

Un dictionnaire de reformulation pour les apprenants du français langue seconde

Jasmina Milićević and Marie-Josée Hamel

Special Issue, 2007

Les apports de la sociolinguistique et de la linguistique à l'enseignement des langues en contexte plurilingue et pluridialectal
Sociolinguistics, Linguistics and the Teaching of Languages in Multilingual and Multidialectal Contexts

URI: <https://id.erudit.org/iderudit/017713ar>

DOI: <https://doi.org/10.7202/017713ar>

[See table of contents](#)

Publisher(s)

Revue de l'Université de Moncton

ISSN

0316-6368 (print)

1712-2139 (digital)

[Explore this journal](#)

Cite this article

Milićević, J. & Hamel, M.-J. (2007). Un dictionnaire de reformulation pour les apprenants du français langue seconde. *Revue de l'Université de Moncton*, 145–167. <https://doi.org/10.7202/017713ar>

Article abstract

This paper describes a research project in lexicology applied to the teaching of French as a second language, which aims at developing an electronic dictionary of a particular type called a *reformulation dictionary*. This dictionary should enable intermediate-to-advanced level students to overcome difficulties in using the French lexicon in a flexible and idiomatic way when performing language production tasks. The following questions are addressed : the theoretical bases and the methodology for the development of a reformulation dictionary, the concept of lexical error and the description of lexical errors with the aid of formal tools that we suggest, and the overall design and the implementation of our reformulation dictionary. The theoretical framework for this research is Meaning-Text linguistic theory.

UN DICTIONNAIRE DE REFORMULATION POUR LES APPRENANTS DU FRANÇAIS LANGUE SECONDE

Jasmina Milićević

et

Marie-Josée Hamel

Dalhousie University

Résumé

Nous présentons une recherche en lexicologie appliquée à l'enseignement du français langue seconde qui vise le développement d'un dictionnaire électronique de type particulier, appelé *dictionnaire de reformulation*. Ce dictionnaire devrait permettre aux étudiants de niveau intermédiaire à ceux de niveau avancé de surmonter des difficultés que présente une utilisation souple et idiomatique du lexique du français dans la production langagière. Les questions abordées sont les suivantes : les fondements théoriques et la méthodologie pour l'élaboration d'un tel dictionnaire, le concept d'erreur lexicale et la description des erreurs lexicales à l'aide d'outils formels que nous proposons, l'architecture du dictionnaire de reformulation et l'implémentation de ce dernier. Le cadre théorique dans lequel se situe notre recherche est celui de la théorie linguistique Sens-Texte.

Abstract

This paper describes a research project in lexicology applied to the teaching of French as a second language, which aims at developing an electronic dictionary of a particular type called a *reformulation dictionary*. This dictionary should enable intermediate-to-advanced level students to overcome difficulties in using the French lexicon in a flexible and idiomatic way when performing language production tasks. The following questions are addressed : the theoretical bases and the methodology for the development

of a reformulation dictionary, the concept of lexical error and the description of lexical errors with the aid of formal tools that we suggest, and the overall design and the implementation of our reformulation dictionary. The theoretical framework for this research is Meaning-Text linguistic theory.

1. Fondements théoriques et méthodologie pour l'élaboration d'un dictionnaire de reformulation

Cet article présente une recherche en lexicologie appliquée à l'enseignement du français langue seconde¹. Nous visons le développement d'un dictionnaire de type particulier, appelé *dictionnaire de reformulation*, qui permettra aux étudiants de niveau intermédiaire à avancé de surmonter des difficultés que présente une utilisation souple et idiomatique du lexique du français dans la production langagière, utilisation qui s'approche de celle des locuteurs natifs.

La thèse centrale sur laquelle s'appuie notre travail est la suivante : bien parler une langue, qu'elle soit maternelle ou étrangère, c'est savoir paraphraser [= produire des paraphrases, ou phrases sémantiquement (quasi-)équivalentes] dans cette langue; cf. Žolkovskij et Mel'čuk (1967), Fuchs (1981), Mel'čuk (1992) et Milićević (2003). En d'autres termes, être à l'aise dans une langue, c'est savoir « dire la même chose autrement », ou encore, savoir *reformuler* son discours. Ainsi, un locuteur ayant une bonne maîtrise du français est capable de produire pour la phrase (1a) au moins quelques-unes de ces paraphrases (1b)-(1f).

- (1) a. *Il faut éviter les abréviations autant que possible.*
- b. *Les abréviations doivent être évitées dans la mesure du possible.*
- c. *Partout où cela est possible, évitez les abréviations.*
- d. *N'utiliser les abréviations que si cela est absolument nécessaire.*
- e. *Il vaut mieux ne pas utiliser les abréviations.*
- f. *L'utilisation des abréviations est fortement déconseillée.*

D'une part, le locuteur a besoin de paraphrases afin de produire au moins une expression correcte pour le sens qu'il veut communiquer. Si

l'on dispose de plusieurs possibilités pour exprimer une idée, on est mieux à même de surmonter les obstacles — cooccurrence lexicale restreinte, lacunes lexicales, paradigmes défectifs — auxquels on se heurte sans cesse dans la production de la parole. La multiplicité des options permet de ne pas rester bloqué, à cause d'un « cul-de-sac », ou à cause d'un problème d'accès en mémoire, avec une variante qui serait la seule utilisable dans une situation donnée.

D'autre part, la paraphrase est nécessaire afin de trouver la meilleure expression pour un contenu sémantique donné dans une situation de communication donnée : il est important de savoir comment « présenter » la phrase pour rendre son idée plus claire, changer le style, éviter des répétitions, etc. Les textes (2b-c) ci-dessous illustrent les reformulations successives déclenchées par la répétition gênante du verbe *ÉVITER* qu'on voit en (2a). (Une autre possibilité de reformuler le texte (2a) serait de substituer à la première phrase une de ses paraphrases (1d-f) en gardant la seconde phrase telle quelle.)

