

KRZYSZTOF BOGACKI

Université de Varsovie

kbogacki@gmail.com

POLYSÉMIE, CONTRAINTES SYNTAXIQUES ET BAISSÉ
DE L'AMBIGUÏTÉ DANS LES LANGUES CONTRÔLÉES

Abstract. Krzysztof Bogacki, *Polysémie, contraintes syntaxiques et baisse de l'ambiguïté dans les langues contrôlées* [Polysemy, syntactic constraints and reduction of ambiguity in controlled languages], *Studia Romanica Posnaniensia*, Adam Mickiewicz University Press, Poznań, vol. XLII/1: 2015, pp. 3-14. ISBN 978-83-232-2854-7. ISSN 0137-2475. eISSN 2084-4158. DOI: 10.14746/strop.2015.421.001

The article considers such peculiarities of the controlled languages as the choice of lexical units and constraints on their syntax. We review various difficulties that arise during the selection of forms we would like to put in the lexicon of a controlled language. We compare the index of multiple meanings of the words used to form the vocabulary of a controlled language before and after the lexical control and calculate the change in the ambiguity rate of sentences due to the elimination of polysemy. The words allowed in the lexicon of a controlled language usually have a high polysemy index in the standard language. This is evident especially as far as grammatical and non-technical words are concerned. In contrast, technical terms are much more often monosemic. Thus the observation of the "one token = one meaning" principle in the lexicon for the controlled language has a great effect on the decline of ambiguity.

Keywords: ambiguity, controlled languages, lexical constraints, polysemy, restrictions on syntax.

Dans cet article nous placerons au centre de nos préoccupations une notion connue depuis plusieurs décennies mais devenue à la mode à la suite d'attentats terroristes spectaculaires, de catastrophes naturelles, d'accidents de toute sorte etc., portés à la connaissance du grand public par les mass medias. Il est hors de doute que les mesures pour minimiser ou prévenir les pertes sont prises en premier lieu par les autorités publiques : police, pompiers, médecins, etc. D'un autre côté cependant, la linguistique peut apporter une contribution importante dans ce domaine en proposant des améliorations au niveau de la langue utilisée pour la transmission des messages qui doivent être clairs, non-ambigus et facilement assimilables même dans une situation de stress et d'extrême urgence¹. Nous nous occuperons de la syntaxe des langues contrôlées (= LC) ou, plus exactement, des restrictions découlant de leur caractère contrôlé. Nous allons nous poser plusieurs questions qui concernent la

¹ Les langues contrôlées ont aussi une autre source. Elles sont nées d'un besoin de traduction rapide et fiable de la documentation technique et se sont développées dans les milieux industriels.

sélection des mots admis dans le lexique des LC et le degré de réduction de l'ambiguïté des messages résultant du passage de la langue standard à sa variante contrôlée.

Les remarques qui vont suivre sont formulées à partir d'observations faites sur un corpus de 322 textes : 119 en polonais, 99 en français et 104 en espagnol². Elaborés par les services compétents et diffusés par les soins des mairies et des préfetures ou des organes de l'administration locale, ils étaient destinés, dans la plupart des cas, au **grand public**. À quelques exceptions près, ils relevaient de deux domaines apparentés: celui de la sécurité civile et celui du secourisme. D'un côté, il y avait donc des consignes de sécurité applicables dans le contexte de la lutte contre le terrorisme, l'incendie et les catastrophes naturelles. Ils concernaient les modalités d'alerte des sapeurs-pompiers, précisaient les dispositions à prendre pour assurer la sécurité du public et du personnel en cas de désastre, les mesures à observer pour faciliter l'évacuation des personnes en situation de handicap ou leur évacuation différée si celle-ci est rendue nécessaire, la mise en œuvre des moyens de secours de l'établissement, l'accueil et le guidage des sapeurs-pompiers, les précautions à prendre en cas de pollution, etc. D'un autre côté, dans nos documents il était question des soins d'urgence, et de façon générale, du service d'aide médicale urgente. Quelques-uns des tracts concernaient la prévention de certaines maladies (grippe porcine, maladies sexuellement transmissibles, recommandations de prévention des intoxications alimentaires), vaccination, précautions à prendre avant et après certains examens médicaux, etc.

Nous disposions aussi d'instructions destinées aux **professionnels** de la lutte anti-feu : très précises, utilisant un vocabulaire rempli de termes techniques, elles rappelaient de façon concise les procédures à observer lors des interventions sur place, et étaient souvent accompagnées de photos ou diagrammes. Nous avons recueilli des consignes formulées par les services hospitaliers et les centres SAMU. Dans la plupart des cas, elles étaient destinées au personnel d'accueil travaillant en jonction avec les médecins et les infirmiers qualifiés ainsi qu'avec les services des ambulances ou véhicules des sapeurs-pompiers.

