

Le choix lexical en français langue seconde :
fondements empiriques et réflexions didactiques

Valérie Flageul

Mémoire

présenté

à

l'École des études supérieures

comme exigence partielle au grade de
maîtrise ès Arts (Programme spécial individualisé)

Université Concordia

Montréal, Québec, Canada

Novembre 2012

© Valérie Flageul, 2012

UNIVERSITÉ CONCORDIA

École des études supérieures

Nous certifions par les présentes que le mémoire rédigé

par Valérie Flageul

intitulé Le choix lexical en français langue seconde : fondements empiriques et réflexions didactiques

et déposé à titre d'exigence partielle en vue de l'obtention du grade de

Maîtrise ès Arts (Programme spécial individualisé)

est conforme aux règlements de l'Université et satisfait aux normes établies pour ce qui est de l'originalité et de la qualité.

Signé par les membres du Comité de soutenance

__Bradley Nelson_____ président

__Djaouida Hamdani Kadri _____ examinatrice

__Denis Liakin_____ examinateur

__Paula Bouffard_____ directrice

Approuvé par __Bradley Nelson_____ Directeur du programme d'études supérieures

22 novembre 2012 __Paula Wood-Adams_____ Doyenne, École des études supérieures

RÉSUMÉ

Le choix lexical en français langue seconde : fondements empiriques et réflexions didactiques

Valérie Flageul

Ce mémoire propose des pistes pour l'enseignement du choix lexical en français langue seconde au niveau intermédiaire-avancé fondées sur une analyse de corpus d'apprenants. L'analyse des erreurs lexicales dans un corpus de productions écrites a permis de dégager des catégories d'erreurs de choix lexical. Cinq catégories d'erreurs de nature sémantique sont caractérisées à l'aide de notions linguistiques, pour être utilisées dans une séquence didactique. Cette séquence comprend la présentation d'un « espace de décision » pour le choix lexical associée à une prise de conscience des paramètres intervenant dans le choix lexical, une démarche d'autocorrection avec utilisation des codes issus de l'analyse du corpus et de stratégies d'autocorrection et enfin un travail sur des champs lexicaux. En effet, la compétence dans le choix lexical dépend à la fois des mots considérés individuellement et des mots insérés dans un réseau lexical.

REMERCIEMENTS

Je remercie du fond du cœur ma directrice, Paula Bouffard, qui a rendu ce projet possible en veillant à l'élaboration du projet SIP et en m'assurant en tout temps de son soutien amical, scientifique et financier (bourses et emplois). Ses suggestions importantes d'utiliser un corpus (en plus de mettre à ma disposition le corpus de son cours de français écrit) et d'utiliser les résultats de l'analyse du corpus comme étiquettes de codes de correction ont grandement contribué à l'orientation de mon travail. Ses commentaires constructifs et nos échanges fructueux m'ont également permis d'améliorer grandement le document.

Je tiens également à remercier Djaouida Hamdani Kadri et Denis Liakin qui ont eu la gentillesse de participer à l'élaboration du projet SIP et de m'avoir fait part de leurs commentaires constructifs sur le projet.

Je remercie Davy Bigot, Ollivier Dyens, Adel Jebali, Svetla Kamenova, Denis Liakin et Maria Trigueiro qui ont contribué à ma survie financière en m'offrant de nombreux contrats de travail au Département d'études françaises pendant mes études de maîtrise.

Mes remerciements vont également à Darlene Dubiel, coordinatrice du programme SIP/INDI, pour son soutien administratif et sa gentillesse.

Je tiens finalement à remercier chaleureusement mes amis et ma famille pour leurs encouragements constants, et tout particulièrement Claudia Bedini, Ginette Dion, Helen El Ketrani, Bénédicte, Marie-Françoise et Maurice Flageul, Maribeth Johnson, Lionel Langlais, Maryam Miraghaie, Michael Ross, Magdalena Shishenkova, Maureen Wahba et Mountaz Zizi.

TABLE DES MATIÈRES

LISTE DES FIGURES	viii
LISTE DES TABLEAUX.....	ix
INTRODUCTION.....	1
1. LE CHOIX LEXICAL EN LANGUE SECONDE : DÉFINITION ET PROBLÉMATIQUE .6	
1.1. L'accès lexical en production.....	7
1.2. Le lexique et les connaissances lexicales	10
1.2.1. Les connaissances lexicales selon Richards (1976).....	11
1.2.2. Les connaissances lexicales selon Nation (2001).....	13
1.3. Le choix lexical chez les apprenants	17
1.3.1. Connaissance lexicale incomplète des apprenants.....	17
1.3.2. Connaissance lexicale imparfaite des apprenants	20
1.4. Le développement du sens lexical en langue seconde.....	24
1.4.1. Le développement des relations concept-mot.....	25
1.4.2. Le développement des réseaux lexicaux	29
1.5. Synthèse.....	32
2. ÉTUDE DU CORPUS : FONDEMENTS EMPIRIQUES ET LINGUISTIQUES.....	35
2.1. Corpus, contexte pédagogique et démarche d'analyse	35
2.2. Présentation synthétique de la classification du corpus.....	40
2.2.1. Erreurs formelles	40
2.2.2. Maladresses stylistiques.....	43
2.2.3. Erreurs pragmatiques	45
2.2.4. Erreurs sémantiques	47

2.2.5. Récapitulatif.....	49
2.3. Analyse des classes sémantiques du corpus	51
2.3.1. CONSTRUCTION INCOMPATIBLE : polysémie et sous-catégorisation	51
2.3.2. CONCEPT INADÉQUAT : concept, composante de sens, champ sémantique et synonymie	54
2.3.3. COLLOCATION ATTENDUE : collocations	56
2.3.4. RELATION INCOMPATIBLE : restrictions de sélection	60
2.3.5. AFFIXE INADÉQUAT : famille de mots	62
2.3.6. Récapitulatif des notions et liens entre les classes	64
3. VERS UNE MÉTHODOLOGIE DE L'ENSEIGNEMENT DU CHOIX LEXICAL.....	66
3.1. Présentation générale de la séquence didactique	67
3.1.1. Pour un apprentissage explicite et en contexte du sens lexical.....	67
3.1.2. Pour le développement de réseaux d'association	71
3.1.3. Les trois étapes de la séquence didactique pour le choix lexical.....	72
3.2. Faire voir l'« espace de décision »	73
3.2.1. Présentation	73
3.2.2. Activités pour l'espace de décision	75
3.3. Autocorrection : mode d'emploi	80
3.3.1. Présentation	80
3.3.2. Activités pour l'espace de résolution de problème.....	84
3.4. Appréhension de la polysémie et construction de réseaux	87
3.4.1. Présentation	87
3.4.2. Sélection du vocabulaire et activités	95

CONCLUSION ET PERSPECTIVES	99
RÉFÉRENCES BIBLIOGRAPHIQUES	103
ANNEXE 1 : Exemples tirés du corpus représentant quelques classes	108
ANNEXE 2 : Exemples de champs sémantiques issus du corpus	116

LISTE DES FIGURES

Figure 1 – Aperçu du modèle de production de Levelt (1989)	8
--	---

LISTE DES TABLEAUX

Tableau 1 : Connaître un mot, d'après Nation (2001) et Richards (1976)	14
Tableau 2 : Catégories d'erreurs de choix lexical	49-50
Tableau 3: Erreurs résultant de la méconnaissance des collocations	58-59
Tableau 4 : Présentation des codes : définitions et exemples	84-85
Tableau 5 : Répartition des champs sémantiques dans différents types de textes	96

INTRODUCTION

Il est maintenant généralement reconnu, à la fois par les linguistes et les didacticiens, que les connaissances lexicales jouent un rôle central dans l'usage d'une langue, maternelle ou seconde (par ex. Levelt, 1989 ; Willis, 1990 ; Tréville et Duquette, 1996 ; Singleton, 1999, 2000 ; Lewis, 2000). En effet, ce qui distingue les apprenants de langue seconde de niveau avancé de ceux de niveau intermédiaire n'est pas tant leur maîtrise grammaticale que l'étendue et la précision de leur vocabulaire (Lewis, 2000). Au niveau intermédiaire-avancé (B2 du Cadre européen commun de référence pour les langues), qui est le seuil d'apprentissage ciblé dans ce mémoire, la compétence lexicale devrait donc faire l'objet d'une attention particulière.

Mais comment définir la compétence lexicale, si tant est qu'elle soit « séparable » de la compétence langagière générale ? Qu'est-ce qu'avoir un « bon » vocabulaire dans une langue étrangère ? Est-ce connaître beaucoup de mots (quantité) ? Est-ce connaître les propriétés des mots en profondeur (qualité) ? Par ailleurs, est-ce reconnaître un mot, en comprendre le sens (réception) ? Est-ce être capable de trouver rapidement le mot correspondant à l'idée que l'on veut transmettre (production, aisance à communiquer) ? Est-ce savoir les utiliser correctement (usage) ? Ou est-ce tout cela à la fois ?

La question de la compétence lexicale peut en premier lieu être examinée sous l'angle de l'étendue du vocabulaire. Si l'on pense d'une manière naïve, ou intuitive, à ce qu'est « avoir » du vocabulaire, c'est souvent la quantité de mots connus qui vient à l'esprit. Il

est en effet indispensable de posséder un certain vocabulaire pour fonctionner adéquatement en langue seconde. Par exemple, Laufer (1989) montre qu'une compréhension raisonnable (c'est-à-dire correspondant à la note minimale de réussite à l'Université Haifa) d'un texte de nature universitaire en anglais langue seconde n'est le plus souvent assurée que lorsque l'apprenant dispose d'une connaissance d'au moins 95 % des mots du texte. Aussi, Hu Hsueh-chao et Nation (2000) mettent en évidence un niveau de connaissance de 98 % des mots d'un texte de fiction (ce qui correspond à un mot inconnu tous les 50 mots) pour la compréhension dans le cadre d'une lecture sans assistance. En-deçà de ces niveaux, la compréhension est généralement partielle, peu probable, voire nulle. Par conséquent, des études sont menées pour chercher à connaître la taille du vocabulaire de l'apprenant qui amènerait une compréhension suffisante ou bien qui correspondrait à une bonne couverture lexicale du texte. Ainsi Laufer (1992) met à jour un niveau seuil de vocabulaire passif de l'apprenant (nombre de mots reconnus et compris) pour une compréhension, raisonnable toujours, de textes de nature universitaire en anglais langue seconde. Il s'agit de la connaissance des 3000 familles de mots les plus fréquentes de l'anglais, une famille de mots étant l'ensemble des formes ayant une racine commune (par ex. *drive* et *driver*).

Comme l'étendue du vocabulaire est un enjeu considérable pour le développement de la compétence langagière, et en particulier de la compréhension, chez les apprenants de langue seconde, de nombreuses études portent sur l'évaluation de la taille du vocabulaire et sur les activités ou méthodes permettant de l'augmenter (cf. Nation (2001) pour une synthèse). Toutefois, de nombreux auteurs s'accordent sur le fait que ce

n'est pas la seule dimension importante, et qu'il est également important de s'intéresser à la qualité, ou la profondeur, de la connaissance du vocabulaire (Richards, 1976 ; Qian, 1999 ; Nation, 2001). La qualité de la connaissance du vocabulaire se définit par le degré de finesse avec lequel l'apprenant connaît les mots, qui correspond à différentes propriétés de ceux-ci, formelles, sémantiques et d'usage. Ces propriétés sont particulièrement importantes en production, comme l'a montré Nicolae (2003), en croisant des tests de connaissance lexicale (quantitatifs et qualitatifs) des verbes en anglais langue seconde avec des tests de reconnaissance syntaxique. Chez des apprenants ayant un niveau de connaissance lexicale intermédiaire-avancé, il est apparu que la connaissance du sens d'un verbe ne garantissait pas la capacité à produire une phrase grammaticalement correcte avec ce verbe. Par exemple, l'analyse de la phrase « *She has cared all animals since she was a child* » indique une certaine connaissance du sens du verbe *care* alors que son utilisation dans la phrase n'est pas correcte du point de vue syntaxique. Ainsi, s'il est important de posséder un vocabulaire étendu et diversifié, il est tout aussi important, en phase de production notamment, d'avoir une connaissance fine des sens des mots, de leur forme et des conditions de leur emploi.

Cependant, la notion de compétence renvoie à celle d'habileté et non à celle d'ensemble de connaissances. La compétence lexicale peut alors être définie comme la capacité à comprendre un mot dans son contexte ou à trouver le mot à utiliser en fonction du contexte et à former la phrase correspondante. Elle repose sur des connaissances

lexicales, quantitatives (nombre de mots connus) et qualitatives (connaissance des différentes dimensions du mot).

En production orale ou écrite, l'apprenant de langue seconde doit utiliser les mots qui expriment au mieux sa pensée et qui, par ailleurs, conviennent au contexte phrastique dans lequel ils s'insèrent. Des erreurs dans le choix d'un mot peuvent créer des problèmes de communication et sont parfois perçues négativement par des locuteurs natifs ou des enseignants (par ex. Salem, 2007 ; Hemchua et Schmitt, 2006), plus négativement que des erreurs grammaticales. Voici des exemples d'énoncés tirés d'un corpus de productions écrites d'apprenants, dans lesquels certains choix lexicaux sont inappropriés.

