelle à choisir.

Les oppositions aspectuelles propres au polonais constituent un autre point de divergence entre l'anglais et le français, d'un côté, et le polonais de l'autre. Elles sont traitées de façon différente. Sans améliorations ci-dessous la phrase *He bought a new vehicle* est traduite par *On kupował nowy samochód* au lieu de *(On) kupił nowy samochód* ce qui est inacceptable sur le plan stylistique. Les traductions que nous venons de mentionner s'expliquent par le fait que le traducteur consulte le dictionnaire bilingue qui, dans l'idéal, stocke pour un verbe anglais (ou français) toutes les formes aspectuelles admises. En l'absence d'algorithme sélecteur d'aspect, tout est donc laissé au hasard. La solution du problème a la forme d'un algorithme en 4 étapes qui prend la forme que voici.

1. Tout d'abord nous construisons une liste de correspondances sous forme de table qui contient : le verbe imperfectif, son équivalent anglais (ou français), une liste des verbes polonais perfectifs. Il est crucial de s'assurer que les verbes anglais (français) dans le dictionnaire de transfert sont toujours traduits par les verbes polonais imperfectifs considérés comme équivalents par défaut.

2. Lors du transfert de l'anglais / du français vers le polonais, la machine consulte la liste des correspondances mentionnée plus haut et mémorise deux formes : l'original anglais (français) et son équivalent polonais par défaut.

3. Nous introduisons des règles qui associent un temps grammatical particulier en anglais (en français) à un aspect polonais donné. En vertu de ces règles, la machine traduira le *present perfect continuous* et le *past continuous*

(*I was doing, I have been doing, I had been doing, I have been living here for 5 years*) par la forme polonaise imperfective du passé (*robilem* ou *zrobilem*), les temps passés à la forme non-progressive (*I did, I have done, I had done*) seront rendus par la forme perfective du passé (*zrobilem*), les temps du futur à la forme progressive par la forme imperfective future (*I will be doing, I will have been doing* – *będę robić, będę robił*) et les temps du futur à la forme non-progressive par les formes perfectives du futur (*I will do, I will have done* – *zrobię*).

Les associations pour la paire: français-polonais seront différentes. Le temps imparfait (*je faisais*) est traduit par la forme polonaise imperfective du passé (*robilem*) et les autres temps par la forme polonaise perfective.

À la fin du traitement, dans le module de synthèse, il est demandé au programme de chercher la forme perfective. Si elle est repérée, la machine sélectionne la forme perfective dans la liste des correspondances aspectuelles. Si l'on retrouve, au contraire, la forme imperfective, on passe à l'étape suivante.

L'algorithme est basé sur l'hypothèse que les temps grammaticaux anglais / français ont, chacun, une traduction privilégiée à l'aide d'un temps grammatical polonais et d'une forme lexicale appropriée. Cette hypothèse est justifiée par les observations faites pour différentes paires de langues. Les exceptions – très nombreuses pourtant! – sont à traiter par des règles d'exception qui, dans la plupart des cas tiennent compte du contexte dans lequel apparaît le verbe. Il s'agit le plus souvent de la présence d'adverbes – fréquentatifs, duratifs, semelfactifs etc. – qui, ayant priorité sur les règles générales de sélection, imposent la forme aspectuelle appropriée corrigent le choix primaire résultant des règles générales. Ainsi la phrase:

*Il est souvent venu.*

est traduite par une forme inaccomplie (*często przychodził*) en dépit de la présence du passé composé *il est venu* qui suggère la traduction par l'accompli polonais (*przyszedł*).

Une procédure semblable pourrait être imaginée pour traiter l'article. Elle aboutirait sur le plan informatique soit à une structure de type **case** avec plusieurs cas soit à une structure de type **if... (then)** à enchâssements multiples. Or pour la réaliser, on aurait besoin de deux types d'informations. Les unes auraient un caractère syntaxique et permettraient d'identifier les fonctions phrastiques, critiques dans le cas de la sélection des articles. En effet, les règles formulées par S. Karolak et M. Nowakowska (1999) font souvent appel à différentes positions syntaxiques pour expliquer l'emploi de telle ou de telle autre forme. Il serait indispensable d'un autre côté de disposer d'informations lexicales qui viendraient enrichir les dictionnaires existants. On y retrouverait des renseignements analogues à ceux proposés par P.-A. Buvet et par X. Blanco.

Au départ on prévoit une règle de placement de l'article défini. Dans le traducteur Systran, elle a la forme suivante:

## Synthesis Rule (Level: S\_Fr\_f): SR\_Art\_Insert

```

void add_article (Word *N)
{ if (VAL(N,f_noun))
{
  if (VAL(N->previous,f_pro)) return; if (VAL(N,r_modified_by_adj_left))
{
  Word *ADJ=WORDREF(VAL(N,r_modified_by_adj_left));
  Fields* F=new syntax_FR_DET;
  Word* NEW=new Word(*F);
  SET(NEW,v_category,f_det);
  NEW->lemma="le";
  InsertBefore(ADJ,NEW);
  SET(NEW,v_gender,VAL(N,v_gender));
  SET(NEW,v_number,VAL(N,v_number));  }

  if (!VAL(N,r_modified_by_adj_left))
  {
    Fields* F=new syntax_FR_DET;
    Word* NEW=new Word(*F);
    SET(NEW,v_category,f_det);
    NEW->lemma="le";
    InsertBefore(N,NEW);
    SET(NEW,v_gender,VAL(N,v_gender));
    SET(NEW,v_number,VAL(N,v_number));
  }
}
}

```

Elle fait partie d'une famille de règles d'insertion qui gèrent différents phénomènes : insertion de la préposition *de* entre deux substantifs voisins pour traiter la structure polonaise  $N + N_{G\acute{e}n}$  (SR\_de\_Insert), insertion de *pas* dans les phrases négatives (SR\_pas\_neg\_Insert) etc.