La plupart des textes ont été trouvés sur différents sites Internet et avaient donc une forme électronique. Ceux, au contraire, qui avaient une forme imprimée (tracts, brochures et pancartes destinées à l'affichage dans des établissements publics) ont été convertis en documents électroniques.

La longueur des textes était variable : de 303 à 992 mots. Tous ces documents ont été « contrôlés » selon les normes LiSe³, et les documents ainsi modifiés ont servi à la création d'une liste de lexèmes qui s'est avérée plus courte que celle

² Ils ont été recueillis sur Internet dans les années 2007-2008 pour les besoins d'un projet international (MESSAGE = Alert Messages and Protocols : JLS/2007/CIPS/022) financé par la Commission Européenne.

³ <http://projet-lise.univ-fcomte.fr/>. Ce site contient nombre d'informations de toute sorte relatives aux recherches récentes sur les langues contrôlées.

obtenue à partir de textes originaux. Dans certains cas, la réduction dépassait 50% (p. ex. 992 mots avant le contrôle vs 449 après le contrôle).

1. QU'EST-CE QU'UNE LANGUE CONTRÔLÉE ?

On connaît les objectifs principaux qui président à la création des langues contrôlées. Ce sont la compréhensibilité des textes, leur traductibilité et leur utilisabilité par les sujets parlant différentes langues maternelles. Cela se ramène à la prévention des erreurs d'interprétation dues aux imprécisions de la langue naturelle, au stress et aux interférences entre les langues. D'un autre côté, les LC sont souvent accompagnées de consignes rédactionnelles ayant en réalité la forme de guides de rédaction qui spécifient, entre autres, la structure des textes (titre, sous-titre éventuel, etc.), la mise en page (alignement des parties constitutives du document), le style et la taille de police utilisée, etc.⁴.

2. L'AMBIGUÏTÉ ET LE CONTRÔLE LEXICAL DANS LES LANGUES CONTROLÉES

Le point crucial dans la structure de LC est sans doute le refus de toute ambiguïté. C'est elle qui, en fin de compte, constitue une pierre d'achoppement à la compréhension des messages et à la traduction automatique. Elle est de différents types. Fuchs (1996) mentionne l'ambiguïté référentielle, pragmatique, morphologique, lexicale et syntaxique. Certaines sont bien décrites et assez faciles à maîtriser. Tel est avant tout le cas de l'ambiguïté morphologique pour laquelle la solution passe par une description méticuleuse de la morphologie qui est à la base d'algorithmes susceptibles de la lever. L'élimination d'autres types d'ambiguïté exige des restrictions sur l'emploi des pronoms, l'utilisation correcte des mécanismes anaphoriques et autres. Nous allons nous attacher à l'étude des ambiguïtés lexicales. Dans la plupart des cas, elles peuvent être évitées lors de l'opération du contrôle du lexique d'une LC avec pour objectif de les réduire au maximum⁵. Leur source est à chercher dans la multiplicité des sens des lexèmes. Dans de tels cas, on pourchasse donc l'homonymie, la polysémie et la polycatégorialité proscrivant p. ex. l'utilisation de mots qui sont substantifs et adjectifs (*ouvrier, protecteur*), participes et adjectifs (*brûlé, changeant*), prépositions et substantifs (*pour, contre*), prépositions et adverbes (*avant, après*), etc. On interdit aussi la synonymie. Face à l'extraordinaire polysémie du

⁴ Cf. J. Renahy (2010).

⁵ L'élimination totale de l'ambiguïté inhérente aux langues naturelles standard semble impossible (elle est d'ailleurs utilisée intentionnellement par les sujets parlants, par exemple dans certains jeux de mots).

langage naturel, différentes recommandations sont formulées. La plus importante sans doute est souvent réduite à la formule : « une entrée lexicale = un sens » et réciproquement : « un sens = une entrée lexicale ». En d'autres termes, elle exige qu'on assigne un seul sens et une seule catégorie grammaticale à un mot, tout en veillant à faire le choix en fonction du domaine d'application (services de sécurité civile, services médicaux d'urgence, transport routier, etc.). Ainsi des trois sens du substantif *incendie*⁶ :

- ‘grand feu qui se propage en causant des dégâts’ ;
- ‘lumière rougeoyante éclairant une grande étendue’ (p. ex. *L'incendie du soleil couchant*) ;
- ‘bouleversement, guerre’

on retiendra le premier pour une LC destinée au domaine de la lutte anti-feu. Le mot *grue* pris au sens de ‘grand oiseau échassier (*gruidés*) qui migre en troupe’ pourrait faire partie du domaine ‘foresterie’ tandis que le même mot sera admis dans celui du bâtiment et des travaux publics avec le sens ‘machine de levage et de manutention’. D'après ce qui vient d'être dit, on voit que sur le plan lexical les LC – « stérilisées », dépourvues de polysémie, sans synonymes – sont autant de versions limitées et donc « appauvries » de la langue standard.