- (2) a. *Il faut **éviter** les abréviations autant que possible. Celles qu'on ne peut pas **éviter** doivent être introduites, même si elles sont triviales pour nous.*
- b. *Il faut **éviter** les abréviations autant que possible. Celles qu'on **doit garder** doivent être introduites, même si elles sont triviales pour nous.*
- c. *Il faut **éviter** les abréviations autant que possible. Celles qu'on **a besoin de garder doivent** être introduites, même si elles sont triviales pour nous.*

La maîtrise de la paraphrase passe par celle du lexique : pour établir des liens paraphrastiques entre expressions d'une langue *L*, il faut connaître les propriétés – sémantiques, syntaxiques, de cooccurrence – des lexies de *L*². En particulier, il faut connaître les *relations lexicales* qu'une lexie donnée entretient avec d'autres dans le lexique. Comme on le sait, ces relations sont de deux types principaux : *relations de collocation* (*L*₁ est un intensificateur, un verbe support, un verbe de réalisation, etc., typiquement utilisé avec *L*₂) et *relations de dérivation sémantique* (*L*₁ est

un synonyme, un antonyme, un conversif, une nominalisation, une verbalisation, etc., de L₂).

Par exemple, l'expression *absolument nécessaire* est une collocation dont l'adjectif NÉCESSAIRE est la base et où le collocatif, l'adverbe ABSOLUMENT, exprime le sens 'intense'/'très' auprès de la base et en fonction de cette dernière. Dans la collocation *avoir besoin*, le collocatif AVOIR est un verbe support, c'est-à-dire un verbe sémantiquement vide, qui sert à « verbaliser » la base BESOIN. Le nom UTILISATION 'fait d'utiliser' est un dérivé nominal exact du verbe UTILISER. Et ainsi de suite.

Le lien paraphrastique entre *ne pas pouvoir éviter P* et *devoir garder P* exploité dans la reformulation indiquée en (2b) se base sur la relation de quasi-antonymie entre POUVOIR 'être.possible' et DEVOIR 'être.nécessaire', d'une part, et ÉVITER et GARDER, d'autre part; il s'agit, *grosso modo*, d'une paraphrase par double antonymie : 'ne pas pouvoir cesser [= ne pas continuer] d'utiliser' ≈ 'devoir continuer à utiliser'. Le lien paraphrastique exploité dans la reformulation en (2c), quant à lui, se base sur le lien de quasi-synonymie entre DEVOIR (*faire P*) et avoir BESOIN (*de faire P*).

Les relations lexicales et les liens paraphrastiques que ces relations permettent d'établir sont universaux (= existant dans toutes les langues) et peuvent être modélisés au moyen de formalismes universels. Nous y reviendrons.

De ce qui vient d'être dit, il s'ensuit que l'acquisition et une exploitation active de liens lexicaux sous-jacents à la paraphrase et de la paraphrase elle-même doivent être des objectifs principaux de l'enseignement de la langue, notamment de la langue seconde (ou étrangère).

Ces objectifs sont à l'heure actuelle de plus en plus reconnus : l'importance du lexique dans l'enseignement d'une langue seconde a été mise en évidence, entre autres, par Wesche et Paribakht (1996), Tréville (2000), Schmitt (2000), Nation (2001), Polguère (2003), Tremblais (2004) et Laniel (2005); celle de la paraphrase a été notée, par exemple, par Daunay (2002) et Polguère (2004). Cependant, ils sont encore loin d'être atteints. Pour s'en convaincre, il suffit de jeter un coup d'œil sur un

quelconque corpus d'apprenants, où abondent les *erreurs lexicales*, c'est-à-dire erreurs ayant comme source des connaissances lexicales insuffisantes.

Une telle erreur, identifiée dans un corpus d'apprenants que nous sommes en train d'analyser dans le cadre de notre recherche, est indiquée en (3a). Le caractère douteux de la phrase (3a) s'explique par la violation d'une contrainte portant sur la collocation *avoir (un) accouchement* : la présence quasi-obligatoire d'un modificateur de la base; cf. une phrase normale, où la contrainte en cause est respectée : *Anne risque d'avoir un accouchement difficile (prématuré)*. Pour corriger l'erreur, il faut remplacer le nom ACCOUCHEMENT 'fait d'accoucher' par le verbe correspondant, ACCOUCHER, qui n'est pas ainsi contraint (et, bien entendu, supprimer le verbe support).

- (3) a. ?**Avoir un accouchement** à domicile n'est pas une décision facile à prendre.
b. **Accoucher** à domicile n'est pas une décision facile à prendre.

Le dictionnaire de reformulation que nous comptons développer guidera l'apprenant dans ses choix lexicaux et lui permettra de faire des reformulations du type illustré ci-dessus.

Le besoin d'offrir un tel dictionnaire nous paraît évident dans le contexte actuel où il existe fort peu de ressources pédagogiques, notamment de ressources électroniques, qui visent spécifiquement l'acquisition de relations lexicales, et encore moins de la paraphrase. Ceci est en particulier vrai dans l'espace francophone. Les dictionnaires d'apprentissage développés jusqu'à présent, tels que *French LEXical dictionary* (Frlex) (Hamel *et coll.*, 1995) et le *Dictionnaire pour l'Apprentissage du Français Langue Etrangère/Seconde* (DAFLES) (Selva, 1999), soit ne sont pas destinés au public qui nous intéresse, c'est-à-dire les apprenants de niveau intermédiaire à avancé, soit traitent la langue de spécialité plutôt que la langue générale.

Le cadre théorique de notre recherche est la *Lexicologie explicative et combinatoire*, ou la LEC (Mel'čuk *et coll.* 1995), la branche lexicologique de la théorie linguistique Sens-Texte. La LEC offre deux outils sophistiqués de description lexicographique particulièrement intéressants

pour nous : les *fonctions lexicales* (Mel'čuk, 1974; Wanner., 1996) pour la description des relations lexicales, et les *règles de paraphrasage* (Žolkovskij et Mel'čuk, 1967; Mel'čuk, 1992; Milićević, 2003) pour la description des liens paraphrastiques.

