

Analyse d'erreurs lexicales d'apprenants du FLS : démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage

Marie-Josée Hamel et Jasmina Milićević

Dalhousie University

Cet article traite d'erreurs lexicales, soit d'erreurs ayant pour cause des connaissances insuffisantes des propriétés — sémantiques, formelles et combinatoires — des unités lexicales. De telles erreurs sont fréquentes dans les productions écrites des apprenants de langue. Elles témoignent de lacunes à combler en ce qui a trait aux connaissances lexicales, notamment, à la maîtrise des relations lexicales et de la paraphrase en production de textes. Notre analyse d'un corpus de textes d'apprenants du français langue seconde, basée sur une typologie d'erreurs lexicales que nous avons développée, révèle un nombre élevé d'erreurs lexicales, en particulier d'erreurs ayant trait aux propriétés sémantiques des lexies ainsi qu'à leurs propriétés de cooccurrence lexicale. Ces résultats serviront à l'élaboration d'un dictionnaire d'apprentissage s'appuyant sur la théorie linguistique Sens-Texte et dont la conception sera orientée de manière à faciliter l'acquisition et l'utilisation active des relations lexicales et de paraphrase.

This article deals with lexical errors, i.e., errors stemming from an insufficient knowledge of properties — semantic, formal and combinatorial — of lexical units. Such errors are frequent in written texts produced by language learners. They indicate that there are gaps to be filled in the learner's lexical knowledge, in particular, when it comes to learners' ability to efficiently use lexical and paraphrastic relations in text production. Our analysis of a corpus of texts produced by learners of French as a second language, based on a typology of lexical errors we have developed, has revealed a high number of lexical errors, in particular those involving meaning properties of lexical units and their restricted lexical cooccurrence properties. These findings will be used for the development of a learners' dictionary based on the Meaning-Text linguistic theory and designed in such a way as to facilitate the acquisition and active use of lexical and paraphrastic relations.

Adresse pour correspondance : Marie-Josée Hamel et Jasmina Milicevic, French Department, Dalhousie University, 6135 University Ave, #1114, Halifax, Nouvelle-Écosse, Canada B3H 4P9
Courriel: marie.hamel@dal.ca ; jmilicev@dal.ca.

Introduction

Dans cet article, nous allons traiter d'erreurs lexicales dans des textes écrits par des apprenants du français langue seconde. Elles ont pour cause des connaissances insuffisantes des propriétés — sémantiques, formelles et de cooccurrence — des unités lexicales, des erreurs comme celle-ci : “Buvez beaucoup de *fluides*” où la lexie FLUIDE_(n) a été utilisée plutôt que la lexie visée LIQUIDE_(n). De telles erreurs sont fréquentes dans les productions écrites (mais aussi orales) des apprenants de langue et témoignent de leurs connaissances lexicales partielles, souvent floues. De fait, Laniel 2005 constate que « les difficultés que rencontrent les apprenants lorsqu'ils ont à faire des choix lexicaux et les fréquentes erreurs qu'ils commettent sont fondamentalement dues à leur méconnaissance du système lexical de la langue cible » (p. 80).

C'est dans le cadre d'un projet de recherche en lexicologie appliquée¹ que nous nous sommes intéressées aux erreurs lexicales des apprenants, en particulier ceux du français langue seconde de niveau intermédiaire-avancé. L'objectif ultime de ce projet est le développement d'une ressource lexicale destinée à cette clientèle d'apprenants et qui sera orientée vers la production, à savoir un dictionnaire de reformulation que nous appelons *Dire Autrement*. Ce dictionnaire d'apprentissage vise spécifiquement l'enrichissement des connaissances lexicales des apprenants, et ce par l'exploitation des *relations lexicales* qu'entretiennent entre eux les mots dans le lexique ainsi que par celle de la *paraphrase* ; leur maîtrise, nous le croyons, se situant au cœur de la compétence lexicale de l'apprenant.

Notre article est organisé en quatre sections. Dans la première, nous faisons le point sur la compétence lexicale de l'apprenant et discutons en particulier de l'importance de la maîtrise des relations lexicales et de paraphrase, notamment chez l'apprenant de niveau intermédiaire-avancé. Dans la deuxième section, nous proposons une définition du concept d'erreur lexicale ainsi qu'une typologie permettant une description fine et détaillée des erreurs lexicales des apprenants. Nous discutons ensuite de la pertinence de l'analyse d'erreurs dans les corpus d'apprenants comme mesure possible de la compétence lexicale et pour le développement de ressources didactiques. Dans notre troisième section, nous présentons l'analyse d'erreurs lexicales que nous avons effectuée sur un corpus d'apprenants du FLS. Nous rappelons d'abord le contexte de notre recherche : cadre théorique et démarche empirique. Nous décrivons ensuite le corpus d'apprenants ainsi que notre méthode d'analyse des données. Enfin, nous présentons nos résultats et notre interprétation de ceux-ci. Dans notre dernière section, nous discutons de la réutilisation de ces résultats à des fins didactiques, c'est-à-dire pour le développement du dictionnaire mentionné plus haut.

La compétence lexicale de l'apprenant

La compétence lexicale est définie par le Conseil de l'Europe comme étant « la connaissance et la capacité d'utiliser le vocabulaire d'une langue » (Conseil de l'Europe, 2000, p. 87). Selon Tréville, elle est formée de « quatre composantes basées sur des connaissances d'ordre linguistique, discursif, référentiel et socioculturel » (Tréville 2000, p. 88). Ces connaissances dynamiques, tant réceptives que productives, tant en étendue qu'en profondeur (ibid., p. 65, citant Wesche et Paribakht, 1996) portent sur le sens et la forme des mots, ainsi que sur leur fonction et leur positionnement linéaire dans l'énoncé (ibid., p. 64, citant Nation, 1990), sur les mots en relation avec d'autres mots (ibid., citant Laufer, 1994), sur leur valeur dans le discours et leur portée en contexte culturel. S'ajoute à ces quatre composantes « une composante dite stratégique » (ibid., p. 88), soit « l'aptitude à manipuler les mots à l'intérieur de leurs réseaux associatifs dans le but de contourner, clarifier, résoudre un problème de communication et l'aptitude à compenser la méconnaissance des mots par des procédés d'inférence à partir d'indices contextuels (en compréhension) ou de formulations approximatives, paraphrases² et définitions (en production) » (ibid., p. 89). En ce qui nous concerne, la compétence lexicale passe essentiellement par la maîtrise des relations lexicales et de paraphrase.

Maîtrise des relations lexicales et de paraphrase

La compétence lexicale présuppose en particulier, et nous l'avons dit plus haut, la maîtrise par l'apprenant des liens entre les lexies de la langue (une lexie est, grosso modo, un mot ou une locution prise dans une acception bien déterminée ; une lexie correspond à un sens particulier d'un mot polysémique), soit plus précisément des relations lexicales suivantes (ex. 1–2).