La procédure de l'établissement du lexique contrôlé commence par la création d'une liste des mots trouvés dans les textes relevant du domaine pour lequel on cherche à définir la LC. Ainsi dans les textes polonais se rapportant à la lutte anti-feu, nous trouvons 18 conjonctions qui sont mentionnées dans *Słownik języka polskiego PWN* (2009)⁷ avec 62 sens: *a* ‘et’ (9), *aby* ‘pour que’ (6), *albo* ‘ou’ (2), *ażby* (2) ‘afin que’, *by* (6) ‘afin que’, *dopóki* (2) ‘tant que’, *gdy* (3) ‘quand’, *i* (6) ‘et’, *jak również* (5) ‘ainsi que’, *jeśli ... to* (3) ‘si’, *jeżeli... to* (5) ‘si’, *lecz* (1) ‘mais’, *lub* (1) ‘ou’, *oraz* (1) ‘ainsi que’, *ponieważ* (1) ‘parce que’, *też* (3) ‘aussi’, *zanim* (1) ‘avant que’, *że* (5) ‘que’. Cette liste peut être réduite à 13 items grâce à l'élimination de *ażby* et de *by* confondus avec *aby*, à celle de *lecz* au profit de *ale* (introduit dans la liste), à celle de *jeżeli* (à sa place on retiendra *jeśli*), à celle de *lub* remplacé par *albo*, et de *oraz*, *jak również* (remplacés par *i*). À chacun de lexèmes de la liste modifiée on assigne un seul sens choisi parmi tous ceux qu'il est susceptible d'exprimer dans la langue standard et en tenant compte de ses propriétés syntaxiques.

Parfois le contrôle monolingue pose problème. Il n'est pas rare de revenir sur les choix faits au fur et à mesure que de nouvelles entités lexicales affluent mettant le concepteur de la LC devant l'alternative : faut-il abandonner les solutions adoptées antérieurement ? ne pas tenir compte du matériel lexical nouvellement apparu ? procéder à des réajustements en matière de sens assigné aux lexèmes admis dans le lexique ? Un seul exemple suffit. Soit le sens ‘avoir quelque chose avec soi, à sa

⁶ Nous empruntons nos exemples et les définitions des mots à l'édition électronique du *Nouveau Petit Robert* 2001.

⁷ Cf. aussi Wajszczuk (1997).

disposition'. Il peut être exprimé, entre autres, par *garder* (*les clés dans un endroit sûr*) ou par *conserver* (*les clés dans un endroit sûr*). Si dans un premier temps c'est *conserver* qui est retenu en remplacement du verbe *garder*, du coup seront exclus tous les autres sens de celui-ci, entre autres 'prendre soin d'une personne' (*garder les enfants*), 'empêcher (une personne) de s'en aller' (*garder un prisonnier*) et 'rester dans (un lieu) pour surveiller' (*garder une maison*). La solution consisterait à allonger la liste des mots admis en y ajoutant le verbe *avoir* en conséquence de quoi on aurait à sa disposition *garder* et *conserver* avec tous leurs sens. On pourrait les éliminer ou les retenir en leur assignant un des sens que chacun d'eux est susceptible d'exprimer.

Les lexèmes français retenus pour former le lexique contrôlé applicable au domaine de la sécurité civile formé à partir de notre corpus sont tous polysémiques dans la langue standard. Par exemple, les 22 verbes qui en font partie, sont classés dans la base de données « Les verbes français »⁸ (= LVF) parmi ceux qui se construisent selon le nombre le plus élevé de schèmes (auxquels correspondent différents sens) : *passer* (61 sens), *prendre* (43), *faire* (42), *monter* (40), *porter* (37), *sortir* (37), *tirer* (36), *tenir* (35), *tomber* (33), *donner* (30), *perdre* (30), *couper* (29), *mettre* (27), *présenter* (27), *tourner* (27), *lever* (26), *revenir* (26), *ouvrir* (25), *couvrir* (24), *jeter* (24), *laisser* (24), *poser* (24). Ils figurent tous dans LVF parmi les 30 verbes les plus diversifiés syntaxiquement. Seulement 8 n'ont pas trouvé leur place dans la liste : *jouer* (33), *coller* (31), *piquer* (31), *relever* (24), *remettre* (24), *rouler* (24) et *sauter* (24). Quant à *lancer* (24), il a été éliminé à cause de sa synonymie, pour un de ses sens, avec *jeter* (24) qui, lui, a été retenu.