Une fonction lexicale appliquée à une lexie L [= FL(L)] établit une correspondance entre L et un ensemble de lexies {L₁, L₂, ..., L_n} qui expriment auprès de L et en fonction de celle-ci un sens assez général tout en remplissant une fonction syntaxique particulière. Par exemple, la FL Magn, dont le sens est 'intense'/'très', « retourne », pour une L donnée, un ensemble de lexies qui fonctionnent comme des intensificateurs typiques de L; cf. : Magn(FROID_(N)) = *de canard, sibérien*; Magn(INTERDIRE) = *strictement, catégoriquement*; Magn(NÉCESSAIRE) = *absolument*; etc. La FL Oper₁ retourne, pour une L nominale prédicative, les verbes supports qui lient L à l'expression de son i^{ème} actant; cf. : Oper₁(BESOIN ~ de X en Y) = *avoir* [~]; Oper₁(DÉCISION ~ de X concernant Y) = *prendre* [ART ~]; Oper₂(OPÉRATION ~ par X de Y) = *subir* [ART ~]; etc. La FL S₀ 'fait de L-er' retourne, pour une L verbale/adjectivale, le nom de l'action/de l'état correspondant; cf. : s₀(ACCOUCHER) = *accouchement*; s₀(TOMBER) = *chute*; s₀(FATIGUÉ_(Adj)) = *fatigue*; etc.

Il existe une soixantaine de FL comme celles qui viennent d'être illustrées, appelées *FL standard simples*. Ces FL peuvent se combiner pour former des FL complexes dont nous verrons quelques exemples plus loin. Les FL sont universelles, c'est-à-dire applicables en principe à n'importe quelle lexie (du sens correspondant) dans n'importe quelle langue. Les règles de paraphrasage sont des règles de (quasi-)équivalence entre configurations lexicales au niveau de la structure syntaxique des phrases. Elles sont formulées en termes de FL et sont donc universelles tout comme ces dernières. On en compte une centaine. À titre d'exemple, la correction de l'erreur lexicale indiquée en (3a) ci-dessus met en jeu la règle de paraphrasage suivante :

[Règle-1] L_[V] ≡ Oper₁(S₀(L_[V])) — II → S₀(L_[V]).

Cette règle, qui décrit une *fission à verbe support*, permet de remplacer (dans un arbre syntaxique) une lexie verbale L_(V) par une configuration lexicale constituée du nom de l'action correspondant [S₀(L_[V])] et d'un verbe support de ce nom [Oper₁(S₀(L_[V]))]. Dans le cas de la phrase (3a), la

règle s'applique de droite à gauche : *avoir* [= $OP_{ER_1}(S_0(L_{[V]}))$] *un accouchement* [= $S_0(L_{[V]})$] \equiv *accoucher* [= $L_{[V]}$]. Voici d'autres équivalences que décrit la présente règle : *décider* ~ *prendre (une) décision*, *opérer* ~ *effectuer (une) opération*, *tomber* ~ *faire (une) chute*, *frapper* ~ *donner (un) coup*, etc.

La LEC a été la base théorique pour le développement de plusieurs dictionnaires, dont, notamment, les dictionnaires suivants pour le français : le *Dictionnaire explicatif et combinatoire du français contemporain*, ou le *DEC* (Mel'čuk *et coll.* 1984, 1988, 1992, 1999), le *Dictionnaire de combinatoire lexicale du français*, ou le *DiCo* (Polguère, 2000a, 2000b) et sa version internet *DiCouèbe* (<http://olst.ling.umontreal.ca/dicouebe>), ainsi que le *Lexique actif du français*, ou le *LAF* (Polguère, 2002a). Ces produits lexicographiques, dont il sera question plus loin, nous serviront comme guide pour l'élaboration de notre dictionnaire de reformulation.

Pour ce qui est de notre méthodologie, nous alimentons notre recherche par les données tirées d'un corpus d'apprenants, comme il est d'usage aujourd'hui; sur l'utilisation de corpus d'apprenants en linguistique appliquée, et notamment, en apprentissage de la langue assisté par ordinateur, voir, par exemple, Granger *et coll.* (2001) et Granger (2003). Il s'agit d'un corpus d'environ 50.000 mots, constitué de textes écrits par des étudiants universitaires de troisième année et qui portent sur le thème de la santé (par la mise en forme, le sport, la bonne alimentation, etc.). L'analyse des erreurs lexicales identifiées dans ce corpus nous sert, d'une part, à développer et valider une typologie des erreurs lexicales et, d'autre part, à établir la nomenclature de notre dictionnaire de reformulation, qui est censée refléter les difficultés spécifiques qu'éprouvent les apprenants de niveaux intermédiaire à avancé.

Soulignons, finalement, que nous développons un dictionnaire de reformulation pour le français, mais qui se base, comme on vient de le voir, sur des concepts et des outils descriptifs universels. Selon ce modèle, il sera possible de développer des dictionnaires de reformulation pour d'autres langues.

Le reste de l'article est organisé comme suit. La section 2 est consacrée au concept d'erreur lexicale et à la typologie des erreurs lexicales. Nous y

montrons comment de telles erreurs peuvent être décrites et corrigées à l'aide de fonctions lexicales et de règles de paraphrasage. La section 3 présente l'architecture envisagée de notre *Dictionnaire de reformulation*, c'est-à-dire sa macrostructure (nomenclature et organisation des lexies en vocables/champs sémantiques) et sa microstructure (format d'un article). La section 4 présente l'implémentation future du *Dictionnaire de reformulation* sous forme d'une base de données électronique.

2. Erreurs lexicales et leur correction à l'aide de règles de paraphrasage

2.1. Caractérisation du concept d'erreur lexicale

Dans le cadre de cette recherche, tout emploi inapproprié d'une lexie L causé par les connaissances insuffisantes des propriétés de L est considérée comme étant une *erreur lexicale*. Un emploi donné d'une lexie est inapproprié non seulement s'il mène à l'agrammaticalité (**obtenir une blessure*) mais aussi s'il résulte en une maladresse (*?avoir un accouchement*; cf. l'exemple (3), p. 149).