À l'étape suivante, l'article de départ serait remplacé, dans certaines conditions, par un indéfini ou par un partitif. À ce moment-là, la tâche consiste à multiplier les descriptions de contextes correspondant à chaque choix. Elles doivent être faites en termes syntaxiques (position dans la phrase: attribut, complément de nom, type de syntagme nominal etc.) et lexicaux (présence de traits spécifiques fournis par les profils lexicaux des substantifs et des verbes) afin de les relier avec l'insertion d'articles précis. Il est entendu que l'insertion ne visera que les noms dont le profil ne contient pas le trait [+ Nom propre] qui par défaut est employé sans article. Les rares cas de présence de l'article

avec les noms de ce type seraient traités par les règles syntaxico-lexicales (*acheter un Picasso, jouer du Bach*) ou au niveau du dictionnaire de transfert (*Le Caire – Kair*).

On peut se faire une idée de ce que seront les traits lexicaux en prenant comme point de référence les règles utilisées par S. Karolak et M. Nowakowska (1999). On y trouve des caractéristiques telles que : noms de personnes, noms de localités, noms de petites îles en Europe, noms masculins des îles éloignées, noms de jours de la semaine. Le trait [ $\pm$ Massif], lui, permettra de gérer l'emploi du partitif. La présence de [+Massif] dans le profil lexical d'un substantif entraînera l'apparition de l'article partitif. La recatégorisation des substantifs de masse en substantifs comptables se traduirait par l'insertion d'un numéral devant un substantif de masse (cf. *boire un whisky, deux bières, ...*). Au contraire, en présence du verbe *manger*, on recatégoriserait le substantif comptable *homme blanc* en substantif de masse en le faisant précéder de l'article partitif (*manger de l'homme blanc*).

Il est clair qu'on n'arrivera pas facilement à régler le problème des articles. D'un côté, la description des contextes dans lesquels ils apparaissent nécessite un nombre élevé de règles<sup>9</sup>. D'un autre côté des moyens plus sophistiqués seraient nécessaires pour gérer les situations où a priori, dans un contexte identique, différents types d'article sont possibles provoquant différentes interprétations, cf. :

*Le mariage et la province vieillissent (un+l') homme.*

*Luc cherche (une+la) maison qui a des volets verts.*

*Z uporem twierdzi, że jest księżniczką.*

→ *Elle s'obstine à dire qu'elle est (la) princesse.*

Il ne fait cependant aucun doute que sans une solution, au moins partielle, du problème du traitement de l'article, on n'arrivera pas à faire un logiciel commercialisable.

## Références

- Blanco X., 1997: «Un dictionnaire électronique des déterminants nominaux en espagnol». BULAG, numéro hors série: *Colloque International Fractal 97*.
- Blanco X., 1998: «Figement et détermination en espagnol». BULAG, 23.
- Blanco X., Buvet P.-A., 1999: «À propos de la traduction automatique des déterminants de l'espagnol et du français». *Métà*, 44, 4, 526–545.

<sup>9</sup> S. Karolak et M. Nowakowska (1999) formulent pour les noms d'objets 42 règles dont 12 s'adressant aux noms propres. Les abstraits, eux, en demanderaient probablement autant.

- Buvet P.-A., 1994: «Détermination : les noms». *Linguisticae Investigationes*, 18, 1.
- Buvet P.-A., 1998: «Détermination et classes d'objets». *Langage*, 131.
- Buvet P.-A., Blanc o X., 1998: «Le projet DétTAL : Détermination et Traitement Automatique des Langues». LLI, Paris 13/UMR 0195 CNRS (INaLF).
- Gniadek S., 1979: *Grammaire contrastive franco-polonaise*. Warszawa: PWN.
- Hutchins W.J., 2003: *Compendium of Translation Software. Commercial Machine Translation Systems and Computer-Aided Support Tools*. [www.eamt.org](http://www.eamt.org).
- Karolak S., Nowakowska M., 1999: *Jak stosować rodzajnik francuski*. Cz. 1. Warszawa: Wydawnictwo Naukowe PWN.
- Wilczyńska W., 1989: *Aprivoiser les articles et les temps*. Poznań: Wyd. UAM.

### Logiciels et pages WEB

- Tłumacz i słownik języka angielskiego*. T. 2. Kompas.
- Tłumacz i słownik języka francuskiego*. T. 1. Kompas.
- [www.ceti.pl/poleng/zasoby/tłumaczenie/english.html](http://www.ceti.pl/poleng/zasoby/tłumaczenie/english.html)
- [www.systransoft.com](http://www.systransoft.com)
- [www.translate.ru/text.asp#tr\\_form](http://www.translate.ru/text.asp#tr_form)