3. LE CONTRÔLE PLURILINGUE

Si la recommandation « un sens = une forme » est difficile à mettre en oeuvre pour une seule langue, elle crée des difficultés pratiquement insurmontables quand on veut tenir compte de deux langues à la fois pour assurer un taux d'ambiguïté zéro en langue cible. Introduisant une dimension contrastive et permettant de mettre l'accent sur les convergences et les divergences entre les deux langues, dans l'idéal, le contrôle simultané devrait simplifier à la fois la langue de départ et la langue cible. La tâche pour un concepteur de LC, dans ce cas-là, peut devenir un véritable casse-tête. Or, cette consigne est justifiée par les exigences de la traduction automatique. En effet, nous avons pu montrer (cf. Bogacki 2009) que les résultats deviennent nettement meilleurs lorsque, dans un dictionnaire de transfert, le nombre d'entrées lexicales ayant plus d'un équivalent dans la langue cible est réduit. C'est donc pour cette raison qu'on postule de procéder au contrôle orienté vers la langue

⁸ <http://rali.iro.umontreal.ca/LVF+1>.

cible. On sait cependant que dans certains cas la bi-univocité parfaite ne peut être obtenue. C'est par exemple le cas du français qui maintient la distinction entre *rivière* et *fleuve* alors que l'anglais se contente d'un lexème unique (*river*). Beddar (2013 : 192) cite le cas du couple français–arabe. Le français masque la distinction entre *papiers* 'ensemble des pièces d'identité' et 'morceaux de matière fabriquée avec des fibres végétales réduites en pâte, étendue et séchée pour former une feuille mince' contrairement à l'arabe où il existe deux mots assurant une distinction au niveau lexical. Le substantif الوثائق désigne les papiers administratifs et se prête, par conséquent, à la traduction de la phrase *présenter ses papiers aux autorités* (p. ex. lors du contrôle de police) alors que الأوراق, ayant le sens général, convient comme traduction dans d'autres contextes (p. ex. *ne pas jeter les papiers par terre*).

Nous avons comparé deux à deux les lexiques de trois LC : le polonais, l'espagnol et le français. Comme on pouvait s'y attendre, nombreux ont été les cas de correspondance non-bi-univoque entre les entrées des deux lexiques. Ainsi face au polonais *ponad*, il y avait deux expressions prépositionnelles espagnoles *por encima de* et *más de* qui semblaient tout aussi indispensables. *Ruch* s'est avéré nécessaire aussi bien dans le sens de 'déplacement (d'un corps) par rapport à un point fixe de l'espace et à un moment déterminé' (cf. l'espagnol *movimiento*) qu'au sens de 'circulation, transport de personnes ou de marchandises' (rendu en espagnol par *tráfico*).

La préposition polonaise *przed* était employée dans nos textes aussi bien dans le sens spatial que temporel ayant pour équivalents espagnols deux expressions prépositionnelles : *delante de* et *antes de* et *avant/devant* en français.

Dans la terminologie utilisée par les sapeurs-pompiers on trouve de nombreux exemples où un terme polonais est rendu par deux équivalents distincts choisis en fonction du contexte : *aparat alarmowy/appareil d'alarme*, *aparat przeciwpożarowy/avertisseur d'incendie*, *droga dojazdu/voie d'accès*, *droga leśna/chemin forestier*, *droga transportowa/voie de transport*, *droga uciezki/chemin de fuite*, *droga wyjścia/chemin de sortie*.

4. LE CONTRÔLE SYNTAXIQUE

Ce qui est frappant, c'est que les auteurs qui sont amenés à parler, d'une manière ou d'une autre, du contrôle syntaxique, tout en reconnaissant l'importance de ce facteur pour minimiser l'ambiguïté d'un message formulé dans une LC y consacrent en fait très peu de place. Deux thèses récentes sont explicites sur les contraintes syntaxiques. La dernière en date, celle de Thongglin (2014) les aborde en parlant des ambiguïtés rencontrées dans les spécifications des besoins des logiciels en thaï. Quant à Beddar (2013 : 193), sa réflexion sur les ambiguïtés syntaxiques se limite aux structures de type *poste de contrôle sanitaire*. Dans la perspective de la traduction automatique entre le français et l'arabe, de telles constructions mettent le système

devant un dilemme qui tient à l'identification de l'élément régissant l'adjectif *sanitaire*. Si en français la morphologie de l'adjectif ne varie pas en fonction du changement du genre du nom qu'il qualifie (aussi bien *poste* que *contrôle* sont masculins), en arabe le genre va changer : *poste* est masculin, *contrôle* est féminin.