(1) Les relations de 'dérivation' sémantique :

- a. synonyme de L (*respecter* ~ *estimer*)
- b. antonyme de L (*respecter* ~ *mépriser*)
- c. conversif de L (*vendre* ~ *acheter*)
- d. nominalisation de L (*respecter* ~ *respect*)
- e. nom d'un actant de L (*bicyclette* ~ *cycliste*)
- f. nom d'instrument de L (*couper* ~ *couteau*)

Un dérivé sémantique d'une lexie L_1 est une lexie ou une expression lexicale L_2 liée à L_1 par : la relation de synonymie, d'antonymie ou de conversion, une nominalisation de L_1 , le nom typique d'un actant de L_1 , le nom d'instrument utilisé pour accomplir l'action décrite par L_1 , comme dans les exemples cités (mais notamment aussi : l'adjectif typique caractérisant un des actants de L_1 , le nom du lieu de l'action décrite par L_1 , etc.).

(2) Les relations de cooccurrence, ou de collocation :

- a. intensifieur de L (une chaleur *accablante*, une montagne de problèmes) ;
- b. verbe support de L (*essuyer* un échec, les applaudissements *retentissent*) ;
- c. verbe de réalisation de L (*remporter* un prix, *résoudre* un problème) ;
- d. verbe marquant le début du fait dénoté par L (une avalanche *se déclenche*, *attraper* un rhume), etc.

Ainsi, une collocation est une expression lexicale semi-phraséologisée constituée d'une base L_1 , choisie pour son sens, et d'un collocatif L_2 (en italiques dans nos exemples), choisi pour exprimer auprès de la base et en fonction de celle-ci un sens assez général. Le collocatif s'exprime de façon contrainte, en fonction des propriétés sémantiques de la base. Dans certains cas, la position grammaticale de la base est contrainte par les propriétés syntaxiques du collocatif. C'est le cas du collocatif verbal, par exemple, qui contrôle la position sujet ou objet de sa base nominale (le soleil_{<subj>} *brille*, *solliciter* un conseil_{<obj>}).

La compétence lexicale présuppose aussi, et nous insistons sur ce point, la maîtrise par l'apprenant de la paraphrase, c'est-à-dire la capacité de reformuler — de 'dire autrement' — soit par exemple :

(3) Paul *respecte* Pierre \cong Paul *éprouve du respect pour* Pierre.

Une paraphrase d'une phrase P_1 est une phrase P_2 ayant (approximativement) le même sens que P_1 (ce qui est marqué par le symbole \cong). Les liens paraphrastiques entre les phrases sont fondés sur l'équivalence entre les lexies considérées au sein de constructions syntaxiques (cf. Milićević 2007). Ce sont des liens récurrents et universels. Dans le cas cité, X (Paul) *respecte* Y (Pierre) et X *éprouve du respect pour* Y, *respect* est la nominalisation du verbe *respecter* et *éprouver* est le verbe support (sémantiquement vide) qui forme avec le nom *respect* une collocation.

La capacité de paraphraser (notamment en L2) est liée à ce que James (1998) considère comme un processus de restauration de la phrase, c'est-à-dire une reformulation de celle-ci la rendant plus naturelle (p. 255). La paraphrase permet en ce sens l'accès à une correction optimale de la phrase, la rendant ainsi plus idiomatique. Nous y reviendrons plus loin.

L'exploitation des liens de dérivation sémantique, de collocation, et de paraphrase confère ainsi au discours son caractère idiomatique et spontané. Leur maîtrise contribue largement à la compétence lexicale laquelle donne lieu à une production langagière fluide et variée, proche de celle du locuteur natif : une production visée par l'apprenant de langue, en particulier l'apprenant de niveau intermédiaire-avancé.

L'apprenant de niveau intermédiaire-avancé

Plusieurs standards existent qui établissent des niveaux de compétence langagière en langue seconde et étrangère. On pense, par exemple aux *Standards linguistiques canadiens 2002* (www.langue.ca), aux *Niveaux de compétence en français langue seconde pour les immigrants adultes* (Laurier, 2000) ou encore au *Cadre européen commun de référence pour les langues* (www.coe.int/t/dg4/linguistic/CADRE/_FR.asp). Ce dernier document retient notre attention puisqu'il propose, de manière spécifique et explicite, des descripteurs ciblant la compétence lexicale, une compétence essentiellement décrite en fonction de l'étendue (voire de la quantité) du vocabulaire maîtrisé, selon une échelle à 6 niveaux. Voir le tableau 1.

La compétence lexicale de l'apprenant se mesure ainsi par « la quantité, l'étendue et la maîtrise du vocabulaire », vus comme étant des « paramètres essentiels de l'acquisition de la langue et, en conséquence, de l'évaluation de la compétence langagière de l'apprenant » (Conseil de l'Europe, 2000, p. 115).

Selon cette échelle, l'apprenant de niveau intermédiaire-avancé — lequel nous intéresse dans le cadre de cette étude — se situe entre les niveaux B1 *Seuil* et B2 *Indépendant (Avancé)*. Il possède un vocabulaire d'une certaine étendue et dont la maîtrise est bonne, en particulier celle du vocabulaire de base (les mots courant du lexique, dont la fréquence est élevée). Il a souvent recours à des périphrases pour contourner ses lacunes lexicales et sa production comporte encore des erreurs lexicales relevant entre autres de la mauvaise sélection de lexies.

Concept d'erreur lexicale

Le concept d'erreurs en apprentissage d'une langue est en général relié aux notions d'ignorance de la norme linguistique en usage et, par conséquent, de déviance par rapport à celle-ci, cette déviance se mesurant en termes de grammaticalité, d'acceptabilité, d'exactitude (James, 1998, p. 64).

Dans le cadre de notre recherche (cf. Milićević et Hamel, à paraître) et tel qu'annoncé dans notre introduction, nous définissons une erreur lexicale comme :

Tout emploi inapproprié d'une lexie L ayant comme cause des connaissances insuffisantes de ses propriétés sémantiques, formelles et/ou de co-occurrence. Un emploi est jugé inapproprié s'il mène à l'agrammaticalité, mais aussi s'il résulte en une maladresse.