- (1) *Le moment indésirable **s'approximait** plus vite chaque jour.* (→ approchait)
- (2) *Subitement, René a **débuté** de bouger.* (→ commencé)
- (3) *Un médaillon en forme de cœur (...) L'inscription sur le **postérieur** montra les mots "Pour Sophie".* (→ envers)
- (4) *En **prenant** le diplôme d'un médecin psychologique neurologue en Égypte (...).*
(→ obtenant)
- (5) *Le voyage, les pays inconnus et les langues étrangères **provoquent** son imagination et sa curiosité.* (→ stimulent son imagination et piquent sa curiosité)
- (6) *Cette **madame** accommodante, qui évoque la tendresse d'une mère (...).* (→ femme)

Comme le laissent paraître ces exemples, ces apprenants disposent déjà d'importantes connaissances lexicales, mais qui sont toutefois incomplètes : certains synonymes sont

utilisés dans un contexte qui ne convient pas ((1) à (3)), certaines collocations ne semblent pas connues ((4) et (5)), le registre n'est pas toujours adéquat ((6)).

Nous nous intéresserons, dans ce mémoire, à la question du choix lexical en français langue seconde chez des apprenants de niveau intermédiaire-avancé. Notre objectif est double. Il s'agit de (i) bien cerner le phénomène du choix lexical afin de (ii) proposer des pistes de nature didactique qui permettront aux apprenants de langue seconde de niveau intermédiaire-avancé d'améliorer leur compétence dans le choix lexical. Pour l'enseignement du vocabulaire en général, et du choix lexical en particulier, on se trouve devant une difficulté importante. D'un côté, les connaissances à acquérir sont infinies, de par la quantité de mots existants et le nombre de connaissances associées à chaque mot. De l'autre, les moyens disponibles sont limités (temps de classe, outils lexicaux, temps disponible pour la lecture...). Dans ce contexte, comment enseigner le choix lexical ? Quels sont les moyens à mettre en œuvre pour développer cette compétence chez les apprenants ? Quels mots ou ensembles de mots faut-il enseigner ?

Afin de mieux appréhender la problématique du choix lexical (section 1.), nous allons caractériser la notion de choix lexical, essayer de comprendre quelles sont les nombreuses connaissances lexicales qui sous-tendent la compétence lexicale, quels types d'erreurs les apprenants commettent réellement et dans quelles proportions, et si leur production est comparable ou non à celle de locuteurs natifs. Comme il se confirme que le choix lexical est source de difficulté pour les apprenants, nous présenterons des hypothèses sur les raisons de ces difficultés. Afin d'être en mesure de proposer des

pistes de méthodologie pour l'enseignement du choix lexical, nous allons travailler à partir d'un corpus de copies d'apprenants (section 2.). L'étude de ce corpus nous permettra de dégager des catégories d'erreurs, en particulier d'erreurs de nature sémantique que nous analyserons à l'aide de notions linguistiques. À partir de l'analyse de travaux en didactique et en acquisition et des résultats de l'analyse de corpus, nous élaborerons une séquence didactique pour l'enseignement du choix lexical (section 3.). En utilisant différents moyens, nous tenterons d'apporter des éléments de réponse à la question suivante : que peut-on faire en salle de classe pour que l'apprenant puisse accroître ses compétences dans le choix lexical ?

1. LE CHOIX LEXICAL EN LANGUE SECONDE : DÉFINITION ET PROBLÉMATIQUE

Le choix lexical s'opère en production, orale ou écrite. La production consiste à passer d'une intention de communication à l'articulation (ou l'écriture) d'un énoncé. Au cours de ce processus, en langue seconde, les apprenants doivent sélectionner, dans leur L2, les mots aptes à transmettre le message qu'ils souhaitent communiquer. Les mots qu'ils vont être amenés à produire font partie de leur vocabulaire actif, par opposition au vocabulaire passif qui est associé à la compréhension. Le vocabulaire actif est généralement plus limité, quantitativement, que le vocabulaire passif. La possibilité de sélectionner un mot sous-entend que l'apprenant possède un vocabulaire en quantité suffisante, qu'il soit capable de différencier des mots faisant partie d'un même champ sémantique par exemple, et cela avec une connaissance des mots suffisamment fine.

Dans cette section, nous allons caractériser la notion de choix lexical en la replaçant dans le contexte de la production d'énoncés (1.1.). Nous allons ensuite présenter les connaissances lexicales sur lesquelles reposent la production de mots et le choix lexical (1.2.). Puis nous allons identifier les difficultés lexicales rencontrées par les apprenants lors de la production de textes (1.3.). Enfin nous tenterons de donner des explications de ces difficultés (1.4.). Nous serons alors en mesure de présenter notre problématique en lien avec le choix lexical (1.5.).

1.1. L'accès lexical en production

Le choix lexical peut être défini comme le résultat de l'accès lexical en production. Si la production de la parole peut se résumer à une série de traitements successifs, l'accès au lexique, qui se traduit entre autres par la récupération d'un mot, a lieu au cours de l'un de ces traitements.

Pour la langue maternelle, Levelt (1989) propose un modèle descriptif de l'ensemble des traitements, ainsi que des connaissances déclaratives utilisées, qui mènent à la production d'un énoncé. Ces différents traitements, ainsi que les connaissances déclaratives associées, sont représentés dans la figure 1, adaptée de Levelt (1989).

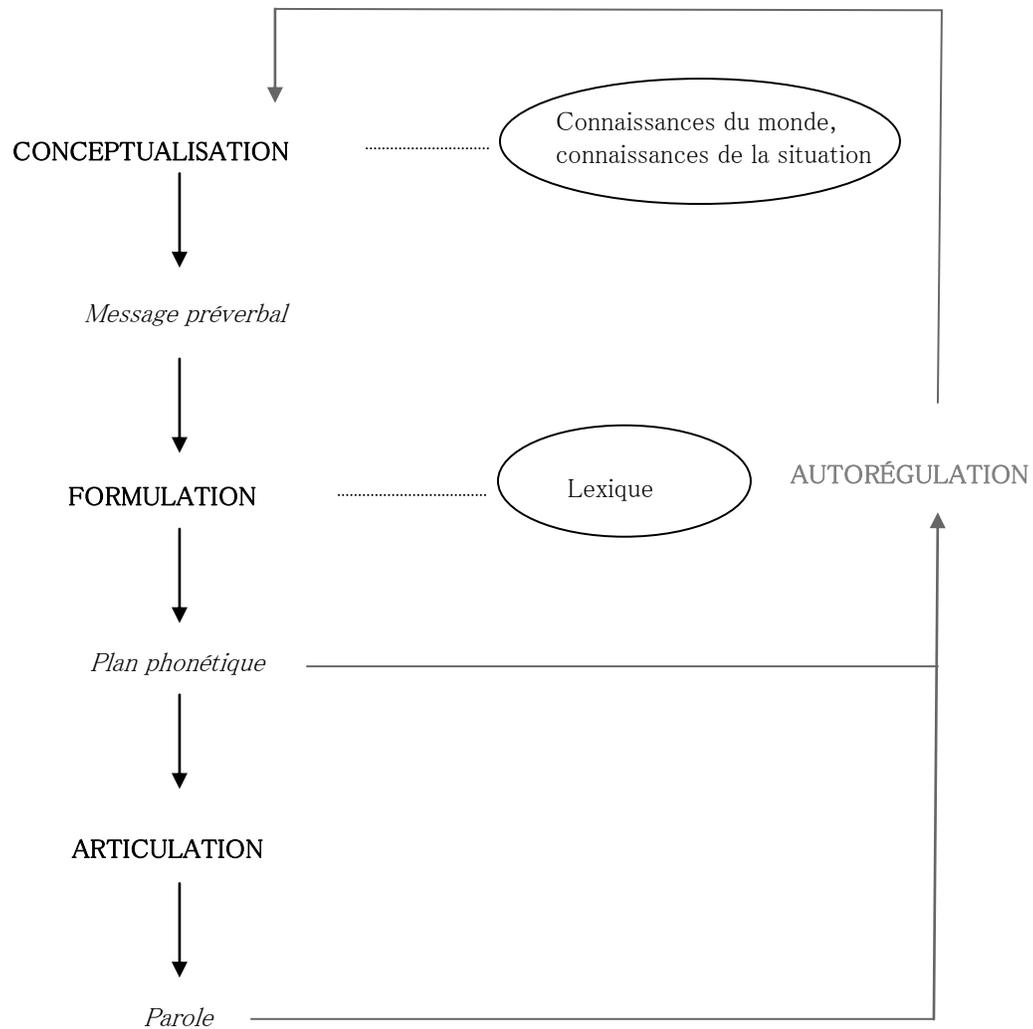


Figure 1 – Aperçu du modèle de production de Levelt (1989)¹

Dans la phase de conceptualisation, le locuteur part d'une intention de communication pour aboutir à un message « préverbal » en se basant sur des connaissances générales et des connaissances de la situation de communication. Notons que pour les locuteurs

¹ Ce schéma est une simplification de la figure « A blueprint for the speaker » dans Levelt (1989), p. 9.

bilingues, De Bot (1992), dans son adaptation du modèle², a posé l'hypothèse que l'une des étapes du processus de conceptualisation est liée à une langue donnée, puisque les associations concepts-mots ne sont pas identiques dans toutes les langues.

Dans la phase de formulation, le locuteur accède au lexique. Dans le modèle de Levelt (1989), le lexique est représenté comme un ensemble de connaissances déclaratives. Chaque unité lexicale contient quatre types d'informations : (i) le sens, qui consiste en une représentation conceptuelle, (ii) les propriétés syntaxiques du mot, c'est-à-dire principalement sa catégorie grammaticale et ses arguments syntaxiques, (iii) des spécifications morphologiques et (iv) des spécifications phonologiques. Des unités lexicales sont associées aux concepts exprimés dans le message préverbal. De plus, à l'aide des informations contenues dans les unités lexicales, le message est encodé grammaticalement, morphologiquement et phonologiquement, pour aboutir à un « plan phonétique ».

Dans la phase d'articulation, les mécanismes articulatoires sont mis en œuvre afin que le plan phonétique puisse donner lieu à la parole. L'autorégulation, correspondant à l'audition et à la compréhension du discours, s'effectue en parallèle.

Ainsi, c'est au cours du processus de formulation que l'accès aux connaissances lexicales prend place. Tout d'abord, une unité lexicale est sélectionnée par l'intermédiaire de son sens, c'est-à-dire de sa représentation conceptuelle. Plus

2 De Bot (1992) a proposé une adaptation du modèle qui puisse rendre compte de la compétence des locuteurs bilingues, quel que soit le niveau de maîtrise de la langue seconde parlée par ceux-ci.

précisément, c'est la présence des concepts du message préverbal dans la représentation sémantique du mot qui déclenche la sélection de celui-ci³. Ensuite, les informations syntaxiques contenues dans l'unité lexicale sélectionnée vont déclencher des procédures d'encodage syntaxique. Enfin, l'encodage phonologique fait appel aux connaissances morphologiques et phonologiques du mot. Le lexique occupe ainsi une place centrale dans ce modèle, puisque, pour la production d'énoncés, l'encodage grammatical en particulier se base sur les connaissances lexicales.

1.2. Le lexique et les connaissances lexicales

En production (expression orale ou écrite), la qualité de la connaissance lexicale revêt une importance particulière, car l'apprenant doit choisir le mot qui correspond au sens qu'il souhaite véhiculer dans un certain contexte phrastique et communicatif, tout en utilisant la forme choisie d'une manière correcte syntaxiquement. Cela suppose qu'il ait une connaissance approfondie du vocabulaire, qui ne se limite pas à la connaissance d'une association forme-sens. Dans cette section, nous présenterons deux modèles (Richards, 1976 ; Nation, 2001) qui articulent les différents types de connaissances impliquées dans la maîtrise de l'usage d'un mot.

Richards (1976) puis Nation (2001) se sont demandé ce que signifie précisément connaître un mot. Ils ont dressé la liste de l'ensemble des connaissances impliquées

3 Toutefois, comme l'a fait remarquer Levelt (1992, citant Ward (1991)), les collocations peuvent également imposer des choix lexicaux, qui dépendent d'un mot se trouvant dans le contexte, et pas uniquement de la représentation conceptuelle.

dans la maîtrise de l'usage d'un mot, que l'on peut regrouper sous le terme de qualité, ou profondeur, de la connaissance lexicale.

1.2.1. Les connaissances lexicales selon Richards (1976)

Richards (1976) a défini un ensemble minimal de connaissances, qui sous-tendent la compétence lexicale, dans le but d'identifier plus précisément des objectifs pédagogiques pour l'enseignement du vocabulaire. Cet ensemble consiste en sept types de connaissances, dont nous reprenons la définition ci-dessous.

- La connaissance de la probabilité d'occurrence, aussi bien d'un mot pris isolément que de mots apparaissant conjointement. Il s'agit plus précisément de la connaissance de la fréquence d'utilisation d'un mot à l'oral et à l'écrit (par ex. *mince* est un mot plus fréquent que *gracile*), ainsi que de la connaissance des mots habituellement utilisés ensemble, c'est-à-dire des collocations (par ex. *brouiller les pistes*, *se couvrir de ridicule*).
- La connaissance de tout ce qui est en lien avec les limitations d'usage d'un mot dans une situation donnée, c'est-à-dire le registre. Cela comprend en particulier la variation régionale (par ex. *mitaine* au Québec et *moufle* en France) et temporelle (par ex. *la gent*), ainsi que les rapports sociaux (par ex. *Salut Paul* par opposition à *Bonsoir monsieur Durand*).
- La connaissance du comportement syntaxique d'un mot, ce qui inclut à la fois ses propriétés grammaticales (par ex. *dire* est un verbe) et les relations qu'il

entretient avec d'autres mots de la phrase (par ex. *dire* a un sujet humain, a un complément direct qui exprime typiquement la chose dite, les paroles prononcées, et un complément indirect humain qui est le récipiendaire des paroles dites).