L'éventail des restrictions syntaxiques est sans doute le plus long dans la thèse de J. Renahy (2010) et dans l'article d'O'Brien (2003). La liste de contraintes proposées par les concepteurs des langues contrôlées s'avère variée⁹ : elles portent avant tout sur l'ordre des termes dans la phrase et l'emploi de diverses catégories de mots à l'exclusion d'autres. Certaines semblent fondées sur le bon sens. Tel est, par exemple, le cas des limites sévères sur la longueur de phrases ou l'obligation de choisir les structures exprimant une condition dans une liste très restrictive (*si/quand/dès que* + [sujet] + [verbe] + [complément] + [complément], *en cas de* + GN etc.). D'autres contraintes syntaxiques sont visiblement motivées par les recherches des psycholinguistes sur l'assimilation par les sujets parlants de certaines structures (elle décroît de façon inversement propositionnelle au nombre de propositions relatives enchâssées). Il y en a aussi dont le rôle dans la désambiguïsation n'est pas évident. Tel est le cas de l'obligation d'écrire les compléments circonstanciels en fin de phrase uniquement, l'interdiction d'employer les adverbes de phrase (dans *Attempto Controlled English*, cf. Fuchs 2005), la préférence accordée à l'infinitif pour l'expression de l'ordre ou enfin l'obligation d'utiliser les noms au singulier et de mettre les verbes au présent (dans *Common Logic Controlled English*, cf. Sowa 2004).

Il est sûr que de telles restrictions, quand elles sont bien ciblées, augmentent la compréhensibilité du texte. Ce qui frappe cependant à l'examen des contraintes syntaxiques proposées dans les LC, c'est le fait qu'elles ne découlent pas des caractéristiques idiosyncrasiques des lexèmes mais touchent soit à la structure globale de la phrase, des groupes nominaux ou verbaux, soit aux catégories grammaticales propres à des classes entières de mots et qui ne sont pas liées à la valeur des lexèmes précis (cf. O'Brien, 2003). Or, il ne fait aucun doute « qu'il y a **adéquation** entre les schèmes syntaxiques d'une 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 entre la destination et l'origine » (J. Dubois & F. Dubois-Charlier, 1997 : 1). Autrement dit, les entrées lexicales et les règles grammaticales – syntaxiques en l'occurrence – sont interdépendantes même s'il serait risqué d'admettre qu'il y a un profil syntaxique par sens et inversement. On peut faire le même constat pour le polonais à la lecture de Polański (1980-1992) et pour le français, en étudiant les tables des verbes élaborées par les chercheurs du LADL (cf. en particulier Gross, 1975) ou celles de la base de données « Les verbes français » (= LVF) de J. Dubois et F. Dubois-Charlier. Les profils syntaxiques des

⁹ Cf. aussi www.muegge.cc/controlled-language.htm

lexèmes y sont décrits en termes de schèmes syntaxiques spécifiant le nombre d'actants connotés, le type de prépositions éventuelles utilisées pour les relier, les traits sémantiques des arguments, les propriétés transformationnelles, etc.¹⁰. En principe, ce sont ces éléments-là qui permettent de différencier les sens de unités lexicales bien que dans certains cas, la désambiguïsation demande une description plus détaillée. Ainsi le verbe *éteindre*, dix fois ambigu selon le *Petit Robert*, se construit selon le schème $N_0 + V + N_1$ admettant en position de N_1 des noms d'humains ce qui dans une phrase comme

On a éteint Paul

autorise deux lectures : « ordinaire » ('arrêter la combustion') et « figurée » ('faire perdre vigueur'). Cette ambiguïté disparaît dans

On a éteint Paul avec cette remarque

en présence du syntagme modificateur *avec cette remarque* qui permet de lever le doute au profit du sens « psychologique ».