Typologie des erreurs lexicales

Les erreurs lexicales peuvent affecter chacune des trois composantes du signe linguistique (son *signifié*, son *signifiant* et sa *forme*). C'est sur cette base que nous distinguons trois classes majeures d'erreurs lexicales :

Tableau 1 : Niveaux de compétence lexicale du *Cadre européen de référence pour les langues* (Conseil de l'Europe, 2000, p. 88–89)

Niveau	Étendue du vocabulaire	Maîtrise du vocabulaire
C2 : Maîtrise	Bonne maîtrise d'un vaste répertoire lexical d'expressions idiomatiques et courantes avec la conscience du niveau de connotation sémantique.	Utilisation constamment correcte et appropriée du vocabulaire.
C1 : Autonome	Bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement les lacunes par des périphrases avec une recherche peu apparente d'expressions et de stratégies d'évitement. Bonne maîtrise d'expressions idiomatiques et familières.	À l'occasion, petites bévues, mais pas d'erreurs de vocabulaire significatives.
B2 : Indépendant (ou avancé)	Bonne gamme de vocabulaire pour les sujets relatifs à son domaine et les sujets plus généraux. Peut varier sa formulation pour éviter des répétitions fréquentes, mais des lacunes peuvent encore provoquer des hésitations et l'usage de périphrases.	L'exactitude du vocabulaire est généralement élevée bien que des confusions et le choix de mots incorrects se produisent sans toutefois gêner la communication.
B1 : Seuil	Un vocabulaire suffisant pour s'exprimer à l'aide de périphrases sur la plupart des sujets relatifs à sa vie quotidienne tels que la famille, les loisirs et les centres d'intérêts, le travail, les voyages, l'actualité.	Une bonne maîtrise du vocabulaire élémentaire, mais des erreurs sérieuses se produisent encore quand il s'agit d'exprimer une pensée plus complexe.
A2 : Survie (ou intermédiaire)	Un vocabulaire suffisant pour mener des transactions quotidiennes courantes dans des situations et sur des sujets familiers. [...]	Un répertoire restreint ayant trait à des besoins concrets.
A1 : découverte	Un répertoire élémentaire de mots isolés et d'expressions relatifs à des situations concrètes particulières.	Pas de descripteur disponible.

1. Les erreurs ayant trait aux connaissances insuffisantes du *sens* de la lexie L — soit de son *signifié*. Ces erreurs ont trait à la propriété centrale de L, soit sa définition.
2. Les erreurs ayant trait aux connaissances insuffisantes de la *forme* de la lexie L — soit de son *signifiant*. Ces erreurs concernent la propriété de codage de L, soit sa forme écrite (orthographe) et orale (prononciation).

3. Les erreurs ayant trait aux connaissances insuffisantes de la *cooccurrence* de la lexie L avec d'autres lexies de la langue. Ces erreurs mettent en jeu les propriétés de combinatoire syntaxique (son régime : c'est-à-dire, *grosso modo*, son cadre de sous-catégorisation) et lexicale de L (ses collocations).

Nos trois classes majeures d'erreurs se subdivisent en plusieurs sous-classes pour donner lieu à une typologie fine et détaillée d'erreurs lexicales, celle-ci s'étant précisée au fur et à mesure d'une analyse empirique menée sur un corpus d'apprenants. Notre typologie est orientée sur l'analyse linguistique de la production de l'apprenant (ce qu'il a écrit et qui est erroné) et non sur celle de son comportement langagier (ce qu'il ne comprend pas et/ou le problème sous-jacent à son incompréhension) ou encore de ses stratégies d'apprentissage (de compensation, de contournement et/ou d'évitement, par exemple). Il s'agit, en l'occurrence, d'une typologie descriptive des erreurs lexicales de l'apprenant par opposition à une typologie explicative de celles-ci. Les erreurs relevées sont classifiées par rapport à la correction visée dans le contexte donné : soit en référence à l'écart entre la lexie erronée et la lexie visée ou encore à l'incompatibilité qui s'établit entre la lexie erronée et une autre lexie dans la phrase ou avec le contexte pragmatique. Les maladroites, créées par exemple par des répétitions abusives ainsi que des variations de registre de langue considérées inappropriées dans le contexte, sont aussi relevées comme des erreurs. Voici la typologie résultante (les exemples sont tirés du corpus d'apprenants décrit dans la section suivante).

1. Erreurs de sens

Nous avons identifié six sous-classes d'erreurs de *sens*. Les quatre premières sous-classes d'erreurs marquent un mauvais choix (paradigmatique) d'une lexie L pour exprimer un sens visé. Ce sont des erreurs où l'on constate un décalage plus ou moins grand entre le sens de la lexie visée L et celui de la lexie choisie par l'apprenant. Les deux dernières sous-classes concernent, respectivement, l'incompatibilité sémantique entre la lexie choisie par l'apprenant et une autre lexie dans la phrase et l'incompatibilité entre la lexie choisie et la situation réelle que la lexie est censée décrire.

(Quasi-)synonyme

L'apprenant choisit, au lieu de la lexie visée L, un (*quasi*-)synonyme de L qui est inapproprié dans le contexte. Les (*quasi*-)synonymes ont en commun la plupart ou toutes les composantes de sens, mais ne sont pas substituables dans tous les contextes. Par exemple :

- (4) a. Buvez beaucoup de *fluides*. → liquide
- b. Plusieurs joggeurs auraient *vu* le crocodile. → aperçu

- c. Les policiers ont *commencé* une recherche. → entamé

Sens proche

L'apprenant choisit, au lieu de la lexie visée L, une lexie de *sens proche*. Les lexies de *sens proche* partagent des composantes importantes de sens, mais ne sont jamais substituables en contexte. Par exemple :

- (5) a. Faites des exercices *nautiques*. → aquatiques
 b. Elle a bouclé ses chaussures de ski et *est partie* vers les pentes. → s'est dirigée
 c. Ils jouent un grand rôle dans la *production* du cancer. → formation

Sens fictif

L'apprenant associe le signifiant d'une lexie existante avec un sens qui 'ne lui appartient pas', créant ainsi une lexie fictive. Par exemple :

- (6) a. Les *écrous* sont des aliments riches en protéines. → fruits à coque/noix
 b. [...] ce qui m'a permis de *trafiquer* dans mon village. → circuler
 c. Il était en train de nourrir le reptile au moment de son *échappement*. → évasion

Générique

L'apprenant choisit un *générique* (un hyperonyme) inapproprié L₂ pour une lexie L₁. Par exemple :

- (7) a. La bicyclette [L₁] est *une méthode* [L₂] de transport. → un moyen
 b. Ce moyen de transport est constitué des pneus [L₁], des pédales [L₁], une selle [L₁] ainsi que d'autres *choses* [L₂] techniques qui [...]. → éléments
 c. [...] ces produits qui se trouvent aussi dans les *objets* [L₂] comme les jeans [L₁]. → vêtements

Sens incompatible

L'apprenant choisit une lexie L₂ dont le *sens* est *incompatible*, c'est-à-dire entre en conflit, avec celui d'une lexie L₁ dans la phrase. Par exemple :

- (8) a. En choisissant de ne plus appuyer [L₁] *la consommation* [L₂] de la viande [...]. → la cause de la consommation
 b. *Une bicyclette* [L₂] est une bonne façon [L₁] de [...]. → Faire de la bicyclette
 c. Cela mène à une augmentation [L₁] *des maladies* [L₂] du cœur. → du taux des maladies

Situation incompatible

L'apprenant choisit, au lieu de la lexie visée L, une lexie dont le sens est 'mal ajusté' à la situation pragmatique, c'est-à-dire à la situation réelle qu'il veut décrire en utilisant la lexie en cause. Par exemple :

- (9) a. La bicyclette permet *d'arriver* à n'importe quel endroit. → de se rendre
 b. J'ai dû *visiter l'hôpital*. → me faire hospitaliser
 c. Conseils pour *être* végétarien → devenir

2. Erreurs de forme

Ces erreurs consistent en la création de formes (de *signifiants*) erronées. Nous avons déterminé trois sous-classes d'erreurs de *forme*.