Comme une lexie L est, *grosso modo*, un ensemble de signes linguistiques ayant une même signification lexicale et ne différant que par leurs significations flexionnelles, et que le signe linguistique est le triplet (signifié, signifiant, syntactique³), les propriétés (du radical) de L peuvent concerner :

1. Le signifié de L, c'est-à-dire la définition de L.
2. Le signifiant de L, c'est-à-dire la prononciation ou l'orthographe de L.
3. Le syntactique de L, c'est-à-dire : a) la combinatoire lexicale de L (ses collocations); b) la combinatoire syntaxique de L (son régime \approx cadre de sous-catégorisation); c) la combinatoire morphologique de L (genre/classe nominal(e), classe de conjugaison, paradigme défectif, etc.); d) la combinatoire stylistique de L (registre : neutre, formel, familier, vulgaire, etc.).

Par conséquent, une erreur lexicale peut être de trois types majeurs : erreur concernant le signifié de L, erreur concernant le signifiant de L et

erreur concernant le syntactique de L. Les erreurs de ce dernier type se divisent en quatre sous-types : erreurs de combinatoire lexicale, erreurs de combinatoire syntaxique, erreurs de combinatoire morphologique et erreurs de combinatoire stylistique. Ceci est l'ossature de notre typologie des erreurs lexicales, que nous présenterons à la sous-section suivante. Une théorisation du concept d'erreur lexicale proche de la nôtre, ainsi qu'une typologie des erreurs lexicales développée pour le français langue maternelle, peuvent être trouvées dans Anctil (2005).

2.2. *Typologie des erreurs lexicales*

Nous avons jeté les bases de notre typologie des erreurs lexicales de façon déductive, en nous appuyant sur la définition du concept de lexie et de celui de signe linguistique. Cette typologie est maintenant en train d'être développée de façon empirique, au fur et à mesure de l'analyse de notre corpus d'apprenants. Dans ce qui suit, nous en présentons un fragment, en excluant de la considération les erreurs ayant trait à la combinatoire morphologique et stylistique des lexies. (Les numéros lexicographiques accompagnant certaines lexies dans les exemples illustratifs sont empruntés au *Nouveau Petit Robert*.)

I. Erreurs ayant trait au signifié de L

Il existe six sous-types d'erreurs liées au signifié de L. Les erreurs des sous-types 1 à 4 sont le résultat du mauvais choix paradigmatique d'une L pour exprimer un sens donné : il y a un décalage entre le sens visé et le signifié de la L sélectionnée. Les erreurs de sous-type 5 résultent de la violation de contraintes sur la combinatoire sémantique de L avec une autre L dans la phrase. Celles de sous-type 6 ont comme source un mauvais ajustement entre le sens de L et la situation réelle qu'elle décrit.

Erreur de sous-type 1 : Mauvais quasi-synonyme

Une erreur de ce type consiste à utiliser, au lieu de la lexie correcte L, une autre lexie fautive L', qui est un (quasi-)synonyme de L. Les quasi-synonymes partagent des composantes sémantiques importantes, mais ne sont pas nécessairement substituables dans tous les contextes (même les synonymes exacts ne le sont pas) : il suffit qu'ils le soient dans au moins quelques contextes.

- (4) **C'est aussi une bonne occasion pour regarder* [L'] \Rightarrow *observer* [L] *la nature*.

Au lieu de la lexie correcte, OBSERVER_{II.1} \approx 'regarder avec une attention soutenue, en vue de connaître', un de ses synonymes plus larges, REGARDER_{I.1}, a été utilisé. En termes lexico-fonctionnels, on a $SYN_{\square}(OBSERVER_{II.1}) = REGARDER_{I.1}$ et $SYN_{\square}(REGARDER_{I.1}) = OBSERVER_{II.1}$. Ces deux lexies sont substituables, par exemple, dans le contexte suivant : *Le chien du voisin me regardait (\approx m'observait) d'un air méchant.*

Erreur de sous-type 2 : Mauvais terme générique

Une erreur de ce type consiste à utiliser, au lieu d'une lexie correcte L, une autre lexie fautive L' en tant que terme générique de la lexie P.

- (5) **La bicyclette* [P] *est un véhicule* [L'] \Rightarrow *moyen de transport* [L] *qui ...*

La bicyclette n'est pas un type particulier de véhicule, comme le démontre l'agrammaticalité de **les voitures, les autobus, les bicyclettes et les autres véhicules*. Cf. l'encodage du lien lexico-fonctionnel correct : $Gener(BICYCLETTE) = _moyen\ de\ transport_$.

Erreur de sous-type 3 : Lexie de sens proche

Une erreur de ce type consiste à utiliser, au lieu d'une lexie correcte L, une autre lexie fautive L', qui, sans être un quasi-synonyme de L, partage avec cette dernière des composantes sémantiques assez importantes. De telles lexies sont appelées *lexies de sens proche*. Deux lexies de sens proche ne sont jamais substituables en contexte, même si elles peuvent appartenir à un même champ sémantique.

- (6) **La bicyclette permet de se rendre à beaucoup d'espaces* [L'] \Rightarrow *d'endroits* [L].

La lexie ESPACE a été choisie au lieu de la lexie ENDROIT pour exprimer le sens 'endroit'; cf. la définition de ENDROIT_{I.1} dans *le Nouveau Petit Robert* : 'partie déterminée d'un **espace**'.

Erreur de sous-type 4 : Sens fictif

Une erreur de ce type consiste à emprunter un signifiant existant (c'est-à-dire le signifiant d'une autre lexie de la langue) pour exprimer un sens qui ne lui « appartient » pas. On obtient de cette façon une lexie fictive.

- (7) *... *la pluie et les bicyclettes sont des **sûretés*** [L'] ⇒ *des **faits*** [L] *de la vie belge.*

La lexie fictive *SÛRETÉ a été créée pour exprimer le sens 'ce qui est sûrII.A.1 d'avoir lieu'. Il s'agit d'un cas assez complexe, étant donné que la lexie fictive a été dérivée à partir de l'adjectif SÛRII.A.1 'très probable' et n'est pas apparentée aux lexies du vocable SÛRETÉ [SÛRETÉII.1.2 (*sûreté publique*), SÛRETÉIII.1 (*sûreté des routes*), etc.], liées à l'adjectif SÛRII.B.1 'sans danger'.