- La connaissance de l'ensemble des dérivations possibles pour un mot, comme la conjugaison (par ex. *marche-marchera*) et l'affixation (par ex. *sur-en-dette-ment*).
- La connaissance des réseaux d'associations entre les différents mots de la langue, les mots se définissant les uns par rapport aux autres par contraste. Les locuteurs d'une même langue font typiquement certaines associations entre les mots, celles-ci pouvant être de nature sémantique ou contextuelle. Des exemples d'associations de nature sémantique sont la synonymie (mots ayant des sens très proches dans un contexte donné, par ex. *incendie* et *feu* dans le contexte *L'incendie/le feu a finalement été contrôlé par les pompiers.*), l'antonymie (mots exprimant des concepts en opposition dans certains contextes, par ex. *petit* et *grand*), l'hyponymie (mots représentant des classes dont l'une est incluse dans l'autre, par ex. *moineau* et *oiseau*). Un exemple d'association contextuelle serait *accident de voiture*. D'après Richards (1976), les connaissances issues de ces réseaux d'associations serviraient, en partie, à sélectionner le mot approprié à un certain contexte, en expression écrite ou orale.
- La connaissance de la valeur sémantique d'un mot. Celle-ci comprend à la fois les traits sémantiques du mot, définis comme des éléments de sens combinés les uns aux autres (par ex. *chaise* est inanimé, non humain, etc. tandis que

époux est animé, humain, etc.) et sa valeur subjective, c'est-à-dire les éléments de connotation du mot (par ex. *maigre* a une connotation péjorative).

- La connaissance d'un certain nombre de sens associés à un mot (sa polysémie), chaque sens étant le résultat d'un processus de construction du sens en contexte (par ex. *monter des bagages, monter un meuble, monter un cheval, monter un coup*).

En conclusion, la connaissance d'un mot implique de très nombreuses dimensions en lien les unes avec les autres. C'est ainsi bien plus que la connaissance d'une forme associée à un unique sens, puisque les mots sont généralement polysémiques, sont soumis à des contraintes grammaticales et d'usage et entrent dans des réseaux d'associations complexes.

1.2.2. Les connaissances lexicales selon Nation (2001)

La classification établie par Nation (2001) des différentes connaissances impliquées dans l'usage d'un mot reprend en grande partie celles déjà dégagées par Richards (1976), comme le montre le tableau 1 ci-dessous, où les correspondances entre les deux classifications sont établies. Les types de connaissances mis en avant par Nation (2001) sont regroupés en trois classes : la forme, le sens et l'usage. Ces regroupements sont justifiés par les résultats de travaux (par ex. Ellis 1995 ; cité dans Nation 2001) qui indiquent que les types d'apprentissage sont différents selon que la forme ou le sens sont concernés.

Nation (2001)		Richards (1976)
Forme	Forme parlée	
	Forme écrite	
	Parties du mot	Dérivations possibles
Sens	Forme et sens	
	Concept et référents	Différents sens Valeur sémantique
	Associations	Réseaux d'associations
Usage	Fonction grammaticale	Comportement syntaxique
	Collocations	Probabilité d'occurrence
	Contraintes d'utilisation (registre, fréquence...)	
		Limitations d'usage

Tableau 1 : Connaître un mot, d'après Nation (2001)⁴ et Richards (1976)

La première classe concerne la forme d'un mot. La forme parlée correspond soit à l'identification ou la reconnaissance d'un mot entendu, soit à la prononciation d'un mot, ce qui comprend notamment la capacité à produire les sons et à placer l'accent tonique (dans les langues où cela est pertinent). La forme écrite est en lien direct avec le système d'écriture de la langue. Les parties du mot sont des éléments formels en lesquels un mot est décomposable ou analysable. L'idée qu'un mot peut dans certains cas être découpé en parties est à la base de la notion de famille de mots où, sous un mot (par ex. *nation*), on regroupe l'ensemble des autres mots qui lui sont reliés morphologiquement (par ex. *national*, *nationaliser*, *international*, etc.).

4 La partie gauche du tableau est tirée de Nation, 2001, page 27.

La deuxième classe regroupe les connaissances de type sémantique. « Forme et sens » renvoie à la possibilité d'établir un lien entre une forme (orale ou écrite) et un sens, car connaître l'un indépendamment de l'autre n'est pas suffisant. La classe « concept et référents » correspond à l'idée qu'un mot, en contexte, peut avoir différents sens, qu'il est possible de regrouper sous un même concept (le mot est alors dit polysémique). Par exemple, le mot *fourche* réfère à la fois à un outil pour l'agriculture et à la séparation d'une route en deux branches⁵. Les associations, quant à elles, représentent l'ensemble des relations sémantiques qu'un mot entretient avec d'autres mots du lexique.

La troisième et dernière classe regroupe les connaissances liées à l'usage. Les fonctions grammaticales comprennent la catégorie grammaticale d'un mot (par ex. nom, verbe, préposition, adjectif, etc.) et les structures grammaticales, qui consistent principalement en la liste des compléments d'un mot. Ces deux aspects déterminent l'usage du mot dans la phrase et, pour les verbes en particulier, une grande partie de la structure de la phrase. Les collocations, ou cooccurrences, indiquent quels mots sont susceptibles d'apparaître avec un mot donné. Les contraintes d'usage sont les contraintes d'ordre sociolinguistique, en particulier le registre (par ex. *voiture* ou *auto* vs. *bagnole* ou *char*), ou bien liées à la fréquence d'utilisation (par ex. l'emploi d'un mot moins fréquent ou vieilli à la place d'un mot plus courant).

Un aspect innovant de la classification de Nation (2001), par rapport à celle de Richards (1976), est qu'elle s'appuie sur la distinction, posée explicitement, entre réception (ce qui

5 L'exemple donné dans Nation 2001 est celui du mot *fork*, qui présente le même type de polysémie en anglais.

est à l'œuvre dans la compréhension orale ou écrite) et production (ce qui est à l'œuvre dans l'expression orale ou écrite). Chacun des types de connaissance lexicale est alors considéré selon cette double perspective, c'est-à-dire à la fois dans sa dimension réceptive et dans sa dimension productive. Cette distinction est établie afin de rendre compte du fait que les deux dimensions renvoient à des compétences différentes. Ainsi, la réception est le processus qui permet, par exemple, d'associer un sens et un concept à une forme lue ou entendue et de reconnaître les collocations dans lesquelles entre cette forme. La production, comme nous l'avons vu dans la section 1.1., est le processus qui aboutit, entre autres, à la sélection de la forme appropriée pour l'expression d'un sens donné, et à l'utilisation de mots entrant dans des relations de collocation avec ce mot. La pertinence de cette distinction trouve un écho dans la remarque de Miller et Fellbaum (1991), qui ont avancé l'idée que, en ce qui concerne l'accès lexical, ce qui est problématique en réception est la polysémie, puisqu'il faut associer à l'occurrence rencontrée un sens parmi plusieurs possibles, alors que ce qui est problématique en production est la synonymie, puisqu'il faut choisir entre plusieurs mots de sens voisin. Cela montre que la réception et la production sont des processus bien distincts qui font intervenir des compétences différentes⁶.

Connaître un mot implique donc de nombreuses dimensions. Toutes ces dimensions interviennent en production, mais certaines interviennent plus spécifiquement dans le choix lexical à proprement parler. Il s'agit des dimensions suivantes : concepts et référents (différents sens, valeur sémantique), associations, collocations, limitations

⁶ Sur la question des liens entre la réception et la production en matière de vocabulaire, on trouvera des éléments de réponse dans Duin et Graves (1987).

d'usage, fréquence, dérivations possibles et parties du mot, ainsi évidemment que l'association forme-sens. Ce sont principalement des critères sémantiques et d'usage.

1.3. Le choix lexical chez les apprenants

Nous avons vu que le lexique faisait intervenir de nombreuses connaissances relevant de plusieurs dimensions, mais on peut se poser la question de savoir si les apprenants de niveau intermédiaire-avancé rencontrent réellement des difficultés particulières lors du choix lexical. Afin d'explorer la connaissance lexicale des apprenants d'un point de vue qualitatif, plusieurs chercheurs procèdent à des analyses de corpus. En particulier, ils essaient d'appréhender la nature de la différence de connaissance lexicale entre les apprenants de niveau intermédiaire-avancé et les locuteurs natifs et à connaître le type d'erreurs que font les apprenants avancés. Dans cette section, nous présentons tout d'abord les conclusions d'études basées sur l'analyse comparative de corpus de locuteurs natifs et d'apprenants. Ensuite nous examinons les résultats d'études qui ont procédé à l'analyse d'erreurs de productions dans des corpus d'apprenants.

1.3.1. Connaissance lexicale incomplète des apprenants

Liu et Shaw (2001) ont cherché à évaluer la qualité de la connaissance lexicale chez les apprenants d'une langue seconde. Pour cela, ils ont comparé deux corpus de productions écrites, le premier appartenant à des locuteurs natifs de l'anglais et le

second à des apprenants de l'anglais, de niveau intermédiaire-avancé, ayant le chinois comme langue maternelle. L'objet de la comparaison était l'utilisation que ces deux groupes de scripteurs faisaient du verbe *make*. Les auteurs ont choisi ce verbe à cause de sa versatilité, tant en ce qui a trait au sens qu'à l'usage grammatical, qui, d'après eux, rend son apprentissage difficile. Ils ont plus spécifiquement cherché à mettre en évidence des aspects cachés de l'interlangue. Ils se sont alors interrogés sur la nature de la différence entre les productions des apprenants et celles des locuteurs natifs. Ils ne se sont pas intéressés aux erreurs mais à la distribution des usages dans les deux corpus. Ils ont quantifié les usages de *make* en regard de la construction dans laquelle il est employé, ainsi qu'en regard de son sens. Les résultats ont été analysés d'un point de vue plutôt qualitatif.

Dans la distribution des constructions du verbe, la différence entre les deux corpus est hautement significative pour certaines constructions, sous-employées ou bien sur-employées par les apprenants. En ce qui concerne les collocations, il y a moins de variation chez les apprenants, certaines étant uniquement employées par les locuteurs natifs. En ce qui a trait aux constructions figées, il y a moins de variation et moins d'utilisation chez les apprenants, et plus de combinaisons libres. Quant à l'aspect sémantique, certains sens de *make* sont à l'origine d'une différence significative entre les deux corpus : le sens causatif (par ex. faire en sorte que quelqu'un/quelque chose fasse quelque chose, forcer quelqu'un à faire quelque chose: *make the poisoned dog vomit*) est beaucoup plus employé chez les apprenants, tandis que le sens correspondant à faire l'action décrite par le nom qui suit (par ex. : *make a*

speech/decision) est beaucoup plus employé chez les locuteurs natifs. Ainsi, la distribution des différents usages du verbe *make* ne coïncide pas dans les deux corpus, certains sens, collocations ou constructions étant plus représentés dans l'un que dans l'autre. La conclusion que les auteurs ont tiré de ces résultats est que les apprenants ne peuvent utiliser toute la richesse sémantique et grammaticale des mots en production, même pour des mots très fréquents qu'ils disent connaître. Ainsi, la compétence lexicale en langue seconde diffère qualitativement de celle des locuteurs natifs.

Bolly (2005) a également comparé deux corpus de productions écrites, en français cette fois, l'un provenant de textes de locuteurs natifs, l'autre d'apprenants de langue maternelle anglaise. Elle a procédé à une analyse de fréquence des emplois du verbe *prendre* dans des constructions figées ou semi-figées de type 'verbe + nom'. Le programme qu'elle a utilisé a permis de comptabiliser les emplois de *prendre* en fonction du nom utilisé (par ex. *prendre conscience*, *prendre patience*) et de la construction (par ex. *prendre forme*, *prendre la forme de*). Bolly (2005) est arrivée au même type de résultats que ceux de l'étude de Liu et Shaw (2001), à savoir le sous-emploi ou le sur-emploi de certaines collocations ou structures chez les apprenants. Par exemple, les séquences du type *prenons l'exemple de*, *prenons un exemple* et *prenons comme exemple* sont beaucoup plus utilisées par les locuteurs natifs, voire même utilisées seulement par eux.

Ces deux études, même si elles ne permettent pas de généraliser puisque chacune d'elles ne se base que sur un seul verbe, tendent à montrer que les apprenants, même à

un niveau intermédiaire-avancé, ne maîtrisent pas le vocabulaire dans toute sa profondeur, même pour des mots extrêmement fréquents. Ces études mettent principalement en évidence des phénomènes de sur- ou de sous-emploi, la manière d'utiliser le vocabulaire étant différente chez les apprenants et les locuteurs natifs. Nous nous tournons à présent vers l'analyse des erreurs lexicales commises par les apprenants, car elle devrait pouvoir nous renseigner sur le type de difficultés auxquelles ils sont confrontés.

1.3.2. Connaissance lexicale imparfaite des apprenants

Un certain nombre de chercheurs ont procédé à une analyse des erreurs de textes d'apprenants de niveau intermédiaire-avancé, afin de mettre en évidence les difficultés rencontrées par les apprenants. Dans un corpus de 200 000 mots (FRIDA) provenant de compositions écrites d'apprenants de français langue étrangère de niveau intermédiaire à avancé, Granger (2003) a compté les erreurs toutes catégories. Les erreurs lexicales représentent 15,51 % des erreurs et sont des erreurs de sens, de cooccurrence et de régime.