Forme erronée

L'apprenant, suite à une faute d'orthographe ou de formation de mot, crée une *forme* lexicale *erronée* plutôt que la forme de la lexie visée L. Par exemple :

- (10) a. un *conduiseur* → conducteur
 b. un régime faible en *grasses* → gras
 c. Pour avoir un corps parfait *instamment* → instantanément

Forme fictive

L'apprenant crée une forme lexicale qui n'existe pas dans la langue plutôt que la forme de la lexie visée L. Par exemple :

- (11) a. les *concernes* de rhume → problèmes
 b. Vous pouvez *cycler* sur la rue. → rouler
 c. Ces pensées sont interrompues par des *souviens* de [...]. → souvenirs

Forme analytique

L'apprenant a recours à une explication qui représente une décomposition maladroite du sens de la lexie visée L. Par exemple :

- (12) a. Sa jambe *est devenue prise* dans la chaîne. → s'est coincée
 b. *une bicyclette d'une seule roue* → un monocycle
 c. [...] qui *a causé lui de voler* au-dessus du guidon. → l'a projeté

3. Erreurs de cooccurrence

Nous comptons deux sous-classes d'erreurs de *cooccurrence* : les erreurs de collocation portant sur les propriétés de combinatoire lexicale restreinte des lexies et les erreurs de régime portant sur leurs propriétés de combinatoire syntaxique restreinte.

Erreur de collocation

L'apprenant fournit, au lieu de la lexie visée L, une lexie inappropriée comme collocatif de la base qu'il a choisie. Par exemple :

- (13) a. Si vous *recevez* un coup de soleil [. . .]. → attrapez
 b. Finalement, je *donne* un cri d'effort. → pousse
 c. le *Président* du zoo → le directeur

Erreurs de régime

Soit l'apprenant utilise une préposition/une conjonction inappropriée pour introduire un actant d'une lexie, soit il omet la préposition/la conjonction ou encore utilise une mauvaise forme grammaticale d'actant. Par exemple :

- (14) a. Cette philosophie vous encourage *de* ne pas modifier [. . .]. → à
 b. Pour *discuter vos options* d'accouchement → discuter de vos options
 c. Les policiers ont demandé à tous de *leur* contacter. → les contacter

Analyse des erreurs lexicales

L'analyse des erreurs lexicales dans les corpus d'apprenants de langue devrait permettre de dégager des problèmes/des lacunes reliés à la compétence lexicale de l'apprenant. Certaines analyses ont démontré l'importance des erreurs lexicales, soit un taux élevé d'erreurs lexicales par rapport à d'autres types d'erreurs, dans les corpus d'apprenants de langue (cf. James, 1998), et en particulier d'apprenants de l'anglais langue seconde (cf. Granger, 1998). En ce qui a trait plus particulièrement au français, une analyse menée sur le corpus FRIDA (Granger, Vandeventer et Hamel, 2001 ; Granger, 2003), un corpus d'apprenants du FLS, révèle que les erreurs ayant trait au *sens* des lexies occupent en importance la troisième place parmi l'ensemble des erreurs analysées (tous types confondus) dans ce corpus. Ajoutées aux erreurs de propriétés de combinatoire des lexies (soit des erreurs de *cooccurrence*), les erreurs lexicales se classent en fait en deuxième position dans le palmarès des erreurs les plus fréquentes de ce corpus d'apprenants du français langue seconde. Une autre étude (Anctil, 2005) menée sur un corpus de français langue maternelle et, contrairement à la précédente, centrée exclusivement sur l'analyse des problèmes lexicaux (erreurs et maladresses lexicales), souligne l'importance des erreurs de *forme* (près du tiers des erreurs commises) ainsi que de combinatoire syntaxique (de *cooccurrence*) dans ce type de corpus.

Les résultats de telles analyses empiriques permettent de dégager des profils plus précis d'apprenants, et peuvent entre autres servir au développement de ressources didactiques 'sur mesure' pour des clientèles bien ciblées. C'est le cas des résultats d'analyse du corpus FRIDA (op. cit.), lesquels ont servis dans l'élaboration des exercices grammaticaux du logiciel FreeText (Hamel,

2006), exercices qui tiennent précisément compte de faiblesses spécifiques relevées dans ce corpus d'apprenants. C'est dans cette même perspective que s'inscrit notre recherche, avec une visée cette fois orientée vers l'élaboration d'une ressource lexicale.

Aucune étude à notre connaissance n'a porté à ce jour sur l'analyse exclusive des erreurs lexicales dans les corpus d'apprenants du français langue seconde (cf. Nesselhauf, 2005, pour les erreurs de collocation en anglais langue seconde). C'est ce que nous avons fait dans le cadre du projet de recherche *Dire autrement*, et dans le but explicite d'exploiter les données à des fins didactiques.

Analyse des erreurs lexicales dans le corpus *Dire Autrement*

Rappel du contexte de la recherche

Cadre théorique

Le cadre théorique sur lequel s'appuie notre recherche est celui de la lexicologie explicative et combinatoire de la théorie linguistique Sens-Texte (Mel'čuk *et al.*, 1995 ; Mel'čuk (dir.), 1984–1999). Il a l'avantage d'une part de s'intéresser à la langue sous l'angle de la production (du sens vers le texte), et de manière spécifique aux relations lexicales et à la paraphrase et d'autre part de disposer d'outils puissants pour leur modélisation. L'encodage des relations lexicales se fait par le biais d'un mécanisme de fonctions lexicales (Wanner, 1996). Celui des liens paraphrastiques se fait au moyen de règles de paraphrasage (Mel'čuk, 1992 ; Milićević, 2007), c'est-à-dire de règles d'équivalence entre configurations lexicales (au niveau de la structure syntaxique des phrases), formulées aussi en termes de fonctions lexicales. Nous n'entrerons pas ici dans le détail en ce qui concerne la description des fonctions lexicales et des règles de paraphrasage. (On consultera Mel'čuk et Polguère, 1995, pour une description du concept de fonction lexicale et Milićević, 2007, pour celui de paraphrase.)