Erreur de sous-type 5 : Incompatibilité sémantique (des sens)

Ce qui est fautif est le lien entre deux sens lexicaux donnés : le sens de la lexie L₁ entre en conflit avec celui de la lexie L₂. Ceci peut être dû, par exemple, à la mauvaise classe sémantique de la lexie L₂ réalisant un actant de L₁. Comme résultat, on obtient une phrase dont le sens est bizarre (voire absurde).

- (8) **La ville n'**encourage pas*** [L₁] *les **bicyclettes*** [L₂] ⇒ *n'**encourage pas l'usage*** *des bicyclettes.*

Le verbe ENCOURAGER exige que son deuxième actant syntaxique soit de type fait (*encourager le port du casque <l'activité physique>*), alors que dans notre exemple cet actant est de type entité.

Erreur de sous-type 6 : Incompatibilité pragmatique (du sens et de la situation)

Ce qui est fautif est le lien entre le sens lexical donné et la situation réelle particulière à laquelle il réfère.

- (9) **Vous pouvez **chercher*** ⇒ ***trouver** la plupart des ingrédients au supermarché.*

Le sens de la lexie CHERCHER ne convient pas dans la situation en cause, où c'est plutôt le résultat de l'action de chercher, exprimable par la lexie TROUVER, qui est pertinent; cf. le lien lexico-fonctionnel entre les deux lexies : $\text{Real}_1(\text{CHERCHER}) = \text{TROUVER}$.

II. Erreurs ayant trait au signifiant de L

Ici, le choix paradigmatique d'une lexie L (ou d'une expression lexicale) pour exprimer un sens donné est déficient du point de vue formel. Les erreurs formelles peuvent être de trois sous-types suivants.

Erreur de sous-type 7 : Forme analytique

Une erreur de ce type consiste à utiliser une forme « maladroite », non idiomatique, pour exprimer un sens donné. On peut parler ici de la forme décomposée, ou de la paraphrase illicite.

- (10) **une bicyclette qui ne bouge pas* \Rightarrow *stationnaire*; **votre nouvel enfant* \Rightarrow *nouveau-né*

Erreur de sous-type 8 : Forme erronée

Le signifiant de L est déformé; il peut s'agir d'une simple faute d'orthographe ou bien d'une faute plus sérieuse, ayant trait à la constitution d'un phrasème, à la structure morphologique d'un lexème, etc.

- (11) **moyen du transport* \Rightarrow *moyen de transport*; **conduiseur* \Rightarrow *conducteur*

Erreur de sous-type 9 : Forme fictive

Un signifié donné est associé à une séquence qui n'est pas un signifiant dans la langue en cause (c'est-à-dire qu'il n'est le signifiant d'aucune lexie de cette langue). Le résultat de cette association est une lexie fictive.

- (12) **... appartenant à toutes les marches de vie* \Rightarrow *couches sociales*

III. Erreurs ayant trait au syntactique de L

Nous présentons ici deux sous-types d'erreurs concernant le syntactique de L : erreurs de combinatoire lexicale (= de collocation) et erreurs de combinatoire syntaxique (= de régime).

Erreur de sous-type 10 : Mauvaise collocation

Avec les erreurs de ce sous-type, il s'agit soit d'un choix inapproprié du collocatif, c'est-à-dire de l'élément de la valeur de la FL appliquée à la base en cause, soit d'un choix inapproprié de la FL, qui retourne une valeur nulle pour la base en cause. Ces deux cas de figure sont illustrés, respectivement, en (13) et en (14).

(13) **Si vous recevez* ⇒ *attrapez un coup de soleil, ...*

Le verbe RECEVOIR n'est pas un élément de la valeur de la FL IncepOper_1 (commencer à avoir) pour la base $_COUP DE SOLEIL_;$ cf. $\text{IncepOper}_1(_COUP DE SOLEIL_) = \textit{attraper}$ [ART ~].

(14) **De temps en temps, je me soumetts à la paresse* ⇒ *je m'abandonne à la paresse et ...*

Selon son sens et sa structure actancielle, le verbe *se soumettre* en (14) devrait constituer la valeur de la FL $\text{Caus}_1\text{Oper}_1$ 'se.causer d'avoir' pour la base PARESSE. Or, la FL en cause n'est pas applicable à cette base; cf. $\text{Perm}_1\text{Oper}_1$ 'se.permettre d'avoir', qui s'applique à PARESSE et retourne la valeur *s'abandonner* [à ART ~].

Erreur de sous-type 11 : Mauvais régime

La préposition ou la conjonction introduisant un actant de L est superflue, manquante ou mauvaise.

(15) * *On se souvient lui* ⇒ ... *de lui* ; **Personne ne s'attendait de ça* ⇒ ... *à ça*.

Le Tableau 1 ci-dessous indique tous les sous-types d'erreurs lexicales figurant dans notre typologie, y compris les erreurs des sous-types 12 à 14, qui n'ont pas été décrites ci-dessus. (On notera qu'il existe un recoupement entre les erreurs ayant trait au signifiant de L et celles ayant

trait à la combinatoire morphologique de L ; ainsi la forme erronée d'une L peut être due au fait que L a été associée à la mauvaise classe flexionnelle. De tels recouvrements sont normaux dans une typologie aussi détaillée que la nôtre.)

Tableau 1 : Typologie des erreurs lexicales

	Erreurs ayant trait au signifié	Erreurs ayant trait au signifiant	Erreurs ayant trait au syntactique
Affectant la nomination	1. QSyn 2. Gener 3. Sens ^{Proche} 4. Sens ^{Fictif}	7. F ^{Analytique} 8. F ^{Erronée} 9. F ^{Fictive}	
Affectant la combinaison	Combinatoire libre 5. Sens ^{Incompatible} 6. Sit ^{Incompatible}		Combinatoire restreinte Lexicale 10. Collocation Syntactique 11. Régime Morphologique 12. Genre 13. Classe flexionnelle Stylistique 14. Registre

Les erreurs qui nous intéressent particulièrement dans le cadre de notre projet de recherche sont celles des sous-types 1 à 6 et 10. Parmi celles-ci, les erreurs les plus fréquentes dans notre corpus appartiennent aux sous-types 1, 3 et 5.