Hamel et Milićević (2007) se sont penchées sur l'analyse des erreurs lexicales uniquement. Elles ont analysé un corpus de 50 000 mots (Dire Autrement) formé de compositions écrites d'apprenants de français langue seconde de niveau intermédiaire-avancé. Elles ont proposé une typologie descriptive d'erreurs lexicales qui repose sur

trois classes et de nombreuses sous-classes : les erreurs liées au sens (quasi-synonyme, sens proche, sens fictif, générique, sens incompatible, situation incompatible), les erreurs de forme (forme erronée, forme fictive et forme analytique) et les erreurs de cooccurrence (collocation et régime). Leurs résultats indiquent que 7,4 % des mots du corpus contiennent une erreur lexicale et que parmi ces erreurs, 63 % sont des erreurs de sens, 25 % des erreurs de cooccurrence et 12 % des erreurs de forme. En particulier, dans la classe *sens*, les sous-classes *sens proche* et *quasi-synonyme* représentent à elles-deux 40 % des erreurs du corpus. Les mots de sens proche « partagent des composantes importantes de sens, mais ne sont jamais substituables en contexte, par exemple « faire des exercices *nautiques*. → aquatiques » » (Hamel et Milićević, 2007, p.32). 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 « Buvez beaucoup de *fluides*. → liquides » » (Hamel et Milićević, 2007, p.31). Si l'on ajoute à ces deux sous-classes la sous-classe *collocation*, alors on aboutit à 56 % des erreurs lexicales du corpus. Les collocations sont composées d'un mot base et d'un collocatif, l'apprenant pouvant associer au mot base un collocatif erroné, comme dans l'exemple « Si vous *recevez* un coup de soleil [...]. → attrapez » (Hamel et Milićević, 2007, p. 34). Aussi, les erreurs de régime et de sens fictif (sens totalement inapproprié, imaginé par l'apprenant pour une forme donnée, par exemple « [...] ce qui m'a permis de *trafiquer* dans mon village. → circuler » (Hamel et Milićević, 2007, p. 32)) représentent chacune 9 % des erreurs du corpus. Les autres types d'erreurs sont en nombre beaucoup plus faible. Les auteures en ont conclu que « 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* » (Hamel et Milićević, 2007, pp. 41-42) représentent des difficultés particulièrement importantes pour l'apprenant.

Dans une autre étude, portant sur l'anglais langue seconde, Hemchua et Schmitt (2006) ont analysé les erreurs lexicales dans les compositions en anglais d'apprenants avancés de langue maternelle thaïe. L'étude s'appuie sur une taxonomie pour identifier les erreurs commises par ces apprenants, déterminer leur fréquence relative et la proportion d'erreurs ayant pour cause le transfert de la L1. Hemchua et Schmitt (2006) reprennent la classification de James (1998) et y ajoutent des sous-catégories inspirées de Leech (1981). La classification des erreurs compte deux catégories principales, les aspects formels et les aspects sémantiques, et 24 sous-catégories. Le corpus analysé est constitué de 20 compositions écrites d'un texte argumentatif d'environ 350 mots réalisées dans un temps limité et sans consultation possible d'un dictionnaire.

Les erreurs de type formel se divisent en trois catégories : les erreurs formelles de sélection (erreur de préfixe ou de suffixe, erreur de voyelle ou de consonne, mot proche phonétiquement, mauvais choix de catégorie grammaticale, mauvais choix de congénère), les malformations, le mot inexistant étant induit par la L1 (emprunt à la L1, arrangement formel d'un mot de la L1 pour qu'il ait une allure de la L2, calque (mot composé ou syntagme lexical)) et les distorsions, qui sont d'autres types de mots inexistantes.

Les erreurs de type sémantique comprennent les confusions entre deux mots liés par le sens (relations d'hyponymie, d'hyponymie, de cohyponymie, de quasi-synonymie), les erreurs de cooccurrence (mauvais choix d'un mot pour des raisons sémantiques, de fréquence ou de combinaisons arbitraires, mauvais choix de préposition), les erreurs de connotation et les erreurs stylistiques (verbosité, sous-spécification du sens). Il est à noter que la taxonomie de James (1998) comprenait deux classes d'erreurs sémantiques : confusion de relation de sens et erreurs de cooccurrence. Dans la catégorie d'erreurs de cooccurrence, Hemchua et Schmitt (2006) ajoutent les erreurs liées à l'utilisation des prépositions, que ce soit avant un nom par exemple (une émission à la télévision) ou après un nom, verbe ou adjectif (sa confiance en lui). Ensuite, parmi les erreurs sémantiques, ils ajoutent les erreurs de connotation et les erreurs stylistiques.

Les résultats de l'étude indiquent que les erreurs de type sémantique sont beaucoup plus nombreuses que celles de type formel, en particulier dans les sous-classes de cooccurrence et de confusion entre des mots de sens reliés, et par conséquent que la maîtrise des aspects sémantiques et de cooccurrence du vocabulaire en usage productif à un niveau avancé est imparfaite. Par ailleurs, parmi toutes les erreurs possibles, ce sont les erreurs de la sous-catégorie « quasi-synonymes » qui sont les plus fréquentes (19,54 %) et qui se retrouvent dans 80 % des copies d'étudiants. Ce sont en général des erreurs de registre, des erreurs liées à la non-équivalence sémantique entre les quasi-synonymes (différents sens sous-jacents, sens qui ne se recouvrent pas complètement), et enfin des erreurs liées à des différences d'usage entre les quasi-synonymes. D'autres

erreurs très fréquentes sont celles relevant d'un mauvais choix de préposition (12,64 %) ou de suffixe (9,20 %). L'étude conclut sur le fait que les erreurs semblent davantage liées à la difficulté intrinsèque que présente la langue seconde aux apprenants qu'au transfert à partir de la L1.

Dans les trois études mentionnées ci-dessus, qui consistent en des analyses statistiques d'erreurs dans des corpus d'apprenants, les chiffres ne sont pas toujours identiques. Cela s'explique par les catégorisations effectuées par les différents auteurs, qui, bien que renvoyant globalement aux mêmes phénomènes, ne se recoupent pas tout à fait. Par exemple, dans Granger (2003), les erreurs de forme, de morphologie, de registre et de style ne sont pas assimilées à des erreurs lexicales et sont comptabilisées à part. Par contre, dans les études de Hamel et Milićević (2007) et Hemchua et Schmitt (2006), les erreurs lexicales incluent, avec les erreurs de sens, les erreurs formelles, de registre et d'affixation. Cependant, le tableau qui se dégage des études est que les erreurs lexicales principales sont celles qui sont liées au sens et en particulier celles qui résultent d'un mauvais usage de quasi-synonymes, ainsi que les erreurs de collocation et de choix de préposition.

1.4. Le développement du sens lexical en langue seconde

Dans cette section, nous présentons des travaux qui apportent des éléments explicatifs à la question de pourquoi le choix lexical est problématique pour les apprenants. Ces

difficultés se situent sur deux plans, que nous exposons séparément bien qu'ils soient reliés. Le premier concerne la manière dont s'effectue le développement du sens lexical (1.4.1.). Le deuxième concerne la manière dont se développent les réseaux lexicaux (1.4.2.).

1.4.1. Le développement des relations concept-mot

Dans cette section, les études présentées expliquent les difficultés des apprenants dans le choix lexical par le fait que les concepts ne sont pas équivalents dans les différentes langues. Ce fait se traduit par des répartitions différentes de la polysémie entre la L1 et la L2, entraînant des erreurs dans le choix d'une forme.

Certaines études (Ijaz, 1986 ; Jiang, 2000) proposent des modèles ou des hypothèses visant à expliquer les phénomènes mentionnés dans la section 1.3. (distribution des sens différente chez les locuteurs natifs et les apprenants, erreurs lexicales liées au sens ou aux collocations), qui étaient alors mis en évidence par des analyses de corpus. Elles s'appuient pour la plupart sur la notion de concept, sachant que les associations concept-mot ne sont pas nécessairement similaires dans deux langues différentes.

Jiang (2000) met de l'avant l'hypothèse du développement du sens lexical en trois stades. Le premier correspond à l'association entre un mot de la L2 et le mot de la L1 qui est sa traduction, l'accès au sens se faisant par le biais de cette traduction. Le

deuxième stade correspond à l'association directe entre un mot de la L2 et le sens de sa traduction en L1, le troisième à l'association entre un mot de la L2 et son sens. Selon Jiang (2000), le troisième stade n'est jamais atteint en contexte d'apprentissage d'une langue seconde avec un enseignement, c'est-à-dire en tant que langue étrangère. Il y aurait donc fossilisation au deuxième stade, avec association de concepts de la L1 aux formes de la L2. Pour passer au stade 3, il faudrait qu'il y ait une restructuration conceptuelle.

S'appuyant sur les travaux de Jiang (2000), Chiu (2009) teste l'hypothèse que les apprenants voient les concepts de la L1 et de la L2 comme équivalents. D'après cette hypothèse, les mots de la L2 sont associés par l'apprenant aux concepts correspondant à leur traduction dans la L1. Cette hypothèse est mise à l'épreuve en examinant la capacité des apprenants à distinguer sémantiquement deux mots de la L2 de sens proches, mais distincts, qui se traduisent par le même mot dans la L1. Cette capacité de distinction représente, selon Chiu (2009), un indice de la qualité de la connaissance lexicale de l'apprenant, qui est corrélée à une mesure de la taille de son vocabulaire. Chiu (2009) a mis en évidence un lien entre le nombre de mots connus et la qualité de la connaissance lexicale, sans toutefois que celle-ci ne se rapproche de celle de locuteurs natifs. Cette étude lui permet de confirmer l'hypothèse de Jiang (2000) selon laquelle le passage du stade 2 au stade 3 n'est pas évident.

Dans Ijaz (1986), la signification d'un mot est vue comme une catégorie, comprenant des sens centraux et d'autres partageant plus ou moins de propriétés (des dimensions

sémantiques) avec le sens central. Le fait de s'éloigner du sens central fait se rapprocher d'autres catégories, donc d'autres mots. Ces sens sont donc plus difficiles à maîtriser par les apprenants. Deux mots de deux langues différentes ne correspondent (en général) pas au niveau de leurs frontières sémantiques. L'apprenant doit donc restructurer les concepts de la L1 ou développer de nouveaux concepts correspondant aux mots de la L2. Au niveau débutant, le transfert de la signification d'un mot de la L1 à un mot de la L2 est courant, comme l'a également suggéré Jiang (2000). Dans son étude, Ijaz (1986) cherche à évaluer les différences de représentation des frontières sémantiques entre des prépositions ayant des sens proches, *on* et *over*, chez des locuteurs natifs et des apprenants. Elle observe des erreurs et des différences concernant la distribution des emplois de ces deux groupes de prépositions : *on* (*on top of, upon, onto*) et *over* (*above*). Cela donne du poids à son hypothèse de différence des frontières sémantiques entre les groupes de participants. Ijaz (1986) montre d'une part qu'il y a plus de cohésion entre les groupes (apprenants vs. locuteurs natifs) pour les usages des sens centraux de ces prépositions que pour les sens périphériques, d'autre part que les concepts de la L1 sont transférés dans la L2 par l'intermédiaire du sens central, ce qu'elle interprète comme étant une stratégie de l'apprenant.

Jiang (2000) décrivait le développement du sens lexical en termes de stades. Une autre manière de décrire le développement des sens associés aux mots est de se baser sur une vision de type processus. Beheydt (1987) propose de décrire l'apprentissage du vocabulaire comme un processus de « sémantisation », c'est-à-dire « a continuing process of getting acquainted with verbal forms in their polysemous diversity within

varying contexts » (Beheydt, 1987, p. 55). De même, Sonaiya (1991) décrit l'évolution de la connaissance sémantique associée aux mots comme un processus de raffinement de concepts et de réajustement des frontières qui les délimitent, en particulier par contraste avec les autres mots qui partagent une partie de leur sens. Elle définit donc l'acquisition du vocabulaire comme « a process of continuous lexical disambiguation » (Sonaiya, 1991), contrairement à Jiang (2000) pour qui il y a une fossilisation.

Sonayia (1991) s'intéresse particulièrement à l'aspect lié à la quasi-synonymie dans l'apprentissage du vocabulaire. En effet, l'acquisition de mots proches sémantiquement dans une L2 est souvent source de difficulté pour les apprenants, en particulier lorsque les concepts associés aux mots ne sont pas équivalents dans la L1 et la L2, et donc le choix effectué parmi différents mots de sens proche peut s'avérer erroné. Sonayia (1991) considère les erreurs comme révélatrices d'un lien sémantique mal maîtrisé entre le mot employé et le mot qui aurait convenu. Ainsi, d'après elle, l'acquisition du vocabulaire résulte de l'apprentissage des relations sémantiques qui existent dans un lexique entre des mots de sens proche.

De ces études, il se dégage l'idée que le système conceptuel déjà existant dans la L1 peut être une source de difficulté à la maîtrise des concepts de la L2, c'est-à-dire de la polysémie en L2. Cette difficulté s'expliquerait par des délimitations de catégories (frontières) différentes dans la L1 et la L2, associées à la mise en équivalence des mots des deux langues par l'intermédiaire du sens central (Ijaz, 1986). Contrairement à Jiang (2000) pour qui il y a fossilisation, Beheydt (1987) et Sonayia (1991) stipulent un

développement continu des concepts chez les apprenants. Enfin, d'après Sonayia (1991), l'apprentissage lexical renvoie à l'apprentissage de relations entre les mots⁷, ce que nous allons développer dans la prochaine section.