Démarche empirique

Notre démarche est empirique en ce qu'elle s'appuie au départ sur des données linguistiques — des données de corpus d'apprenants — pour déterminer (en partie) le stock lexical du dictionnaire d'apprentissage que nous comptons développer. C'est plus précisément au moyen d'une analyse détaillée des erreurs lexicales relevées dans ce corpus d'apprenants que seront identifiées en partie les lexies qui figureront dans la ressource en question. Une fois sélectionnées, ces lexies seront décrites d'après un format de description lexicographique adapté du modèle Sens-Texte. L'ensemble de ces descriptions constituera une base de données lexicales que l'apprenant pourra consulter en ligne. Nous y reviendrons brièvement dans la conclusion.

Description du corpus « Dire Autrement »

Le corpus que nous avons compilé est un corpus d'apprenants du français langue seconde dont la taille est d'environ 50 000 mots (précisément : 48 114 mots, soit 3066 phrases, soit 183 textes répartis en 6 sous corpus). Il s'agit d'un corpus écrit consistant en une collection de textes électroniques produits dans le cadre d'un cours universitaire de français de 3^e année. Les textes sont de types narratif, injonctif, persuasif et informatif. Ils ont été écrits avec le loisir d'accéder à des ressources lexicales et grammaticales. Les étudiants (environ une soixantaine) qui ont produit ces textes sont des hommes et des femmes, de vingt ans en moyenne et en majorité anglophones (à l'exception de quelques francophones de milieux minoritaires). Ils sont inscrits dans un programme de baccalauréat avec concentration en français. Le cours en question, qui s'échelonne sur une année, est obligatoire pour eux. Il porte précisément sur la compréhension et la production de divers types de textes en français.

Méthode d'analyse des données

L'analyse des données du corpus d'apprenants a comporté plusieurs étapes que voici :

1. Repérage de l'erreur lexicale

L'erreur lexicale est repérée dans le corpus d'apprenants, comme suit :

ma vitesse était trop forte

2. Identification de l'erreur lexicale

L'erreur lexicale repérée est identifiée (selon notre typologie décrite plus haut), comme suit :

ma <Coll = vitesse était trop forte>

3. Correction « minimale » de l'erreur lexicale

Une correction minimale de l'erreur lexicale identifiée est proposée. Il s'agit d'une correction obligatoire, essentiellement locale, comme suit :

ma <Coll = vitesse était trop forte = élevée>

4. Description de l'erreur lexicale

Une description de l'erreur lexicale en fonction de la correction minimale visée est proposée. Celle-ci peut comprendre une analyse du problème rencontré, des descriptions formulées à l'aide de fonctions lexicales, comme suit :

MAGN(VITESSE)= élevée [Vs MAGN(SENSATION)= forte]

(MAGN est une fonction lexicale dont le sens est 'très' ou 'intense' et dont les éléments de la valeur fonctionnent comme modificateurs de la base.)

5. Vérifications

Les étapes 1 à 4 sont vérifiées au moyen de justifications et de contre-corrrections, comme suit :

- Consultation de dictionnaires unilingues et bilingues, de cooccurrences, de synonymes, etc. ;
- Consultation de bases de données terminologiques ;
- Recherche Internet Google/Yahoo (pour la fréquence d'emploi, l'usage, etc.) ;
- Contre-codage (par au moins 2 autres correcteurs).

6. Correction « optimale » de l'erreur lexicale

Finalement, une correction maximale, c'est-à-dire optimale, est suggérée dans le cas où elle est pertinente et celle-ci peut comporter une reformulation mettant en jeu des règles de paraphrasage, comme suit :

Ma vitesse était trop élevée. → **Je roulais trop vite.**

La règle de paraphrasage correspondante :

'vitesse de X (se déplaçant en un moyen de transport α que X manœuvre) est Z'

\cong 'X se déplace (en α) à une vitesse Z'

Cette règle stipule une équivalence (\cong) entre les deux phrases de l'exemple analysé. Le sens est décomposé et met en évidence les actants (X : *Je*, α : *un moyen de transport sous-entendu* et Z : *la qualité de la vitesse*) impliqués dans les deux phrases équivalentes. À titre illustratif, le tableau 2 explique une phrase telle qu'extraite du corpus d'apprenants analysé.

Il s'agit de la phrase identifiée sous le numéro 458 d'un des sous-corpus. En 458.0, on voit la phrase originale, telle qu'elle a été produite par l'apprenant. En 458.1, les erreurs lexicales ont été relevées, identifiées et des corrections sont proposées (avec parfois des choix). En 458.2, la phrase est corrigée de façon minimale. Elle ne contient plus d'erreurs, mais n'est pas nécessairement naturelle. En 458.3, on trouve une analyse des problèmes rencontrés, et une description des relations lexicales mises en jeu proposée en termes de fonctions lexicales (qui décrivent ici des verbes support). En 458.4, la phrase corrigée de façon maximale, c'est-à-dire reformulée, 'dite autrement'.

Résultats

Les erreurs ont été extraites par type dans chacun des sous-corpus, à l'aide d'un concordancier, pour être ensuite compilées dans des fichiers Excel, et comptabilisées. Sur un nombre total de 48 114 mots dans le corpus d'apprenants *Dire autrement*, on relève 3580 erreurs lexicales standard⁴. Le tableau 3

Tableau 2 : Phrase extraite du corpus analysé *Dire autrement*

458.0 Après que j'ai lu à propos des horreurs qui se prendre place sur les fermes industrielles j'ai immédiatement décidé de couper la viande de mon régime et j'ai commencé à manger des écrous, des haricots et le tofu dans sa place.

458.1 Après que j'ai lu à propos <Coll = des <SProche = horreurs = **cruautés**> qui se prendre place sur = **qui sont commises dans**> les fermes industrielles j'ai immédiatement décidé de <Coll = couper la viande = **éliminer la viande**> de mon régime et j'ai commencé à manger des <SFictif = écrous = **noix/fruits oléagineux/fruits à coque**>, des <Gener = haricots = **légumineuses**> et le tofu <FErr = dans sa place = à la place>.

458.2 Après avoir lu à propos des cruautés qui sont commises dans les fermes industrielles, j'ai immédiatement décidé d'éliminer la viande de mon régime et j'ai commencé à plutôt manger des fruits à coque, des légumineuses et du tofu à la place.