2.3. *Correction des erreurs lexicales à l'aide des règles de paraphrasage*

Nous utilisons deux types de corrections : des *corrections minimales* (= *obligatoires*) et des *corrections optimales* (= *suggérées*). Une correction obligatoire est une correction impliquant le minimum d'interventions nécessaires pour rendre la phrase grammaticale. Une telle correction a été illustrée plus haut, p. 149. Si la phrase — telle quelle ou après une correction obligatoire — est acceptable mais pas idéale (elle est

stylistiquement lourde, par exemple), on a recours à une correction suggérée.

Une correction obligatoire peut mettre en jeu une règle de paraphrasage, c'est-à-dire constituer une véritable reformulation, mais pas forcément : dans bien des cas, la correction obligatoire consiste à remplacer un mauvais élément de la valeur d'une FL, une mauvaise préposition régie, etc. Une correction suggérée, quant à elle, qui implique nécessairement le recours à une ou plusieurs règles de paraphrasage, est toujours une reformulation véritable.

Dans ce qui suit, nous illustrons une correction obligatoire d'une erreur lexicale suivie d'une correction suggérée.

- (16) a. *[*Parfois, il m'est arrivé de tomber de ma bicyclette et] d'obtenir une blessure.*
b. [...] *de subir une blessure.*
c. [...] *de me faire une blessure.*
d. [...] *de me blesser.*

Commençons par donner les informations lexicographiques pour la lexie BLESSURE₁ qui nous seront utiles dans la discussion qui suivra.

- Structure actancielle de BLESSURE₁ : ~ de individu X [par individu Y] à partie du corps Z(X) [au moyen de W].
- Fonctions lexicales de BLESSURE₁ : Oper₁ 'avoir' = *avoir* [ART ~], *porter* [ART ~], *souffrir* [de ART ~]; IncepOper₁ 'commencer à avoir' = *subir* [ART ~] | B. dans le contexte de sports compétitifs ou d'équipe ; Caus₁Oper₁ 'se.causer de commencer à avoir' = *se faire* [ART ~].

L'erreur en (16a) est due au mauvais choix de l'élément de la valeur de la FL IncepOper₁ pour BLESSURE₁. La correction minimale, dont on voit le résultat en (16b), consiste à choisir un bon élément de la valeur de cette FL de BLESSURE₁. Cependant, cette correction laisse à désirer, car *subir une blessure*₁ se dit surtout dans le contexte de sports compétitifs ou d'équipe (cf. ci-dessus, la condition sur l'usage de *subir* en tant qu'élément de la valeur de la FL IncepOper₁ de BLESSURE₁), alors que dans notre cas il s'agit d'une simple promenade à vélo. Pour effectuer une

correction optimale, deux possibilités se présentent, chacune mettant en jeu une règle de paraphrasage.

Pour passer de (16b) à (16c), c'est-à-dire de *subir* [IncepOper₁] à *se faire* [Caus₁Oper₁] *une blessure*₁, on se sert de la règle de quasi-équivalence suivante :

[Règle-2] Caus(X) \Rightarrow Incep(X).

Cette règle décrit une implication : causation d'un état de choses implique le début de cet état de choses. Dans notre cas, X = Oper₁ et la règle s'applique de droite à gauche.

Les implications sont normalement non réversibles ; ainsi, la causation de X implique le début de X, mais l'inverse n'est pas vrai. Cependant, dans ce contexte particulier, il est possible d'« inverser » l'implication parce qu'on sait que la personne allait à bicyclette et donc qu'elle était responsable de la blessure (cf. l'indice 1 auprès de Caus).

Pour passer de (16c) à (16d), c'est-à-dire de *se faire une blessure*₁ à *se blesser*, on se sert de la règle d'équivalence suivante :

[Règle-3] L_[V] \equiv Caus₁Oper₁(S_{Res}(L_[V])) — **II** \rightarrow S_{Res}(L_[V]).

Il s'agit d'une fission à verbe support, similaire à celle décrite par la **Règle-1**, p. 150. Dans notre cas, L_[V] = *se blesser* et S_{Res} 'nom du résultat' (L_[V]) = *blessure*₁; la règle s'applique de droite à gauche.

3. Architecture du « Dictionnaire de reformulation »

3.1. *Nomenclature et macrostructure du « Dictionnaire de reformulation »*

Les lexies que nous avons l'intention de décrire appartiennent au français standard international; toutefois, certaines particularités lexicales du français canadien seront prises en compte.

Le *Dictionnaire de reformulation* contiendra à peu près 1.000 lexies. De ce nombre, 30 % appartiendront au lexique de base et le reste au lexique intermédiaire à avancé (ciblant directement le profil de nos apprenants). Avec un taux de polysémie moyen de quatre lexies par vocable, le nombre de vocables décrits dans le dictionnaire devrait être d'environ 250. Ce nombre pourrait être révisé à la hausse, sachant que nous ne décrivons que

les acceptions courantes des vocables (en excluant les acceptions vieilles, trop recherchées, etc.). Bien entendu, il ne s'agit que d'un fragment du lexique censé être connu par les apprenants du FLS de ce niveau, le nombre attendu étant de 6.000 (niveau-seuil) à 10.000 lexies. Une extension du dictionnaire est envisageable dans une phase ultérieure.

La nomenclature du *Dictionnaire de reformulation* sera déterminée sur la base de l'analyse de notre corpus d'apprenants; elle inclura, notamment, des lexies dont l'utilisation correcte s'avère particulièrement difficile et leurs « voisines » dans des champs sémantiques correspondants.