5. RÉDUCTION DE LA POLYSÉMIE ET BAISSÉ DE L'AMBIGUÏTÉ

Le lexique d'une LC contient les mots qu'on pourrait grouper en trois classes : mots grammaticaux (conjonctions, déterminants, prépositions, etc.), ceux du langage standard (verbes, adjectifs, certains substantifs, etc.) avec un seul sens et des termes techniques. Le tableau ci-dessous présente les données concernant les mots extraits des documents « contrôlés » réparties en 3 classes :

	Total	Mots grammaticaux	Lexique standard	Termes techniques
Espagnol	1164	88	899	177
Français	1099	79	895	125
Polonais	1151	65	945	141

Chacun de ces lexèmes a son profil syntaxique et on peut remarquer que, dans leur usage usuel (non-contrôlé), les mots admis dans le lexique d'une LC se caractérisent d'habitude par un indice de polysémie élevé. Cela est visible surtout dans les mots grammaticaux et dans le lexique standard. Les termes techniques, au contraire, sont beaucoup plus souvent monosémiques. L'opération du contrôle lexical qui s'applique aux items des 3 groupes a pour effet de réduire l'indice global de polysémie pour tout le lexique. Cependant il n'est pas possible de prévoir le taux de

¹⁰ La description de deux verbes français : *monter* et *sortir* proposée par Żłobińska-Nowak (2008) offre un degré de granularité plus fin. La spécification des profils syntaxiques des verbes polonais nécessiterait systématiquement des précisions sur le cas des arguments nominaux.

réduction qui est variable selon la LC prise en compte. Ainsi l'indice de polysémie¹¹ pour l'ensemble des 98 mots du lexique de la LC 'incendie' établi à partir des textes français avant le contrôle était de 987 tandis que celui de 'médecine d'urgence' était de 1041 pour 97 mots. Bien qu'impressionnante dans les deux cas, la baisse de l'indice de polysémie est différente: $987/98 = 10,07$ dans le premier cas et $1043/97 = 10,73$ dans l'autre. De toutes façons, si chaque sens est corrélé avec un profil syntaxique unique, la baisse de l'indice de polysémie – qui, dans l'idéal, devrait passer à zéro – diminue le nombre de structures syntaxiques autorisées. Or une des conséquences est la baisse spectaculaire de l'ambiguïté des messages. Nous allons illustrer ce phénomène avec trois exemples.

Le premier concerne les messages réduits à un seul mot tels que : *Feu !*, *Attention !* On passe, dans ce cas-là, de 24 sens (pour *feu*) à 1 et de 3 (dans le cas de *attention*) à 1. Les deux autres sont constitués par des structures phrastiques et par les syntagmes nominaux de type Adj + N.

Considérons les phrases construites sur le verbe *sortir*. Nous le retiendrons au sens de 'monter, quitter un endroit' illustré par la phrase *La fumée sort de la cheminée* construite selon le schème syntaxique $N_0 + V + de + N_1$. Le même schème se retrouve avec le verbe *sortir* exprimant d'autres sens :

*sortir*₁ – 'aller hors d'un lieu (en parlant des animés)' – *Sortir de l'immeuble, sortir du lit* ;

*sortir*₂ – 'dépasser les limites d'un lieu abstrait' – *L'orateur sort du sujet, de sa compétence, du temps imparti* ;

*sortir*₃ – 'se répandre hors de' – *La rivière sort de son lit. Le sang sort de la plaie* ;

*sortir*₄ – 'ne plus être dans' – *On sort de l'hiver, d'une grève* ;

*sortir*₅ – 'être frais émoulu' – *Cet ingénieur sort d'une grande école* ;

où la différence entre les sens est assurée au niveau des arguments connotés et peut être décrite par un jeu de contraintes sémantiques. En effet, pour les sens ci-dessus, en position N_0 on trouve respectivement [+liquide] (par opposition à [+gaz], [+animé] ou [+humain]) d'un côté et de l'autre : [+concret] ou [+abstrait-temps], [+temps], [+école] en position N_1 .

Cela étant, l'ambiguïté théorique de la phrase examinée est de 20 sens ($4*5$) étant donné que le schème phrastique a deux positions N_0 et N_1 et que *fumée* a 4 sens alors que *cheminée* en a 5. Cf. :

*fumée*₁ – 'mélange plus ou moins dense et de couleur variable de produits gazeux et de très fines particules solides, qui se dégage des corps en combustion ou portés à haute température' ;

¹¹ Nous avons tenu compte des données fournies par l'édition électronique du *Nouveau Petit Robert*, v. 2.1.