458.3 CRUAUTÉS commises par X envers Y

Oper₁(CRUAUTÉS) = *commettre* [des ~], *infliger* [des ~ à N = Y], *faire subir* [des ~ à N = Y]

Oper₂(CRUAUTÉS) = *être victime* [de ~], *subir* [des ~], *souffrir* [des ~]

RÉGIME de X consistant en Y

Liqu₁Oper₂(RÉGIME) = *éliminer* [N = Y de A_{poss} ~]

ÉCROU : 'bolt' ang.

458.4 Après m'être documentée à propos des cruautés (envers les animaux) commises dans les fermes industrielles, j'ai décidé sur-le-champ d'éliminer la viande de mon alimentation, et j'ai commencé à manger à la place des substituts de viande (fruits oléagineux, légumineuses et tofu).

présente par classes majeures et sous-classes d'erreurs lexicales, le détail du décompte obtenu.

La figure 1 présente la répartition des erreurs lexicales par classes majeures, tandis que la figure 2 présente la répartition des erreurs lexicales par sous-classes.

Analyse quantitative des résultats

Le pourcentage total d'erreurs lexicales standard dans le corpus d'apprenants est de 7.4%. On trouve ainsi en moyenne une erreur lexicale par phrase d'apprenant dans le corpus.

Classes majeures d'erreurs

Les erreurs lexicales de la classe *sens* sont majoritairement dominante dans le corpus d'apprenants, avec un taux de 63%. Viennent ensuite les erreurs

Tableau 3 : Résultats des classes majeures et sous-classes d'erreurs lexicales

Classe d'erreurs	No. d'erreurs	
	Sous-classe	Total
1. Erreurs de sens		2237
(quasi-)Synonyme	676	
sens proche	787	
sens fictif	328	
générique	91	
sens incompatible	204	
situation incompatible	151	
2. Erreurs de forme		431
forme erronée	237	
forme fictive	99	
forme analytique	95	
3. Erreurs de cooccurrence		(912)
collocation	572	
régime ³	(340)	

lexicales de la classe *cooccurrence*, avec un taux de 25%. Les erreurs lexicales appartenant à la classe *forme* sont les moins fréquentes du corpus d'apprenants, avec un taux de 12%.

Sous-classes d'erreurs

Trois sous-classes d'erreurs lexicales sont prédominantes dans le corpus d'apprenants et se situent nettement au-dessus de la moyenne d'erreurs (7.4%) : ce sont d'abord les erreurs de *sens proche* (21%), puis les erreurs de (*quasi*-)*synonyme* (19%) et enfin, les erreurs de *collocation* (16%). Ces trois sous-classes d'erreurs comptent ensemble pour plus de la moitié des erreurs lexicales du corpus (56%). Les deux premières sous-classes d'erreurs appartiennent à la classe d'erreurs lexicales de *sens* et totalisent 40% des erreurs lexicales rencontrées dans le corpus.

Viennent ensuite les erreurs dont le pourcentage est nettement moins élevé par rapport aux trois sous-classes les plus fréquentes d'erreurs. Les erreurs de *sens fictif* (9%) sont les quatrièmes en importance dans le corpus d'apprenants, de même que les erreurs de *régime* (9%). Leur pourcentage est tout de même au-dessus de la moyenne d'erreurs (7.4%). Arrivent en cinquième position les erreurs de *forme erronée* (7%). Finalement, les erreurs restantes, soit celles des sous-classes : *sens incompatible* (6%), *situation incompatible* (4%), *forme fictive* (3%) et *forme analytique* (3%), se situent toutes en dessous de la moyenne d'erreurs lexicales dans le corpus d'apprenants, sans être pour autant résiduelles.

Répartition des erreurs lexicales par classes majeures

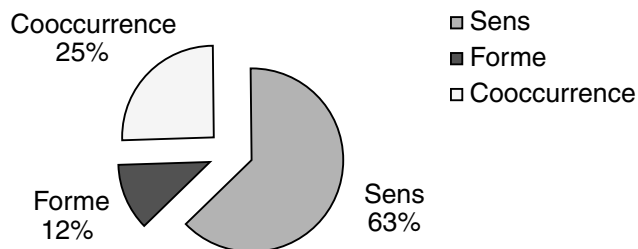


Figure 1 : Répartition des erreurs lexicales par classes majeures dans le corpus *Dire autrement*

Répartition des erreurs lexicales par sous-classes

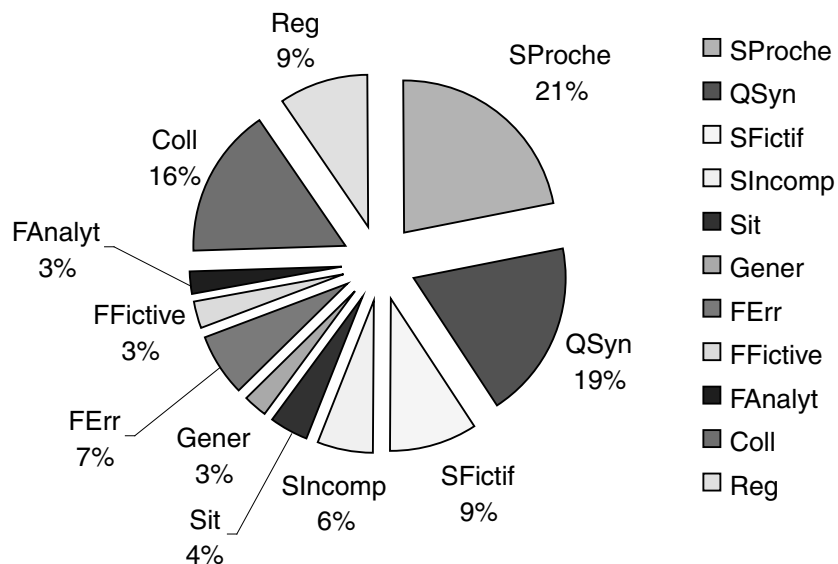


Figure 2 : Répartition des erreurs lexicales par sous-classes dans le corpus *Dire autrement*

Interprétation qualitative des résultats

Notre analyse détaillée des sous-classes d’erreurs lexicales dans le corpus *Dire autrement* nous permet de relever des problèmes importants chez l’apprenant, notamment en ce qui concerne la sélection du mot juste parmi des lexies de

sens proche ou des lexies (*quasi*-)synonymes et celle du collocatif d'une base dans une *collocation*.