L'élaboration du dictionnaire se fera par champs sémantiques, comme il est d'usage dans le cadre de la LEC. Cette façon de faire présente l'avantage de mettre mieux en évidence les similarités et les différences entre lexies sémantiquement apparentées et mène ainsi vers une cohérence descriptive plus grande. Par exemple, la description des lexies appartenant au vocable POUVOIR_(V) se fera en même temps que celle des lexies appartenant à leurs champs sémantiques respectifs; cf. :

Tableau 2 : Trois acceptions du vocable POUVOIR et leurs champs sémantiques respectifs

Acceptions du vocable POUVOIR _(V)	Champs sémantiques auxquels appartiennent les acceptions du vocable
<p>POUVOIR₁ <i>Je ne peux pas soulever cette valise.</i></p>	<p>'capacité' POUVOIR_{(V)1}, SE DÉBROUILLER, CAPACITÉ, POUVOIR_{(N)1}, HABILITÉ, DEXTÉRITÉ, CAPABLE, HABILE, DÉBROUILLARD, etc.</p>
<p>POUVOIR₂ <i>Tous les citoyens âgés d'au moins 18 ans peuvent voter.</i></p>	<p>'droit' POUVOIR_{(V)2}, DROIT, PERMISSION, AUTORISATION, PRÉROGATIVE, PRIVILÈGE, etc.</p>

POUVOIR₃

Cela peut très bien arriver.

'probabilité'

POUVOIR_(V)₃, PROBABILITÉ, POSSIBILITÉ, ÉVENTUALITÉ, CHANCE, RISQUE, PROBABLE, ENVISAGEABLE, PEUT-ÊTRE, PROBABLEMENT, etc.

La consultation du dictionnaire, quant à elle, pourra se faire soit par vocables soit par champs sémantiques, soit encore suivant d'autres modes de recherche de l'information lexicographique, dont il sera question à la section 4.

3.2. *Microstructure du « Dictionnaire de reformulation » : article de dictionnaire standard*

Nous comptons adopter le format général de l'article de dictionnaire développé dans le cadre de la LEC. Ce format diffère quelque peu dans les trois dictionnaires qui ont été mentionnés plus haut, p. 150-151. Le DEC utilise une formalisation mettant en jeu des définitions paraphrastiques (c'est-à-dire faites par décomposition du sens de la lexie-vedette en termes de sens plus simples) et des relations lexicales décrites au moyen du formalisme classique de fonctions lexicales. Le DiCo se sert de pseudo définitions mettant en jeu les étiquettes sémantiques (caractérisation taxinomique de la lexie-vedette et de ses actants : fait, acte, processus, événement, relation, caractéristique, etc.), en gardant le formalisme classique des fonctions lexicales pour la description des relations lexicales. Finalement, le LAF, une version grand public du DiCo, utilise le même format de définitions que ce dernier, mais encode les relations lexicales en termes de fonctions lexicales « vulgarisées », c'est-à-dire adaptées à une clientèle non initiée (voir à ce propos Popovic 2003). Nous opterons pour le style DiCo, avec les modifications suivantes :

- La description de relations lexicales de la lexie-vedette en termes de fonctions lexicales vulgarisées, avec l'indication, en parallèle, de l'encodage classique.
- L'ajout de mises en garde dans des cas particulièrement problématiques (par exemple, cas d'interférence avec la langue maternelle des apprenants, les paronymes, etc.).

- L'ajout de la composante reformulative : chaque article contiendra un renvoi explicite aux règles de paraphrasage s'appliquant à la lexie-vedette.

Nous donnons maintenant, à titre d'illustration, l'article de dictionnaire d'une des acceptations du vocable POUVOIR dont il a été question plus haut.

Identification de la lexie-vedette L

POUVOIR₁, verbe transitif

Caractérisation sémantique de L

capacité: être X ~ Y-er

Relations lexicales de L

Quasi-synonyme	(QSyn) : <i>savoir</i> _{IL.1}
Caractéristique de X peut faire Y	(S ₀) : <i>capacité</i>
X tel qu'il peut faire Y	(A ₁) : <i>capable</i>
Y tel que X peut le faire	(A ₂) : <i>possible</i>
Question posée à X en tant que demande polie de faire Y	<i>Pouvez-vous <Peux-tu> Y-er ?</i>

Schéma de régime de L

X = sujet = N Y = CO^{dir} = V_{Inf}, LE_{Pron.neut}

Particularités syntaxiques de L

En style soutenu, POUVOIR₁ accepte la négation sans PAS :

Je ne pourrais vous le dire.

Exemples d'usage typique de L

Balthazar peut résoudre ce problème. | Cet enfant ne peut pas encore marcher.

Paraphrases mettant en jeu L

X peut Y-er ≡ X sait_{IL.1} Y-er ≡ X a la capacité de Y-er ≡ X est capable de Y-er ≡ Y-er est possible pour X; X peut Y-er ≈ X habituellement fait Y.

4. Implémentation du « Dictionnaire de reformulation »

Nous avons l'intention de développer une base de données électronique (MSAccess) sur le fond de celle du Dicouèbe (cf. section 1) et dont l'interface d'interrogation (MySQL) permettra une consultation de l'ensemble des lexies du dictionnaire en fonction d'un certain nombre de critères de recherche, qui pourront être combinés. Ainsi, par exemple, l'apprenant aura la possibilité de faire une consultation :

- En mode *classique* : une recherche portant sur une lexie donnée (avec possibilité d'une recherche plus fine avec l'aide de jokers – permettant d'ignorer le début/la fin d'un mot, etc.).
- En mode *alphabétique* : une recherche à partir d'une liste de lexies classées par ordre alphabétique.
- En mode *partie du discours* : une recherche permettant d'afficher l'ensemble des adjectifs, des noms, des verbes, etc.
- En mode *champs sémantique* : une recherche portant sur une lexie donnée à l'intérieur de son champ sémantique (liste prédéfinie de champs sémantiques); cf. section 3.
- En mode *collocations* : une recherche d'un collocatif ou d'une base pour une lexie donnée (liste prédéfinie de collocations).
- En mode *relations lexicales* : une recherche portant sur une FL et qui permettra d'afficher l'ensemble des lexies de la base de données répondant aux critères de sélection (tous les $Oper_i$ de la base de données, tous les $Magn$, etc.).

Une fois la recherche d'une lexie réussie, l'apprenant pourra consulter *l'article de la lexie-vedette* (cf. section 3) *dans sa totalité* (tous les champs qui la décrivent apparaîtront à l'écran) ou encore sélectionner, parmi les champs disponibles, *un ou des champs en particulier* qu'il souhaiterait consulter (définition, relations lexicales, exemples d'usage, emploi fautif/à éviter, etc.).