1.4.2. Le développement des réseaux lexicaux

Dans cette section, les études présentées mettent en évidence l'importance des réseaux lexicaux dans le développement lexical et le choix lexical. En effet, une manière d'appréhender la connaissance lexicale consiste à représenter le lexique, non pas à travers la somme de ses unités considérées individuellement, mais comme un tout formant un réseau (par ex. Meara, 1983, 1996 ; Henriksen, 1999). Ces réseaux sont constitués de liens mettant en relation les unités lexicales. Comme le rappellent Henriksen (1999) et Read (2004), ces liens peuvent être de nature paradigmatique ou syntagmatique. Les relations de type syntagmatique concernent les liens entre des mots au sein d'une phrase. Un exemple typique de relation syntagmatique est la relation de cooccurrence. Les relations de type paradigmatique, quant à elles, renvoient aux liens sémantiques entre des mots du lexique, comme dans les relations de synonymie ou d'antonymie. Ces deux types de relations, syntagmatiques et paradigmatiques, contribuent à former un réseau lexical. Ainsi, les réseaux d'associations partiels, centrés sur un seul mot, qui faisaient partie des connaissances sur un mot listées dans les

7 La justesse de la représentation sémantique chez l'apprenant est associée à la maîtrise des frontières des concepts pour Ijaz (1986) et à la maîtrise des relations entre des mots de sens proche pour Sonayia (1991).

modèles de Richards (1976) et de Nation (2001), pourraient être considérés comme des extraits de ce grand réseau lexical que forme le lexique.

Dans la mesure où un apprenant doit choisir un mot parmi d'autres mots, la justesse du choix lexical dépend de la qualité du réseau lexical dont il dispose. Dans ce sens, différents auteurs (par ex. Meara, 1983 ; Henriksen, 1999) associent étroitement la compétence lexicale à l'organisation du lexique ou au réseau sémantique reflétant la connaissance lexicale de l'apprenant. Henriksen (1999) appelle processus de sémantisation les aspects de développement du lexique qui incluent la définition, la référence d'une part et les relations sémantiques avec les autres mots d'autre part. D'après elle, les apprenants ont deux tâches principales : la première consiste à développer des concepts et à associer une forme à un concept, la seconde à construire des réseaux sémantiques. Cela renvoie à l'idée que la langue forme un système dont les unités entrent dans des rapports d'opposition et dans lequel la valeur d'un mot s'appréhende par ses différences avec les autres mots (Saussure, 1916/1995).

Dans le domaine de la recherche sur l'acquisition du vocabulaire, certaines études se penchent sur la notion de réseau lexical. Meara (1983, 1996) s'est intéressé au fait que les réseaux lexicaux diffèrent chez les locuteurs natifs et les apprenants, et que leur organisation se reflète dans la compétence. Meara (1983) se demande ce qui caractérise le lexique mental d'un apprenant et ce qui le différencie de celui d'un locuteur natif. Pour répondre à ces questions, il analyse les résultats de tests d'association chez ces deux catégories de locuteurs. Les tests d'association consistent à présenter un mot

au participant, qui doit indiquer le premier mot qui lui vient à l'esprit. Les réponses peuvent être en relation syntagmatique aussi bien que paradigmaticque avec le mot présenté. Meara (1983) constate qu'alors que les résultats de tests d'association sont statistiquement extrêmement prévisibles chez les locuteurs natifs, les réponses sont très variées et hétérogènes chez les apprenants, malgré le fait que la taille de leur vocabulaire soit plus petite. De plus, les mots produits par les apprenants en réponse à un stimulus peuvent ne jamais apparaître dans les résultats des locuteurs natifs. Meara (1983) suppose alors que les mots constituant le lexique des apprenants n'ont pas entre eux des liens sémantiques forts. Dans une étude ultérieure publiée en 1996, il pose l'hypothèse qu'outre sa taille, l'organisation du lexique est déterminante pour la performance de l'apprenant. Ainsi, plus les apprenants possèdent un lexique structuré, plus leur performance dans des tâches linguistiques est élevée. Il pose une seconde hypothèse qui est que les aspects liés à l'organisation interne du lexique gagnent en importance au fur et à mesure que la taille du lexique augmente.

Haastrup et Henriksen (2000) se sont penchées sur l'acquisition d'adjectifs dans une étude longitudinale. Elles pensent que l'acquisition de vocabulaire concerne davantage le système (relations avec les autres mots) que les unités lexicales prises individuellement. Elles recourent donc au concept d'intégration, qui est une composante de l'acquisition, et qui consiste à insérer le nouveau mot dans des réseaux de relations entre mots (paradigmatiques et syntagmatiques). Elles considèrent les relations de synonymie, d'antonymie et de gradation. Elles s'intéressent en particulier à l'évolution du

champ lexical pour *happy*. Elles concluent que ce processus est lent et difficile, et que le développement à travers le temps est faible.

Comme l'a fait remarquer Meara (1996), la notion de réseau lexical s'articule avec celle de la taille du lexique. Dans l'optique du choix lexical, la notion de taille apparaît fondamentale. En effet, la connaissance d'une certaine quantité de mots est nécessaire pour permettre la différenciation entre des mots candidats à la production. Si le lexique est de petite taille, alors l'éventail d'unités lexicales pouvant permettre un choix est lui-même restreint, limitant de ce fait les possibilités de choix, qui pourraient même dans certains cas être inexistantes, en particulier dans le cadre d'une production spontanée. En production écrite cependant, l'apprenant peut utiliser des outils variés (dictionnaires, manuels, Google, etc.). Il se peut donc qu'il effectue un choix portant sur une forme qui ne faisait pas partie de son vocabulaire. Cela étant dit, ce mémoire se centre sur les aspects sémantiques, conceptuels et de réseaux, liés au choix lexical.

1.5. Synthèse

Nous avons vu en 1.1. le rôle central joué par le lexique dans la production d'énoncés. L'accès lexical en production étant guidé par le sens (Levelt, 1989), les représentations sémantiques du lexique des apprenants doivent être suffisamment élaborées pour permettre la sélection des mots adéquats pour l'expression d'une idée. À travers les modèles présentés dans la section 1.2., qui établissent les nombreuses dimensions

participant à la connaissance d'un mot, il est possible d'entrevoir toute la complexité de la maîtrise du vocabulaire. Chez des apprenants de langue seconde, le résultat de l'accès lexical repose en grande partie sur la qualité des connaissances lexicales qui, d'après Henriksen (1999), incluent la connaissance des concepts associés aux mots ainsi que celle du réseau sémantique. En effet, la sélection du mot à utiliser se fait parmi d'autres mots au sein du lexique mental (Levelt 1989, De Bot 1992). Dans la section 1.3., il a été établi que la compétence lexicale des apprenants de niveau intermédiaire-avancé diffère de celle des locuteurs natifs, et ceci par des études ayant des objets et méthodes d'analyse différents. Des analyses contrastives de corpus mettent en évidence la non-maîtrise de toute la profondeur de mots très fréquents, avec en particulier des phénomènes de sur- et de sous-emploi. Des analyses d'erreurs révèlent la maîtrise imparfaite des connaissances sémantiques et de cooccurrence, avec certaines difficultés particulières mentionnées par les auteurs : les mots de sens proche, les collocations, le régime, la suffixation. Deux explications principales de ces difficultés ont été présentées en 1.4., tout d'abord la non-équivalence des concepts entre les langues associée au développement incomplet des concepts de la L2 chez les apprenants, qui se réfèrent aux concepts de leur langue maternelle, mais aussi le développement insuffisant des réseaux lexicaux chez les apprenants, les liens sémantiques entre les mots n'étant pas assez forts. Dans ces conditions, il apparaît clairement que le choix lexical est source de difficulté pour l'apprenant.

Les apprenants tentent de verbaliser au mieux une idée, alors que leur connaissance des moyens existant pour l'exprimer est partielle et approximative. Le problème ne

réside pas nécessairement dans le fait qu'ils effectuent un choix erroné, comme pourrait le faire un locuteur natif, mais dans le fait que leur connaissance est trop lacunaire pour leur permettre d'effectuer un réel choix.

Il ressort de tout cela que le phénomène de choix lexical sur lequel nous travaillons concerne l'étape de la formulation en production, étape durant laquelle l'apprenant repose sur la quantité et la qualité de ses connaissances lexicales pour pouvoir transmettre son intention de communication. Ayant vu dans cette section qu'au niveau intermédiaire-avancé, le système lexical n'est généralement pas maîtrisé dans toute sa richesse et sa complexité, des lacunes dans la connaissance lexicale des apprenants subsistent et donnent lieu à des productions imparfaites. Par ailleurs, si l'on considère la connaissance lexicale d'un point de vue évolutif (Schmitt, 1998 ; Haastrup et Henriksen, 2000), il semble que le développement du sens lexical chez les apprenants avancés progresse relativement lentement. Ces difficultés ne se résolvent donc pas toutes seules avec le temps et nécessitent une intervention pédagogique. Nous poserons alors l'hypothèse qu'une approche didactique visant à renforcer les connaissances lexicales d'une manière active peut contribuer à l'accroissement de la compétence des apprenants dans le choix lexical. Cependant, comme il y a beaucoup à apprendre, relativement peu d'outils, de moyens, et de temps de classe, comment s'y prendre ? Avant de proposer des pistes de méthodologie pour l'enseignement du choix lexical, nous allons procéder à une étude empirique de corpus.

2. ÉTUDE DU CORPUS : FONDEMENTS EMPIRIQUES ET LINGUISTIQUES

Afin de proposer des pistes d'intervention pour l'enseignement du choix lexical, nous allons nous appuyer sur l'étude d'un corpus d'apprenants. Ce corpus nous renseigne sur les erreurs de choix lexical réellement commises par les apprenants dans le cadre d'un cours de français écrit de niveau intermédiaire-avancé. La caractérisation linguistique de ces erreurs nous a permis de mieux cerner les sources de difficulté des apprenants dans le choix lexical et sous-tend la séquence didactique que nous exposerons par la suite, dans la section 3.

En 2.1., nous présentons brièvement le corpus, le contexte pédagogique du cours dont il est tiré, ainsi que notre méthode d'analyse. En 2.2., nous donnons une vue d'ensemble des catégories d'erreurs de choix lexical, qu'elles soient formelles, stylistiques, pragmatiques ou sémantiques. En 2.3., nous analysons de manière plus approfondie les catégories de nature sémantique en nous servant de notions linguistiques.

2.1. Corpus, contexte pédagogique et démarche d'analyse

Le corpus est constitué d'une sélection de textes d'apprenants du français langue seconde produits dans le cadre d'un cours de communication écrite de niveau intermédiaire-avancé. Le spectre du niveau des étudiants inscrits à ce cours est assez large, correspondant aux niveaux B1 et B2 du Cadre européen commun de référence

pour les langues. Pour pouvoir s'inscrire à ce cours, les étudiants doivent avoir complété l'équivalent de 360 heures de cours de français, certains d'entre eux ayant toutefois à leur actif au moins 450 heures d'étude du français. Les étudiants sont testés lors du premier cours et peuvent éventuellement se voir conseiller de changer de cours si leur niveau n'est pas jugé en adéquation avec celui du cours (qu'ils soient pas assez ou trop avancés). Cependant, les étudiants dont le niveau serait jugé insuffisant mais qui vérifient les conditions d'acceptation officielles au cours doivent y être acceptés. En outre, les niveaux atteints par les apprenants après 360 heures de cours présentent naturellement une grande variabilité. Les apprenants sont des étudiants d'une université anglophone et sont de diverses langues maternelles.

Dans le corpus, trois types de textes ont été retenus pour la présente étude : des textes descriptifs, narratifs et informatifs. Ils ont été produits dans un intervalle de temps n'excédant pas deux semaines, à l'extérieur de la salle de classe et avec la possibilité d'utiliser des outils tels que des dictionnaires, des grammaires et des correcteurs (par ex. le logiciel *Antidote*). Le corpus comprend les textes de deux sessions universitaires et totalise ainsi un peu plus de 108 200 mots. Les textes du corpus ont été anonymisés.

Le cours d'écrit s'articule autour de la notion de type de texte et utilise la démarche d'autocorrection. Pour chaque texte soumis par l'étudiant, le professeur utilise des codes (Bouffard, 2003) pour signaler des erreurs de différentes natures. Il y a au total 27 codes, rendant compte de phénomènes aussi divers que l'accord, l'orthographe, le vocabulaire, le style, le choix des prépositions et des pronoms, la construction des

phrases, etc. Les codes qui rendent compte des erreurs de vocabulaire sont m-1 (barbarisme), m-2 (mot passe-partout) et m-3 (anglicisme). L'apprenant doit ensuite corriger les erreurs de son texte grâce aux indices fournis par les codes (une description exemplifiée des codes leur est fournie dans le recueil) et le soumettre à une nouvelle évaluation. Pour les erreurs signalées par le professeur, il est ainsi souvent possible d'accéder à l'autocorrection de l'apprenant, ce qui peut nous renseigner sur son intention de communication.