- fumée*₂ – ‘vapeur qu’exhale un liquide ou un corps humide dont la température est plus élevée que l’air ambiant’ ;
*fumée*₃ – ‘vapeurs qui sont supposées monter au cerveau sous l’effet de l’alcool, brouillant ainsi les idées’ ;
*fumée*₄ – ‘excréments des cerfs et autres bêtes sauvages’ ;
*cheminée*₁ – ‘dispositif formé d’un foyer et d’un tuyau qui sert à évacuer la fumée’ ;
*cheminée*₂ – ‘partie inférieure de la cheminée qui sert d’encadrement à l’âtre’,
*cheminée*₃ – ‘partie supérieure du conduit qui évacue la fumée et que l’on voit sur le toit’ ;
*cheminée*₄ – ‘trou, conduit vertical’ ;
*cheminée*₅ – ‘ouvrage (ou formation naturelle) servant de régulateur de pression dans un système hydraulique’.

En réalité cependant toutes les combinaisons ne sont pas possibles : semblent exclues celles où *fumée* aurait le sens 3. Seuls les sens ‘concrets’ de *fumée* sont acceptables: le passage à l’abstrait se solde par l’agrammaticalité de la phrase. L’indice d’ambiguïté réelle passe dans ce cas-là de 20 à 15 car on a 3 possibilités en position N₀ et 5 autres en position N₁ (3*5 = 15).

Avec le même verbe employé au sens de ‘dépasser les limites d’un lieu abstrait’, l’ambiguïté varie en fonction du nombre des sens de ses arguments. Considérons

L’orateur sort du sujet (‘l’orateur dépasse les limites du sujet’)

Le professeur sort du sujet

Le président sort du sujet

avec le substantif *sujet* en position N₁. Si des 13 sens qu’il peut avoir on en exclut d’avance 7 qui désignent une personne ou véhiculent un sens grammatical (‘terme considéré comme le point de départ de l’énoncé’), seuls les 6 sens ci-dessous sont compatibles avec la fonction de lieu abstrait :

- sujet*₁ – ‘ce qui est soumis à l’esprit, à la pensée; ce sur quoi s’exerce la réflexion ; ce dont il s’agit, dans la conversation, dans un écrit’ ;
*sujet*₂ – ‘ce qui, dans une œuvre littéraire, constitue le contenu de pensée sur lequel s’est exercé le talent créateur de l’auteur’ ;
*sujet*₃ – ‘ce sur quoi s’applique la réflexion, dans un travail scientifique, une œuvre didactique’ ;
*sujet*₄ – ‘thème ou motif principal, spécialement dans la musique contrapuntique’ ;
*sujet*₅ – ‘ce qui est représenté ou évoqué dans une œuvre graphique, plastique’,
‘représentation d’un motif anecdotique, littéraire, d’une action, ce motif’ ;
*sujet*₆ – ‘ce qui fournit matière, occasion à (un sentiment, une action)’.

La phrase avec le substantif *orateur*, doublement ambigu, en position de sujet présentera 12 lectures. Le mot *président*, ayant 3 sens, fera monter le nombre de lectures à 18 tandis que *professeur*, qui est monosémique, fera descendre le nombre de lectures possibles à 6.

Les résultats sont différents si on fait varier le lieu abstrait tout en gardant les mêmes mots en position de sujet. Ainsi avec $N_1 = \textit{compétence}$ (4 sens dont 2 compatibles) et $N_0 = \textit{orateur}$ (2 sens) on aura 4 lectures de la phrase tandis que $N_1 = \textit{temps}$ (14 sens dont 4 compatibles) combiné avec $N_0 = \textit{orateur}$ donne 8 lectures (4×2).

Le même mécanisme se retrouve dans les syntagmes. Avant le contrôle lexical, l'indice d'ambiguïté théorique correspond au produit d'indices de polysémie de ses composants. Par exemple dans le syntagme composé d'un adjectif et d'un substantif : *de lourdes présomptions*, l'adjectif est 11 fois ambigu alors que le substantif ne peut convoquer que 2 sens ('opinion fondée seulement sur des signes de vraisemblance' et 'opinion trop avantageuse que l'on a de soi-même'). Étant donné que l'adjectif n'est retenu qu'avec les sens compatibles avec la classe sémantique du substantif auquel il se rapporte, l'ambiguïté réelle va diminuer. Seront éliminés les sens « physiques » de l'adjectif *lourd* : 'difficile, pénible à porter, à déplacer, en raison de son poids', 'dont le poids est élevé ou supérieur à la moyenne' et les sens « moraux » maintenus, p. ex. 'au figuré, grand (en parlant de ce qu'on supporte)'.