Une possibilité de multi-fenêtrage est ici considérée de façon à ce que l'apprenant puisse consulter, pour les comparer, deux articles de dictionnaire à la fois (par exemple, ENDROIT et PLACE ou bien PERSONNE et INDIVIDU, etc.).

Hypertexte aidant, l'apprenant pourra naviguer au sein de la base de données en cliquant sur les diverses lexies se trouvant à même l'article d'une lexie-vedette consultée, lorsqu'elles font l'objet d'articles du dictionnaire.

Le chemin parcouru à travers la base de données s'affichera au fur et à mesure en haut d'écran de sorte que l'apprenant puisse à tout moment revenir en arrière sans perdre le fil de sa recherche.

Une aide en ligne présentera à l'apprenant des explications concernant les objectifs du dictionnaire, l'étendue et la structure de sa base de données, la façon de l'interroger, la nature des formalismes employés, en particulier de ceux utilisés pour décrire les relations lexicales et paraphrastiques.

Remerciements

Cet article a été lu et commenté par Dominic Anctil, Yves Bourque, Igor Mel'čuk et Muriel Peguret, que nous remercions chaleureusement.

Bibliographie

- Anctil, D. (2005). *Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et stratégie de résolution de problèmes lexicaux*. Montréal : Université de Montréal. Mémoire de maîtrise. [www.olst.umontreal.ca/pdf/AnctilMA2005.pdf]
- Daunay, B. (2002). *La paraphrase dans l'enseignement du français*. Bern : Lang.
- Fuchs, C. (1981). *La paraphrase*. Paris : Ophrys.
- Granger, S. (2003). Error-Tagged Corpora and CALL : A promising synergy. *CALICO*. 20:3.465-480.
- Granger, S., Vandeventer, A., et Hamel, M.-J. (2001). Analyse de corpus d'apprenants pour l'ELAO basé sur le TAL. *TAL*. 42:2.609-622.
- Hamel, M.-J., Nkwenti-Azeh, B., et Zahner, C. (1995). The Conceptual Dictionary in CALL. *In Proceedings of the EuroCALL95 Conference*, Valencia : SPUPV. 509-518.

- Laniel, D. (2005). Le vocabulaire du français langue seconde : de parent pauvre à invité d'honneur des communications assistées par ordinateur. *Revue de l'AQFLS*. 25:2.73-100.
- Mel'čuk I. (1974). *Opyt teorii lingvističeskix modelej Smysl ~ Tekst*. Moskva : Nauka.
- Mel'čuk, I. (1992). Paraphrase et lexique : la théorie Sens-Texte et le « Dictionnaire explicatif et combinatoire ». In Mel'čuk, et coll. 9-59.
- Mel'čuk, I., et coll. (1984, 1988, 1992, 1999). Le dictionnaire explicatif et combinatoire du français contemporain. *Recherches lexicosémantiques I-IV*. Montréal : Presses de l'Université de Montréal.
- Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Miličević, J. (2003). *Modélisation sémantique, lexicale et syntaxique de la paraphrase*. Montréal : Université de Montréal. Thèse de doctorat.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge : CUP.
- Polguère, A. (2000a). Towards a Theoretically Motivated General Public Dictionary of Semantic Derivations and Collocations for French. In *Proceedings of EURALEX 2000*. Stuttgart. 517-527.
- Polguère, A. (2000b). Une base de données lexicale du français et ses applications possibles en français. *Revue de linguistique et de didactique des langues*. 21.75-97.
- Polguère, A. (2003). Collocations et fonctions lexicales : pour un modèle d'apprentissage. In Grossman, A., et Tutin, A. (dir.). *Les collocations, analyse et traitement*. Amsterdam : Éditions De Werelt. 23-34.
- Polguère A. (2004). La paraphrase comme outil pédagogique de modélisation des liens lexicaux. In Calaque, E., et David, J. (dir.). *Didactique du lexique : contextes, démarches, supports*. Bruxelles : De Boeck. 115-125. [coll. Savoirs en pratique]
- Popovic, S. (2003). *Paraphrasage des liens de fonctions lexicales*. Montréal : Université de Montréal. Mémoire de maîtrise.
- Selva, T. (1999). *Ressources et activités pédagogiques dans un environnement informatique d'aide à l'apprentissage lexical du*

français langue seconde. Thèse de doctorat. Université de Franche-Comté.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge : CUP.

Tremblais, O. (2004). *Une approche structurée de l'enseignement/l'apprentissage du lexique en français langue maternelle basée sur la lexicologie explicative et combinatoire*. Montréal : Université de Montréal. Mémoire de maîtrise.

Tréville, M.-C. (2000). *Vocabulaire et apprentissage d'une langue seconde*. Montréal : Logiques.

Wanner, L. (dir.). (1996). *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia : Benjamins.

Wesche, M., et Paribakht, T. S. (1996). Assessing second language vocabulary knowledge : depth versus breath. Harley, B. (dir.). *Vocabulary Learning and Teaching in a Second Language*. CMLR. 53:1.13-40.

Žolkovskij, A., et Mel'čuk, I. (1967). O semantičeskom sinteze. *Problemy kibernetiki*, 19.177-238. [Traduction française : Sur la synthèse sémantique, 1970. *TA Informations*. 2.1-85].

¹ Le projet *Dire autrement : Vers un dictionnaire de reformulation pour les apprenants du français langue seconde* (2005 ~ 2008) bénéficie d'une subvention du Fonds CCRSH (n° 410 2005 0177); l'adresse internet du projet : <http://direautrement.french.dal.ca>.

² Une lexie [= unité lexicale] est un mot ou une locution pris(e) dans une acception bien déterminée. Dans notre approche, une lexie est l'unité de base de description lexicographique : elle est décrite dans un article de dictionnaire. Un vocable est un ensemble de lexies ayant le même signifiant et partageant des composantes sémantiques importantes. Un vocable forme un superarticle de dictionnaire.

³ Le syntactique d'un signe linguistique comporte les informations sur les traits idiosyncrasiques de ce dernier, c'est-à-dire les traits que l'on ne peut déduire ni de son signifié ni de son signifiant.