Notre étude du corpus consiste à catégoriser les erreurs de choix lexical afin de bâtir une intervention pédagogique basée sur les principales difficultés auxquelles les apprenants sont confrontés. Les catégories d'erreurs devraient fournir une réponse à la question : « quel est le type de 'problème' qui fait que ce mot ne marche pas dans ce contexte-ci ? ». La caractérisation des erreurs devrait ainsi orienter l'apprenant vers une solution à son problème de choix lexical. Elle devrait l'éclairer sur les raisons réelles de l'inadéquation de son choix, tout en le mettant sur la piste d'un mot qui conviendrait. Pour cette raison, il nous a semblé important de définir les classes d'erreurs sémantiques à l'aide de notions linguistiques qui puissent avoir un caractère opérationnel, c'est-à-dire qui permettront d'agir, de prendre des décisions.

Le plus souvent, les étudiants se sentent désemparés lorsqu'une erreur lexicale de nature sémantique leur est signalée, puisqu'il n'existe pas de « règles » lexicales auxquelles ils peuvent se référer pour analyser leur erreur. Ils se trouvent devant la mention d'une erreur en étant parfois incapables de comprendre pourquoi c'est une

erreur et comment la corriger. Les catégories d'erreurs sémantiques que nous allons présenter dans cette section ont en partie pour vocation de servir d'outil d'analyse aux apprenants lorsqu'ils se trouvent devant leur erreur.

Notre objectif, par cette classification d'erreurs, est donc de pouvoir indiquer à l'apprenant comment résoudre son problème de choix lexical. Comme les classifications de Hamel et Milićević (2007) et Hemchua et Schmitt (2006) présentées dans la section 1.3. n'avaient pas les mêmes objectifs et que, dans une catégorisation, les catégories sont liées aux objectifs, l'ensemble des classes dégagées ne répond pas précisément à notre besoin. La taxonomie utilisée par Hemchua et Schmitt (2006) était établie a priori et utilisait celle de James (1998), qui reposait sur la notion de connaissance lexicale, avec quelques ajouts. Elle était ensuite utilisée pour analyser les données de leur corpus. Certaines classes qu'ils ont dégagées soulignent des aspects qui ne sont pas nécessairement pertinents à notre étude, ou qui ne mériteraient pas nécessairement une catégorie. Le but de la classification de Hamel et Milićević (2007) était de servir de base à la conception d'un dictionnaire d'apprentissage basé sur la maîtrise des relations lexicales et de la paraphrase. Leurs catégories reflètent leur objectif, puisque certaines sous-classes obtenues expriment bien les relations entre le mot choisi et soit le mot qui aurait été adéquat, soit le contexte. La distinction qu'elles effectuent par exemple entre la catégorie « (quasi-)synonyme » et « sens proche » se révèle trop détaillée pour nos besoins. Par ailleurs, certaines classes que nous avons proposées ne se retrouvent pas dans leur classification (par ex. AFFIXE INADÉQUAT, dans le rapport sens/forme), alors qu'elles nous paraissent utiles pour résoudre des problèmes de choix lexical. Certaines

de nos catégories mêlent étroitement forme et sens, comme nous le verrons plus loin, ce qui n'est pas nécessairement le cas dans les autres classifications.

En résumé, les classifications d'erreurs présentées en 1.3, ainsi que les types de connaissances lexicales tels que présentés en 1.2. par Richards (1976) et Nation (2001) nous ont inspirée, mais eux-même ayant établi ces classifications pour d'autres fins, elles ne conviennent pas pour une utilisation directe dans un but d'amélioration du choix lexical. Toutefois, nous pouvons remarquer que la plupart des notions abordées dans ces différents modèles ou classifications se retrouvent d'une façon ou d'une autre, mais ne sont pas regroupés de la même manière ou abordées selon le même angle.

Nous avons procédé à une analyse de corpus également pour avoir des données brutes sur les erreurs d'apprenants, ce qui facilite grandement la construction d'une intervention pédagogique. Nous ne souhaitons pas définir une classification a priori, mais préférons que celle-ci émerge des données du corpus en étant orientée par la question de l'aide à la résolution du problème de choix lexical chez les apprenants. Ainsi, notre classification des erreurs de corpus a été établie de manière inductive. Nous essayons de présenter une analyse des erreurs basée sur des notions linguistiques. Nous croyons que la compréhension de ces notions s'avérera opérationnelle pour le choix lexical.

Tous les exemples présentés dans ce chapitre sont tirés du corpus. Les mots en gras dans les exemples identifient les erreurs lexicales. La correction de l'apprenant ou la forme attendue sont données entre parenthèses.

2.2. Présentation synthétique de la classification du corpus

Les catégories d'erreurs que nous avons dégagées dans le corpus peuvent être regroupées sous différentes classes : formelle, stylistique, pragmatique et sémantique. Comme il est important d'avoir une vue d'ensemble et qu'une intervention pédagogique pourrait cibler toutes les classes, nous les présentons toutes ici. Cependant seules les classes d'erreurs de nature sémantique seront analysées en 2.3. et nous ne retiendrons qu'elles pour la proposition didactique à la section 3.

2.2.1. Erreurs formelles

Nous avons dégagé sept catégories d'erreurs de nature formelle : MOT INEXISTANT, PARONYME, FAUX AMI, CATÉGORIE GRAMMATICALE, SOUS-CATÉGORISATION, EXPRESSION IDIOMATIQUE et CALQUE. Ce sont des erreurs de récupération de la forme : la forme du mot pour MOT INEXISTANT, PARONYME et FAUX AMI, la forme grammaticale pour CATÉGORIE GRAMMATICALE et SOUS-CATÉGORISATION, et enfin la forme de l'expression figée pour EXPRESSION IDIOMATIQUE et CALQUE. Les trois premières catégories regroupent des erreurs formelles qui se limitent au mot lui-même, sans faire intervenir le contexte. Les quatre catégories suivantes sont des erreurs formelles de cooccurrence, interprétables à la lumière du contexte. Les deux

premières sont d'ordre grammatical, tandis que les deux suivantes sont liées au figement. Nous allons maintenant présenter et définir ces catégories une à une.

Dans la catégorie MOT INEXISTANT, la forme choisie est inexistante en français. On ne la retrouverait pas dans un dictionnaire. C'est le cas dans l'exemple : *une **médecine*** (médecin).

Dans la catégorie PARONYME, la forme choisie par l'apprenant présente une ressemblance phonétique avec la forme attendue. Il ne semble y avoir aucun lien sémantique entre le mot choisi et le mot attendu. Des exemples d'erreurs liées à la paronymie sont : *au **contour** de la douane* (comptoir), *contrôler la **prorogation** du VIH* (propagation), *une **cadence** en fer* (carence), ***gager** de l'argent* (gagner), *un autre luxe informatique s'est **ajointé*** (ajouté), *comme un **poison** dans l'eau* (poisson).

Dans la catégorie FAUX AMI, la forme choisie est utilisée non avec le sens qui lui est associé en français, mais avec le sens d'un mot de forme similaire (congénère) en anglais. Il s'agit de ce que l'on appelle communément des faux amis. En voici des exemples : *le placement des **avertissements*** (publicités) pour « advertisement », *qui ont **définitivement** un effet néfaste* (sans aucun doute) pour « definitely », ***dépendant de leur importance*** (selon, en fonction de) pour « depending on », *ce **stage** de développement* (stade) pour « stage ». Les erreurs de cette catégorie sont bien des erreurs de forme, et non de sens, car le mot choisi n'a aucun rapport sémantique avec le

mot attendu. C'est en effet la proximité de forme entre un mot anglais et un mot français, malgré des sens différents, qui a provoqué l'erreur.

Dans la catégorie CATÉGORIE GRAMMATICALE, la catégorie grammaticale utilisée par l'apprenant est erronée. La famille de mot choisie est correcte, mais le mot sélectionné ne convient pas au contexte syntaxique, comme nous pouvons le voir dans les exemples suivants : *chaque **trouve*** (trouvaille), *le terminus n'était plus **lointain*** (loin), *vivre **pauvre*** (pauvrement). Ici, un verbe est utilisé à la place d'un nom et un adjectif à la place d'un adverbe.

Dans la catégorie SOUS-CATÉGORISATION, la construction syntaxique utilisée n'est pas correcte pour le mot choisi. Le schéma de sous-catégorisation n'est pas le bon. Nous considérons que la construction, ou le régime, fait partie de la forme. Voici quelques exemples d'erreurs de sous-catégorisation : ***permettre** la respiration de s'écouler naturellement* (permettre à), *son bureau **lui accommodait** avec trois journées [...]* (l'accommodait), *Valérie **portait son attention à** un couple* (portait son attention sur). Ici, ce sont des erreurs d'omission ou de mauvais choix de préposition. Ces erreurs peuvent avoir un effet sur le choix d'un pronom objet, comme dans l'exemple du verbe *accommoder* ci-dessus, où *lui* est utilisé au lieu de *l'*.

Dans la catégorie EXPRESSION IDIOMATIQUE, la forme exacte de l'expression ou de la locution figée n'est pas retrouvée. Ces erreurs sont, comme les collocations, des erreurs de cooccurrence, mais dans une expression dont le degré de figement est plus

important. En voici des exemples : *faire pignon sur rue* (avoir), *le laisser des yeux* (quitter), *la goutte qui fait déborder le verre* (vase), *donner l'eau à la bouche* (mettre). Il existe un lien sémantique entre le mot choisi et le mot attendu, mais l'expression étant figée, ces erreurs se classent parmi les erreurs de forme.

Dans la catégorie CALQUE, la forme de l'expression est calquée sur l'anglais. C'en est une traduction mot à mot, comme cela transparait dans les exemples suivants : *essayer son mieux* (faire de) pour « try his best », *comme magique* (par magie) pour « like magic », *pour le garder simple* (faire) pour « to keep it simple », *vivre jour par jour* (au jour le jour) pour « live day by day ». Dans cette catégorie, les expressions présentent un degré de figement dans les deux langues.

2.2.2. Maladresses stylistiques

Trois autres catégories, PLÉONASME, PÉRIPHRASE et MOT VAGUE, sans toujours être à proprement parler des erreurs, sont plutôt des maladresses stylistiques susceptibles d'être relevées par le professeur. Nous les présentons ci-après.

Dans la catégorie PLÉONASME, le mot choisi est redondant sémantiquement avec un mot du contexte. Voici quelques exemples de pléonasmes : *l'aura qu'elle émet* (son aura), *ses spécificités uniques* (ses spécificités), *une longue durée de temps* (période), *rester ensemble avec ceux qui sont importants pour soi* (près de).

Dans la catégorie PÉRIPHRASE, une périphrase est utilisée à la place d'un mot existant, comme dans les exemples suivants : **chose difficile** (difficulté), **vie lorsqu'elle était jeune** (jeunesse), **faire connaître sa colère** (exprimer), **faire des choses pour les autres** (aider), **amener sa tête plus proche** (approcher), **l'obscurité de l'auditorium empirait** (s'obscurcissait), **le train recommençait à bouger** (redémarrait), **rôle comme vedette** (premier rôle), **le train a commencé à se déplacer encore** (redémarré), **elle fait cinq pieds et six pouces de hauteur** (mesure), **entendre son nom déclaré** (être appelé), **les semaines sont venues et ont quitté vite** (ont passé), **un membre ancestral** (un ancêtre), **nous faisons toutes les choses ensemble** (tout).

Dans la catégorie MOT VAGUE, un mot très fréquent est utilisé alors qu'un mot plus précis, plus spécifique serait disponible, comme dans les exemples suivants : **elle a eu une grande influence sur ma vie** (a exercé), **ils pouvaient aller chez eux passer la veille de Noël avec leur famille** (rentrer), **elle admirait le travail qu'ils avaient fait** (accompli), **la moitié de leurs choses étaient absents** (affaires), **pour ne pas faire partir les ours** (fuir), **la majorité d'autres restaurants ont au moins plusieurs options pour végétariens** (offrent), **il a toujours été remarquable dans tout ce qu'il faisait** (entreprenait), **quelque chose qui changerait n'importe qui** (une expérience), **il était dans son pays maternelle** (habitait), **il y a des solutions** (existe), **avoir un sentiment** (éprouver), **avoir un emploi** (décrocher, accéder à), **avoir de l'audace** (faire preuve de). De nombreuses maladresses de ce type proviennent de la sur-utilisation des mots *chose*, *faire*, *être*, *avoir*, soit à l'intérieur d'une collocation, soit dans une expression libre. Il ne s'agit pas

d'erreurs, mais d'un manque de précision ou d'une faiblesse de style, qui peut affecter la qualité du texte produit.

Les erreurs stylistiques peuvent être interprétées comme des manifestations de la difficulté rencontrée par l'apprenant à utiliser le mot adéquat ou à varier son expression. Par exemple, le recours à une périphrase ou à un mot peu précis peut être un indice de connaissance lexicale incomplète ou du fait qu'un mot ne fait pas partie du vocabulaire actif de l'apprenant. Ces maladresses pourraient également être dues à une stratégie d'évitement de la part de l'apprenant, à une absence de prise de risque. La présence de pléonasmnes souligne également l'imperfection de la connaissance lexicale de l'apprenant.

2.2.3. Erreurs pragmatiques

Trois catégories, REGISTRE, CONNOTATION et FRÉQUENCE, relèvent de considérations pragmatiques. En effet, certains mots sont plus ou moins adaptés à un contexte communicatif donné.