On note de manière spécifique des lacunes chez l'apprenant quant à ses connaissances lexicales portant sur des groupes de lexies en particulier, des lexies qui se trouvent en fait appartenir à des champs sémantiques précis. Ces sont par exemples les lexies du champ 'espace', qui comprennent notamment les lexies suivantes : 'endroit', 'lieu', 'place', 'destination', etc. C'est un champ clairement mal maîtrisé par l'apprenant qui choisit souvent à l'intérieur de celui-ci une lexie (*quasi*-)synonyme ou encore de *sens proche*, laquelle se révèle inappropriée au contexte qu'il cherche à décrire. C'est aussi le cas des lexies du champ 'personne' qui comprennent notamment les lexies suivantes : 'gens', 'individu', 'population', 'on', etc. Il en va de même pour les lexies du champ 'communication', avec des lexies telles que : 'conseil', 'instruction', 'information', 'renseignement', etc.

On relève aussi des problèmes ayant trait à certains types de *collocation* :

- celles dont la base est en position objet — dans ce contexte, le collocatif précède linéairement la base malgré que celle-ci en contraint le choix (*mettre* de l'importance → *accorder* de l'importance).
- celles dont le collocatif visé est peu transparent c'est-à-dire plus idiomatique donc moins prévisible (*devenir* enceinte → *tomber* enceinte).
- celles dont le collocatif est plus transparent mais reste tout de même moins prévisible car la lexie en question est moins récurrente en français (une avalanche *a lieu* → une avalanche *se déclenche*).

Le fait que l'on trouve un taux plus élevé d'erreurs lexicales de *sens proche* et (*quasi*)synonyme, donc d'erreurs où l'écart sémantique entre la lexie erronée et la lexie visée est moins grand (par rapport à celui qui existe entre la lexie erronée de *sens fictif* et la lexie visée, par exemple) indique chez l'apprenant la présence d'un certain bagage de connaissances lexicales. Lorsque qu'il crée une lexie de *sens fictif*, dans plusieurs cas il emprunte la première traduction/le premier sens fourni par son dictionnaire bilingue pour la lexie qu'il cherche à traduire (*nuts*₁ → *écrous* Vs *nuts*₂ → noix). Il fait très peu de véritables créations lexicales. Ses erreurs de *forme fictive* s'avèrent souvent des emprunts (quasi-)direct de l'anglais (*concernes* → problèmes). Ses erreurs de *forme erronée* sont causées par une orthographe déficiente, mais aussi par une mauvaise dérivation morphologique (*construiseur* → constructeur). L'apprenant choisit souvent par défaut un (semi-)auxiliaire — soit une lexie plus productive, de haute fréquence — comme verbe support dans une *collocation* (*avoir* → ressentir de la fatigue), créant ainsi des tournures maladroites. Il utilise de temps en temps une forme analytique (votre *nouvel enfant* → nouveau-né) pour contourner l'obstacle du mot/de l'expression inconnu/e. Il a tendance à répéter les mêmes mots et parfois à se tromper dans sa reprise

lexicale — au niveau par exemple de la sélection du *générique* (les *individus* → gens). En somme se dessine le profil d'un apprenant qui possède bel et bien des connaissances lexicales, mais pour lequel on constate des lacunes importantes à combler.

Vers un dictionnaire d'apprentissage du FLS basé sur la TST

Le développement de connaissances lexicales qui mènent à la compétence lexicale est un objectif reconnu de l'enseignement de la langue (Tréville, 2000 ; Nation, 2001 ; Laniel, 2005). Nous constatons toutefois que les ressources/moyens qui permettent de l'atteindre en français langue seconde sont déficients. En effet, il existe peu de ressources, entre autres électroniques, facilitant la production textuelle (notamment la révision et l'autocorrection par le biais de la reformulation en L2) et ciblant particulièrement les apprenants de niveau intermédiaire-avancé en français langue seconde. C'est dans cette perspective — et nous l'avons annoncé plus haut — que nous comptons développer, en partie sur la base de notre travail d'analyse d'erreurs lexicales, un dictionnaire d'apprentissage visant, par le biais de la reformulation, l'amélioration de la compétence lexicale de production de cette clientèle d'apprenants. Ce dictionnaire, nous l'avons dit, sera conçu dans le cadre de la théorie Sens-Texte (TST). Des dictionnaires existent qui ont été élaborés dans ce cadre — le *DEC* (Mel'čuk, 1984–1999), le *DiCo* (Polguère, 2000). Il s'agit essentiellement d'outils de recherche, peu conviviaux. Or, notre visée est celle d'un dictionnaire d'apprentissage, qui s'inscrira dans la lignée du *Longman Activator* ou encore du *Dictionnaire du français usuel* (Picoche et Rolland, 2001). On trouve aussi, en TST, des initiatives de développement de dictionnaires pédagogiques telles le *LAF* (Mel'čuk et Polguère, 2007), le *FrLEX* (Hamel, Nkwenti-Azeh et Zahner, 1996), le *DAFLES* (Selva, 1999 ; Selva, Verlinde et Binon, 2002), le *Diccionario de collocationes españolas* (Alonso-Ramos, 2004), mais dont l'approche, la clientèle, le contenu et/ou le format diffèrent de ce que nous envisageons. Notre dictionnaire d'apprentissage basé sur la TST prendra compte de difficultés attestées d'apprenants du FLS et c'est ce qui le distinguera, entre autres, des initiatives louables susmentionnées. Sa nomenclature sera organisée en champs sémantiques et les lexies qui forment ces champs sont en partie des lexies relevées du corpus d'apprenants (avec leurs voisines dans des champs sémantiques correspondants). Nous avons sélectionné une quinzaine de champs sémantiques à travers lesquels se répartissent environ 1000 lexies, lesquelles seront décrites dans un format lexicographique convivial, adapté de celui proposé par la TST. Les lexies seront rangées au fur et à mesure dans une base de données qui pourra être interrogée en ligne au moyen de divers types de requêtes. Une interface d'ALAO permettra des accès multiples et variés vers les données lexicales de la base. Elle comprendra

aussi un module didactique axé sur les relations lexicales paradigmatiques et syntagmatiques où les données lexicales de la base seront exploitées dans des exercices, notamment de reformulation (de paraphrasage).

Conclusion

Notre analyse d'erreurs lexicales d'un corpus d'apprenants du FLS nous a permis de mesurer certains aspects de la profondeur (soit la connaissance des relations sémantiques et syntagmatiques) de la compétence lexicale de ces apprenants, et ce en contexte de production. Il est entendu qu'il existe d'autres mesures de la compétence lexicale des apprenants, des mesures portant sur l'étendue et/ou la profondeur des connaissances lexicales, en contexte de compréhension ou de production. À titre d'exemples : les *Euralex French Tests* (Bogaards, 2000 : mesure de la profondeur/contexte de compréhension) ou encore le *Profil de Fréquence Lexicale* (Cobb et Horst, 2004 : mesure de la répartition de l'étendue/en contexte de production). Ces mesures sont complémentaires à l'analyse d'erreurs lexicales dans les corpus d'apprenants. Elles nous intéressent dans la perspective où nous comptons poursuivre notre travail de description de la compétence lexicale des apprenants du FLS de niveau intermédiaire-avancé.