6. CONCLUSION

Le contrôle lexical a deux visages : sémantique et syntaxique. Le premier, sémantique, difficile à observer surtout dans la perspective bilingue, obéit au principe « une entrée lexicale = un sens ». Or, ce choix a des répercussions sur le plan de la syntaxe. En effet, avec le sens retenu est sélectionné en même temps un schème syntaxique donné. L'opération permet de réduire la polysémie et par là-même de faire baisser l'ambiguïté.

Celle-ci n'est ni la simple somme ni le simple produit des indices de polysémie de ses composantes lexicales et encore moins celle des lexèmes formant le lexique entier de la LC. Elle est par contre fonction du nombre de sens compatibles avec les caractéristiques des positions ouvertes par le terme régissant : substantif-tête dans le syntagme nominal et verbe au niveau de la phrase. S'il est certain que l'ambiguïté peut être réduite grâce au contrôle lexical, l'ampleur de la baisse d'ambiguïté obtenue par ce moyen-là ne peut pas être prévue d'avance. Elle dépend du degré de polysémie calculé avant le contrôle des lexèmes formant le lexique d'une LC.

RÉFÉRENCES

- BEDDAR, Mohand (2013) : « Vers un prototype de traduction automatique contrôlée français/arabe appliquée aux domaines de sécurité ». [Thèse de doctorat]. Besançon.
- BOGACKI, Krzysztof (2009) : « Controlled languages and Machine Translation ». Dans : *ISMTCL Proceedings*, International Review Bulag, PUFC, 49-55.

- DUBOIS, Jean ; DUBOIS-CHARLIER, Françoise (1997) : *Les verbes français*. Paris : Larousse-Bordas.
- EUN, Soon You (2007) : « Le traitement des unités lexicales polysémiques (l'adjectif et le verbe). Vers un système de traduction automatique ». [Thèse de doctorat]. Besançon.
- FUCHS, Catherine (1996) : *Les ambiguïtés du français*. Paris : Éditions Ophrys, Coll. L'Essentiel français.
- FUCHS, Norbert E. (2005) : *Attempto Controlled English*, ms. www.ifi.unizh.ch/attempto/talks/files/Talk.Stanford.05.pdf
- GROSS, Maurice (1975) : *Méthodes en syntaxe. Le régime des constructions complétives*. Paris : Hermann.
- KIATTIBUTRA-ANANTALAPOCHAI, Raksi (2011) : « Analyse lexicale morphologique et syntaxique du thaï en vue de la traduction automatique appliquée au domaine de l'administration publique ». [Thèse de doctorat]. Besançon.
- O'BRIEN, Sharon (2003) : *Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets*, EAMT-CLAW 2003, www.mt-archive.info/CLT-2003-O'Brien.pdf
- POLAŃSKI, Kazimierz (1980-1992): *Słownik syntaktyczno-generatywny czasowników polskich*. Wrocław : Zakład Narodowy im. Ossolińskich.
- ROBERT, Paul (2001) : *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Paris : Le Robert.
- RENAHY, Julie (2010) : « Conception d'une langue contrôlée généralisante. (Application aux domaines de la santé, sécurité civile). Établissement des normes de contrôle pour la création de protocoles de sécurité pour assurer compréhensibilité et traductibilité ». [Thèse de doctorat]. Besançon.
- Słownik języka polskiego PWN* (2009). Warszawa : Wydawnictwo Naukowe PWN.
- SOWA, John F. (2004) : *Common Logic Controlled English*, ms. www.jfsowa.com/clce/specs.htm
- THONGGLIN, Kiatibutra (2014) : « Controlled Language for Thai Software Requirements Specification » [Langue contrôlée pour la spécification des besoins du logiciel en thaï]. [Thèse de doctorat]. Besançon.
- WAJSZCZUK, Jadwiga (1997) : *System znaczeń w obszarze spójników polskich. Wprowadzenie do opisu*. Warszawa : Katedra Lingwistyki Formalnej UW.
- WU, Xiaohong (2006) : « Conception d'une langue contrôlée pour un système de traduction automatique de protocoles médicaux : applications aux domaines de l'échinococcose et au clonage moléculaire ». [Thèse de doctorat]. Besançon.
- ŻLOBIŃSKA-NOWAK, Aleksandra (2008) : *Désambiguïsation des expressions lexicales des opérateurs de l'espace dans le cadre d'une approche orientée objets : les verbes de mouvement "monter" et "sortir" et leurs équivalents polonais*. Katowice : Wydawnictwo Uniwersytetu Śląskiego.

RÉSEAUGRAPHIE

<http://projet-lise.univ-fcomte.fr/>
<http://rali.iro.umontreal.ca/LVF+1>
www.muegge.cc/controlled-language.htm