Dans la catégorie REGISTRE, le mot utilisé n'est pas du registre approprié au contexte. Un mot familier est employé dans un contexte où l'on aurait attendu un mot plus formel. Voici quelques exemples d'erreurs appartenant à cette catégorie : *What ? a répliqué la **madame*** (oral, familier, alors que le niveau de langue du texte est relativement

soutenu), *cette **madame** accommodante* (femme), *quand mon père **gueule** après elle* (crie), ***turbiner** ensemble* (travailler), *la seule à avoir les **couilles** de patrouiller l'école* (le courage, le cran), *une soirée **de même*** (telle), *j'entendais les **docteurs** ainsi que la **police*** (médecins), *les coups de pied au visage d'un adversaire « **évaché** »* (étendu au sol).

Dans la catégorie CONNOTATION, le mot choisi est connoté d'une manière non prise en compte par l'apprenant. Par exemple, dans *crier dans la **face** de cette **femelle*** (au visage de cette femme), *face* et *femelle* sont péjoratifs, et par ailleurs également de style familier, alors que le style du texte est plutôt poétique, avec un vocabulaire recherché. Voici d'autres exemples : *cet **individu** est un membre de la famille* (homme), *l'élevage industriel fait la plus grande **contribution** à la pollution d'eau douce* (cause), où contribution a une valeur positive qui ne cadre pas avec le contexte.

Dans la catégorie FRÉQUENCE, le mot utilisé est un mot rare ou désuet, qui n'est pas adapté au style du texte, comme dans les exemples suivants : *les années **subséquentes*** (suivantes), ***reposer** sur une chaise longue ou un coussin* (se reposer), *je remarque **son habit*** (ses vêtements), *plusieurs **gens*** (personnes).

2.2.4. Erreurs sémantiques

Nous avons mis en évidence cinq catégories d'erreurs de nature sémantique : CONSTRUCTION INCOMPATIBLE, RELATION INCOMPATIBLE, COLLOCATION ATTENDUE, CONCEPT INADÉQUAT, AFFIXE INADÉQUAT. Toutes ces catégories font intervenir le sens d'une manière ou d'une autre, ainsi que le contexte d'utilisation.

Dans la catégorie CONSTRUCTION INCOMPATIBLE, le mot est polysémique et les différents sens entrent dans différentes constructions. La construction dans laquelle le mot est employé ne fonctionne pas avec ce sens. Par exemple, dans *elle n'a pas réalisé le camion* (réalisé qu'il y avait un camion), le verbe dans la construction choisie signifie « faire », alors que le sens recherché est celui de « remarquer ». De même, dans *filer les ciels gris* (fuir), le verbe, dans cette construction, renvoie à l'univers textile, alors que le sens recherché est celui de « fuir » qui correspond, avec la forme lexicale *filer*, à un emploi intransitif.

Dans la catégorie RELATION INCOMPATIBLE, l'association entre un argument et son prédicat ne fonctionne pas pour des raisons sémantiques, de restrictions de sélection. Il peut s'agir de la relation d'un verbe avec son sujet, avec l'un de ses compléments, ou encore de la relation entre un adjectif et le nom qu'il modifie, ou enfin de la relation entre un nom et son complément également nominal. Voici quelques exemples d'erreurs de cette catégorie : *une maison âgée* (*âgé* s'applique à un être humain), *le teint de ses*

cheveux (le teint est associé au visage), *c'est la première fois qu'un chien l'affecte autant* (touche, émeut) (le sujet de *affecter* doit désigner un événement ou être abstrait lorsque l'objet désigne un humain), *apprendre des connaissances* (le complément de *apprendre* doit désigner la matière enseignée, l'objet de l'apprentissage ; la connaissance est le résultat de l'apprentissage, le verbe utilisé avec *connaissances* devrait être *acquérir*).

Dans la catégorie COLLOCATION ATTENDUE, un mot est utilisé alors qu'il existe avec un mot du contexte une collocation que l'on s'attendrait à voir utiliser. Le sens du mot choisi est cependant correct. Il s'agit le plus souvent d'un synonyme du collocatif attendu. Voici quelques exemples d'erreurs de cette catégorie : *emporter l'attention de quelqu'un* (attirer), *faire des décisions* (prendre), *un bonheur énorme* (extraordinaire, formidable, immense).

Dans la catégorie CONCEPT INADÉQUAT, le concept associé au mot choisi ne correspond pas à l'intention de communication de l'apprenant. Toutefois, il existe des éléments de sens communs au mot utilisé et au mot attendu, qui sont des quasi-synonymes se différenciant par des nuances de sens. En voici quelques exemples : *mettre fin à un rapport* (relation/amitié), *Il est difficile de trancher dans cette industrie* (se démarquer), *enlever son meilleur enquêteur* (retirer).

Dans la catégorie AFFIXE INADÉQUAT, la forme utilisée se différencie de la forme correcte par un changement d'affixe. Le sens est ainsi modifié, même si un élément de

sens est commun aux deux formes. Les mots appartiennent à la même famille. Voici des exemples relevant de cette catégorie : *Elle **se renfermait** dans sa chambre* (s'enfermait), *des sentiments **changeables*** (changeants), *des femmes **rondelles*** (rondelettes), *ouvrir un **concessionnaire** d'automobiles* (une concession). Les préfixes ou suffixes des verbes, noms et adjectifs sont modifiés, supprimés ou ajoutés, occasionnant une différence sémantique.

2.2.5. Récapitulatif

Les catégories d'erreurs représentatives du corpus étudié et présentées dans les sections précédentes sont rassemblées dans le tableau 2 ci-dessous. Les classes d'erreurs ne sont pas nécessairement exclusives. Par exemple, *déléguer l'équipe* peut être analysé comme un mélange de RELATION INCOMPATIBLE (on ne peut pas déléguer une équipe, on délègue un travail) et de CONSTRUCTION INCOMPATIBLE (on délègue à son équipe).

Erreurs formelles	
MOT INEXISTANT	Ex : médecine (médecin)
PARONYME	Ex : <i>une cadence en fer</i> (carence)
FAUX AMI	Ex : <i>le placement des avertissements</i> (publicités)
CATÉGORIE GRAMMATICALE	Ex : <i>chaque trouve</i> (trouvaille)
SOUS-CATÉGORISATION	Ex : <i>permettre la respiration de s'écouler</i> (permettre à)
EXPRESSION IDIOMATIQUE	Ex : <i>faire pignon sur rue</i> (avoir)

CALQUE	Ex : pour le garder simple (faire)
Maladresses stylistiques	
PLÉONASME	Ex : ses spécificités uniques
PÉRIPHRASE	Ex : mettre toute sa concentration sur quelque chose (se concentrer)
MOT VAGUE	Ex : faire un travail (accomplir)
Erreurs pragmatiques	
REGISTRE	Ex : avoir les couilles de (cran)
CONNOTATION	Ex : cette femelle (femme)
FRÉQUENCE	Ex : les années subséquentes (suivantes)
Erreurs sémantiques	
CONSTRUCTION INCOMPATIBLE	Ex : réaliser le camion (réaliser que / remarquer)
RELATION INCOMPATIBLE	Ex : des lèvres roux (rouges)
COLLOCATION ATTENDUE	Ex : faire des décisions (prendre)
CONCEPT INADÉQUAT	Ex : enlever son enquêteur (retirer)
AFFIXE INADÉQUAT	Ex : se porter avec de l'aisance (se comporter)

Tableau 2 : Catégories d'erreurs de choix lexical

D'un point de vue quantitatif, les catégories les plus importantes sont COLLOCATION ATTENDUE, RELATION INCOMPATIBLE, CONCEPT INADÉQUAT et AFFIXE INADÉQUAT. Les erreurs de type CONSTRUCTION INCOMPATIBLE sont plus marginales.

Dans ce mémoire, nous nous intéressons au choix lexical d'un point de vue sémantique. Nous laisserons donc de côté les erreurs purement formelles, qui révèlent une erreur dans la récupération de la forme mais pas nécessairement dans le choix du mot d'un point de vue conceptuel, de même que les erreurs stylistiques, qui sont des

maladresses, et que les erreurs pragmatiques⁸. Nous nous concentrerons sur les erreurs qui ont une racine sémantique.

2.3. Analyse des classes sémantiques du corpus

Dans cette section, nous reprenons les catégories d'erreurs de choix lexical de nature sémantique en nous intéressant, cette fois-ci, aux notions linguistiques qui se trouvent derrière elles. Il s'agit d'aboutir à une caractérisation plus précise et plus opérationnelle, qui se révélera utile pour la didactique du choix lexical que nous développerons dans la partie 3. Après avoir analysé les cinq classes, nous tenterons d'établir des liens entre les classes. Des tableaux comprenant de nombreux exemples de la plupart de ces catégories d'erreurs se trouvent dans l'annexe 1.

2.3.1. CONSTRUCTION INCOMPATIBLE : polysémie et sous-catégorisation

La catégorie CONSTRUCTION INCOMPATIBLE peut être caractérisée grâce aux notions de polysémie, d'unité lexicale et de sous-catégorisation.

La polysémie se définit par le fait que des mots admettent plusieurs sens reliés de manière plus ou moins forte. Pour représenter les différents sens d'un mot, et ce faisant

⁸ Les classes dégagées par l'analyse de corpus pourraient cependant servir comme étiquettes de codes de correction, comme nous l'explicitons en 3.3.

spécifier une unité de base pour l'apprentissage du vocabulaire, Bogaards (1994) utilise le concept d'unité lexicale tel que défini par Cruse (1986), « l'union d'une forme lexicale et d'un sens unique ». Une unité lexicale est (Bogaards (1994, citant Cruse (1986)) :

« les plus petites parties qui répondent aux deux critères suivants :

- une unité lexicale doit être au moins un constituant sémantique
- une unité lexicale doit être au moins un mot ».

Une unité lexicale peut donc être une séquence de mots fonctionnant en bloc comme *lune de miel*, puisque le sens de l'expression n'est pas dérivé de celui de ses parties, mais pas un morphème comme *mé-*, puisque cela ne constitue pas un mot. Suivant cette définition, il est possible de faire correspondre à un mot plusieurs unités lexicales, comme dans l'exemple donné par Bogaards (1994) du mot *rayon* qui admet une multitude de contextes d'emploi. Parmi ces emplois on peut citer « *Quelques rayons de soleil perçaient la verdure. Un des rayons de ma roue avant est cassé. Vous trouverez cet article au rayon des bagages.* » (Bogaards, 1994, p. 13), qui correspondent à des unités lexicales différentes.

La polysémie se caractérise par une forme à laquelle sont associées plusieurs unités lexicales reliées entre elles par le sens. On distingue ainsi la polysémie de l'homonymie par le fait que dans le second cas, il n'est pas possible de trouver des liens sémantiques entre les unités lexicales. Ainsi, dans l'exemple du mot *rayon* ci-dessus, les deux premiers emplois relèvent de la polysémie, alors que le troisième relève par rapport aux deux premiers de l'homonymie.

La notion de sous-catégorisation renvoie au fait que des contraintes de construction, d'insertion dans la phrase, pèsent sur les unités lexicales, tout particulièrement pour les verbes, mais aussi pour les adjectifs et les noms. Par exemple, le verbe *dépendre* s'utilise avec la préposition *de* alors que le verbe *penser* se construit, entre autres, avec la préposition *à* ; l'adjectif *furieux* s'emploie avec la préposition *contre*. Or, pour certains mots polysémiques, les différentes unités lexicales peuvent entrer dans des schémas de sous-catégorisation différents, le sens de chacune dépendant alors de sa construction. Par exemple, les trois constructions suivantes du verbe « compter » correspondent à trois unités lexicales différentes : *compter des choses* (associer une valeur numérique), *compter sur quelqu'un* (tenir pour acquise l'aide de quelqu'un), *compter faire quelque chose* (avoir l'intention de faire quelque chose). Ainsi, les différentes constructions dans lesquelles entre un mot peuvent correspondre à des sens différents.

Les erreurs classées CONSTRUCTION INCOMPATIBLE correspondent aux cas où la construction utilisée existe, mais ne coïncide pas avec le sens recherché par l'apprenant. Un exemple en est : *Ma mère était trop fatiguer pour conduire, c'est pour ça qu'elle n'a pas réalisé le camion*. Dans cette construction, le verbe prend le sens de « faire » alors que le sens recherché est celui de « remarquer »/« voir ». La construction dans laquelle le verbe aurait dû être employé est : *elle n'a pas réalisé qu'il y avait un camion*. Ces erreurs se distinguent des erreurs formelles classées SOUS-CATÉGORISATION en ce qu'elles provoquent une modification du sens du mot. En effet, les erreurs classées SOUS-CATÉGORISATION, comme par exemple *dépendre*

sur, sont des erreurs de construction, celle-ci n'étant pas adaptée au mot choisi, et non de sens.

2.3.2. CONCEPT INADÉQUAT : concept, composante de sens, champ sémantique et synonymie

La catégorie CONCEPT INADÉQUAT peut être mieux comprise grâce aux notions de concept, de composante de sens, de champ sémantique et de synonymie.

Dans la catégorie CONCEPT INADÉQUAT, le sens du mot choisi par l'apprenant est proche de celui qui aurait convenu, mais différent. Le mot utilisé ne reflète pas exactement le sens que l'apprenant veut véhiculer dans la phrase. L'erreur est de nature conceptuelle. Considérons quelques exemples. Dans **distinguer sa richesse** (percevoir), le concept *distinguer* comporte la composante de sens « discrimination », qui n'est pas présente dans le verbe *percevoir* et qui est responsable de l'inadéquation du verbe dans ce contexte. Dans **obtenir un verre d'eau à la cuisine** (prendre), le concept *obtenir* suppose « de quelqu'un » qui ne correspond pas au contexte et ne se retrouve pas dans le concept *prendre*. Dans **collecter ses bibelots préférés** (rassembler), *collecter* suppose « de différents propriétaires », notion non présente dans *rassembler*. Dans **apercevoir ce grand homme** (observer), le concept *apercevoir* ne contient pas la composante « avec attention » qui serait nécessaire dans ce contexte. Dans ces exemples, le mot choisi et le mot juste se différencient par une composante de

sens. Il semble alors qu'un élément entrant dans la définition du mot choisi ne soit pas perçu par l'apprenant.