Avant tout, notre analyse d'erreurs lexicales d'un corpus d'apprenants du FLS s'est montrée profitable en ce qu'elle a pu nous fournir des données pertinentes à l'élaboration d'un dictionnaire d'apprentissage du FLS axé sur les relations lexicales et de paraphrase. Cette démarche empirique, en amont dans le processus d'élaboration d'une telle ressource, nous a ainsi permis de faire non seulement une sélection mieux éclairée du stock lexical à décrire, mais aussi de définir le format du dictionnaire (sa macro- et sa microstructure). En aval, les données nous serviront sans doute aussi à l'élaboration du module didactique, qui comportera des exercices ciblant certaines difficultés relevées dans le corpus analysé. La ressource lexicale résultante ne sera qu'un fragment de ce qu'elle devrait être, notre objectif principal étant de démontrer la validité, notamment de notre démarche empirique, mais encore et surtout de celle du cadre théorique sur lequel s'appuie toute notre recherche.

Notes

Nous remercions chaleureusement nos assistants de recherche, et en particulier Yves Bourque (pour les statistiques et les figures) et Muriel Péguret (pour la relecture attentive de ce manuscrit).

¹ Le projet *Dire Autrement* (direautrement.french.dal.ca/) bénéficie d'une subvention du CSRH (410-2005-0177).

² Le terme *paraphrase* est utilisé ici dans le sens de 'phrase explicative', de périphrase : un sens différent de celui que nous lui conférons.

- ³ Parce qu'elles sont moins intéressantes du point de vue de la description lexicale et de la nature de l'outil envisagé, les erreurs de type *régime* n'ont pas été systématiquement identifiées dans l'ensemble du corpus et demeurent donc comptabilisées de façon non formelle à ce jour. Nous en avons relevé 340.
- ⁴ Ce sont uniquement les erreurs qui se répartissent dans nos 3 classes majeures d'erreurs lexicales. On compte en fait un total de 3810 erreurs lexicales dans l'ensemble du corpus. Une proportion de ces erreurs correspond à des erreurs non standard par rapport à notre typologie. Ce sont principalement des erreurs d'omission de mot ou de mot superflu (qui ne sont pas des erreurs de *régime*).

Références

- Alonso Ramos, M. 2004. Elaboración del Diccionario de colocaciones en español y sus aplicaciones. Dans P. Bataner et J. de Cesaris (dir.), *De Lexicographia. Actes del I symposium internacional de lexicografía*, Barcelone, Edicions Petició, pp. 149–162.
- Anctil, D. 2005. Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et analyse des stratégies de résolution de problèmes lexicaux. Mémoire de maîtrise, Université de Montréal.
- Bogaards, P. 2000. Testing L2 vocabulary knowledge at a high level : The case of the Euralex French Tests. *Applied Linguistics*, vol. 21, n^o4, pp. 490–516.
- Conseil de l'Europe. 2000. *Cadre européen commun de référence pour les langues*. Disponible à : www.coe.int/t/dg4/linguistic/CADRE_FR.asp
- Gillard, P. et A. Gadsby. 1998. Using a learners' corpus in compiling ELT dictionaries. Dans S. Granger (dir.), *Learner English on Computer*. London, Longman, pp. 159–171.
- Granger, S. (dir.). 1998. *Learner English on computer*. London, Longman.
- Granger, S. 2003. Error-tagged learner corpora and CALL : A promising synergy. *CALLICO*, vol. 20, n^o3, pp. 465–480.
- Granger, S., A. Vandeventer et M.-J. Hamel. 2001. Analyse de corpus d'apprenants pour l'ELAO basé sur le TAL. *Traitement Automatique de la Langue*, vol. 42, n^o2, pp. 609–621.
- Hamel, M.-J. 2005. Grammaire de texte en contexte d'ALAO : une expérience avec FreeText. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, vol. 62, n^o1, pp. 221–233.
- Hamel, M.-J., B. Nkwenti-Azeh et C. Zahner. 1996. The conceptual dictionary in CALL. Dans *Selected Proceedings of the EuroCALL95 Conference*. Valencia University, Espagne, pp. 509–518.
- James, C. 1998. *Errors in language and use*. Don Mills, Ontario, Longman.
- Laniel, D. 2005. Le vocabulaire en français langue seconde : de 'parent pauvre' de l'enseignement à 'invité d'honneur' des communications assistées par ordinateur. *Revue de l'Association Québécoise des Enseignants de Français Langue Seconde*, vol. 25, n^o2, pp. 73–100.
- Laurier, M. (dir.). 2000. *Niveaux de compétence en français langue seconde pour les immigrants adultes*. Ministère des Relations avec les citoyens et de l'Immigration, Gouvernement du Québec.

- Mel'čuk, I. 1992. Paraphrase et lexique : la théorie Sens-Texte et le « Dictionnaire explicatif et combinatoire ». Dans Mel'čuk (dir.), vol. III, pp. 9–59.
- Mel'čuk, I. (dir.). 1984, 1988, 1992, 1999. *Le dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I–IV*. Avec la collaboration de Nadia Arbatchewsky-Jumarie, Lidija Iordanskaja et Adèle Lessard. Montréal, Presses de l'Université de Montréal.
- Mel'čuk, I., A. Clas et A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve, Duculot.
- Mel'čuk, I. et A. Polguère. 2007. Lexique actif du français L'apprentissage du vocabulaire fondé sur 20 000 dérivation sémantiques du français. Louvain-la-Neuve, deBoek Duculot.
- Milićević, J. 2007. *La paraphrase. Modélisation de la paraphrase langagière*. Berne, Peter Lang.
- Milićević, J. et M.-J. Hamel. À paraître. Un dictionnaire de reformulation pour les apprenants du français langue seconde. Dans *Actes du 29^e colloque de l'Association linguistique des Provinces de l'Atlantique*, Université de Moncton, 2005.
- Nation, I.S.P. 2001. *Learning vocabulary in another language*. Cambridge, Cambridge University Press.
- Nesselhauf, N. 2005. *Collocations in a learner corpus*. Amsterdam, John Benjamins.
- Picoche, J. et J.C. Rolland. 2001. *Dictionnaire du français usuel*. Paris, Duculot.
- Polguère, A. 2000. Une base de données lexicales du français et ses applications possibles en didactique. *Revue de Linguistique et de Didactique des Langues (LIDIL)*, n°21, pp. 75–97.
- 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, France.
- Selva T., S. Verlinde et J. Binon. 2002. Le DAFLES, un nouveau dictionnaire pour apprenants du français. Dans *Actes du dixième congrès EURALEX 2002*, Copenhague, pp. 199–208.
- 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*. Philadelphia, Benjamins.