La notion de champ sémantique permet de rendre compte de l'ensemble des relations entre des mots dans un domaine particulier (par ex. le champ sémantique des verbes de perception). L'idée saussurienne sous-jacente est que les mots, ou les signes linguistiques, d'une langue s'opposent au sein de réseaux lexicaux par une composante de sens ou plusieurs, et que c'est par ce jeu d'oppositions que les mots prennent leur sens. De ce point de vue, le lexique constitue un ensemble structuré.

Par ailleurs, la notion de concept sous-entend un ensemble d'unités lexicales associées à une forme linguistique polysémique. Or, la synonymie se définit en relation avec la polysémie. Deux mots sont synonymes s'il existe deux de leurs unités lexicales qui possèdent le même sens et sont substituables⁹ dans un énoncé sans en altérer le sens. Comme il est rare que deux unités lexicales aient exactement le même sens (en général, le sens est légèrement différent, même lorsque les unités lexicales sont substituables, par ex. *Cette institution a pour mandat de promouvoir/favoriser la recherche.*), nous parlerons plutôt de quasi-synonymes (cf. également Hamel et Milićević, 2007). Les mots étant généralement polysémiques, deux mots quasi-synonymes dans un certain contexte ne le seront pas nécessairement dans un autre (par ex. *Une campagne publicitaire a pour but de promouvoir/*favoriser un nouveau produit. Une tisane en fin de repas, cela *promeut/favorise la digestion !*). Dans la

⁹ Il se peut que la substitution nécessite des modifications de la construction (par ex. *à la hâte* vs *avec précipitation*)

catégorie CONCEPT INADÉQUAT, l'erreur est souvent une conséquence de la polysémie lexicale, où un quasi-synonyme est employé dans un sens pour lequel les mots ne peuvent être utilisés l'un à la place de l'autre. Cela résulte du fait que la valeur du mot n'est pas maîtrisée dans le système lexical.

2.3.3. COLLOCATION ATTENDUE : collocations

Les collocations font partie des cooccurrences, c'est-à-dire des groupes de mots qui, par convention, sont fréquemment utilisés ensemble. Elles peuvent être définies comme se situant sur un continuum entre d'une part, les combinaisons libres, transparentes, et d'autre part, les locutions, figées et opaques (Bossé-Andrieu et Mareschal, 1999) ; elles sont identifiées par leur caractère à la fois quasi-transparent et prévisible.

Une collocation (par ex. *lancer un défi*, *relever un défi*) comporte un mot-base (*défi*) et un collocatif (*lancer*, *relever*), choisi en fonction du mot-base et dont l'interprétation sémantique en dépend (Mel'čuk, 1998 ; Bossé-Andrieu et Mareschal, 1999). Par exemple, *piquer*, dans la collocation *piquer la curiosité*, ne s'interprète pas de la même manière que dans le contexte *piquer quelqu'un avec un objet pointu*, où la combinaison est libre. C'est la présence du mot *curiosité* qui provoque l'interprétation spécifique de *piquer*, qui ne se combine pas avec beaucoup d'autres noms dans ce sens-là (stimuler, provoquer). Par ailleurs, un mot-base peut être associé à plusieurs collocatifs. Par exemple, pour le mot-base *défi*, un ensemble de collocatifs possibles et de sens proches

(avec l'idée d'initiation) inclut *adresser, jeter, lancer, mettre à*¹⁰. D'autres collocatifs de *défi*, dont les sens sont différents les uns des autres, sont par exemple *avoir le goût de, constituer, faire face à, relever*¹¹.

La notion de fonction lexicale (Mel'čuk, 1998) pourrait permettre d'opérationnaliser la description du phénomène de collocation. Une fonction lexicale est une fonction qui prend pour argument une unité lexicale et qui a pour valeur un ensemble d'unités lexicales synonymes entre elles. Les unités lexicales argument et valeur peuvent entrer en relation syntagmatique ou paradigmatic. Dans le cas des collocations, l'argument sera le mot base et la valeur un ensemble de collocatifs. Il existe de nombreuses fonctions lexicales (cf. par ex. Mel'čuk, Clas et Polguère, 1995 ; Mel'čuk, 1998). Par exemple la fonction 'Magn' est une fonction lexicale syntagmatique qui a pour sens « 'très', 'intense', 'à un degré élevé' » et « **Magn**(amour)=ardent, fou [...] **Magn**(boire)=comme un trou [...] **Magn**(souffrir)=atrocement » (Mel'čuk, Clas et Polguère, 1995, p.136).

On remarque que, dans une collocation, les combinaisons de mots sont contraintes, mais pas totalement, dans la mesure où il existe des choix à effectuer parmi un ensemble limité de collocatifs possibles. L'apprenant est confronté à un double choix. Tout d'abord le mot-base appelle certains collocatifs. Ensuite, parmi les collocatifs possibles, certains sont plus adaptés que d'autres, parce que les sens sont différents ou qu'il existe des nuances de sens. Nesselhauf (2003, 2007) signale que les collocations

10 *Dictionnaire des combinaisons de mots*, Éditions Le Robert, France

11 *Dictionnaire des combinaisons de mots*, Éditions Le Robert, France

où l'un des mots se combine avec plus de termes sont plus difficiles à maîtriser par les apprenants que celles où l'un des mots peut se combiner avec moins de termes. Bolly (2005) va également dans ce sens en posant l'hypothèse que plus les séquences semi-figées présentent de la variabilité, c'est-à-dire moins de figement, plus elles sont difficiles à acquérir.

Dans la catégorie d'erreurs COLLOCATION ATTENDUE, le mot utilisé est incorrect, mais un collocatif de sens proche fonctionnerait. En effet, le sens du mot utilisé est le plus souvent approprié, mais la combinaison obtenue est maladroite, du fait qu'elle entre en concurrence avec une collocation. Cela transparait dans les exemples du tableau 3 ci-dessous. Lorsque l'on considère, pour un mot base donné, le collocatif existant et le mot utilisé par l'apprenant, on remarque qu'ils sont proches sémantiquement. Mais seul l'un des deux est utilisable, à cause de l'existence d'une collocation avec le mot base.

Exemple	Mot base	Collocatif existant	Mot utilisé
<i>emporter l'attention</i>	attention	attirer	emporter
<i>fortifier la vérité</i>	vérité	renforcer, rétablir	fortifier
<i>faire un engagement</i>	engagement	prendre	faire
<i>prendre une maladie</i>	maladie	attraper	prendre
<i>donner un regard</i>	regard	lancer	donner
<i>créer une réunion</i>	réunion	organiser	créer
<i>risquer la stabilité de sa famille</i>	stabilité	compromettre, menacer	risquer
<i>abréger/brimer les droits civils</i>	droit	bafouer	abréger, brimer
<i>la tradition se maintient</i>	tradition	se perpétuer	se maintenir

<i>la rumeur se diffuse</i>	rumeur	se répandre, se propager	se diffuser
<i>une vaste générosité</i>	générosité	grand, immense	vaste
<i>son titre céleste</i>	titre	prestigieux	céleste
<i>ses traits visuels</i>	trait	physique	visuel
<i>une douleur insurmontable</i>	douleur	insupportable, insoutenable, intolérable	insurmontable
<i>un bonheur énorme</i>	bonheur	extraordinaire, formidable, immense	énorme
<i>gagner sa vie aisément</i>	gagner sa vie	bien	aisément

Tableau 3: Erreurs résultant de la méconnaissance des collocations

Dans le domaine des collocations, on trouve les cinq cas de figure suivants : (i) le verbe support est neutre mais ne fonctionne pas : **provoquer sa curiosité** (piquer), **protéger son temps** (préserver), **répéter l'expérience** (revivre), **faire (prendre) une décision** ; (ii) le verbe faisant partie d'une collocation dont le mot base est le sujet ne fonctionne pas : *L'amour n'est pas resté* (n'a pas duré) ; (iii) l'adjectif modificateur d'un nom ne fonctionne pas avec ce nom : *inspiration grandiose* (importante, incroyable), *un caractère puissant* (affirmé), *un chômage grave* (important, massif, considérable) ; (iv) l'adverbe modificateur d'un adjectif ou d'un verbe ne fonctionne pas : **spécifiquement** (particulièrement) *fier de ses légumes*, *apprécier extrêmement* (énormément) ; (v) la structure 'Nom de Nom' ne fonctionne pas : *des lignes de bronzage* (marques). Les types d'erreurs les plus courants dans le corpus correspondent aux trois premiers cas de figure. Les deux autres apparaissent de manière anecdotique.

Dans quelques cas, le mot choisi par l'apprenant est bien un collocatif du mot-base, mais ce choix parmi plusieurs collocatifs est cependant erroné en lien avec l'intention de communication. En voici quelques exemples : *des souvenirs à **garder** (conserver?) pour la vie, **lancer** (sortir) un album, un mystère à **dévoiler** (déchiffrer), Son regard me **donne** une impression (produit une impression sur moi)*. Cependant, dans notre corpus, la majorité des erreurs résultent de la sous-utilisation des collocations, et non d'erreurs dans le choix du collocatif. Cela peut sans doute s'expliquer par le fait que, comme les collocations ne sont pas maîtrisées, il y a peu d'emploi de « mauvais » collocatifs (collocatif existant mais sémantiquement incorrect), de même que de « bons » collocatifs (corrects sur tous les plans).

2.3.4. RELATION INCOMPATIBLE : restrictions de sélection

La catégorie RELATION INCOMPATIBLE peut être clarifiée grâce à la notion de restriction de sélection.

Nous avons vu dans la section précédente que les collocations se trouvaient dans une zone d'un continuum. Les restrictions de sélection sont également des contraintes qui pèsent sur la cooccurrence, mais cette fois au bout du continuum, sur les combinaisons libres et transparentes. Sur le plan sémantique, la plupart des verbes, noms ou adjectifs peuvent être considérés comme des prédicats (ce qui est dit sur quelque chose) possédant un certain nombre d'arguments (ce dont il est question), verbalisés ou non.

Comme le prédicat impose généralement des contraintes sémantiques sur ses arguments, les mots utilisés pour représenter des arguments doivent posséder certaines composantes de sens ; les verbes, de même que certains noms et adjectifs, imposent des restrictions de sélection à leurs arguments. Par exemple, si l'on considère le verbe *louer* dans le sens 'louer quelque chose à quelqu'un', le sujet et le complément indirect doivent désigner des humains ou des organisations d'humains ; ils possèdent donc la composante de sens « humain ».

Les erreurs classées RELATION INCOMPATIBLE résultent d'une incompatibilité entre le prédicat et son argument, dans le sens où les contraintes sémantiques imposées par le prédicat sur son argument sont violées. Dans le corpus étudié, c'est le plus souvent l'argument qui est adéquat par rapport à l'intention de communication de l'apprenant et le prédicat qu'il faudrait corriger. En effet, le prédicat choisi par l'apprenant possède des contraintes sémantiques sur ses arguments qui ne peuvent être vérifiées étant donné le mot choisi pour jouer le rôle de l'argument. Le mot choisi pour représenter le prédicat doit tenir compte des caractéristiques sémantiques des « arguments ».

Le prédicat peut correspondre à un verbe, un adjectif ou un nom. Voici quelques exemples tirés du corpus. Le prédicat est un verbe : *aider son investigation* (faciliter), l'objet-argument de *aider* devant posséder le trait « humain ». Le prédicat est un adjectif : *des contenus âgés* (anciens), l'argument de *âgé* devant posséder le trait « humain ». Le prédicat est un nom : *période d'agrandissement* (croissance),

l'argument étant un enfant, le nom prédicat doit être compatible avec un argument ayant le trait « humain ».

Cependant, c'est parfois le prédicat qui reflète l'intention de communication de l'apprenant et l'argument qui doit être corrigé, puisqu'il ne satisfait pas aux contraintes imposées par le prédicat. C'est le cas dans l'exemple *raccrocher la ligne* (le combiné), où le tout est utilisé à la place d'une de ses parties.

Il est parfois difficile de distinguer entre ce qui relève des critères sémantiques (RELATION INCOMPATIBLE) et ce qui relève de la collocation (COLLOCATION ATTENDUE). Ceci s'explique chez Bossé-Andrieu et Mareschal (1999) qui suggèrent que les collocations se trouvent dans la partie centrale d'un continuum transparent/fortuit–opaque/figé. Ces dimensions sont donc, dans certains cas, intriquées.

2.3.5. AFFIXE INADÉQUAT : famille de mots

La catégorie AFFIXE INADÉQUAT renvoie à la notion de famille de mots. Une partie des règles morphologiques permettent de former des mots par des processus d'affixation, soit l'ajout d'un préfixe ou d'un suffixe, ou les deux à la fois, à une racine. Comme le signale Bogaards (1994), ces procédés sont en partie irréguliers et non prévisibles. Il fournit plusieurs raisons pour lesquelles la dérivation présente peu de transparence pour

l'apprenant : la pluralité de sens de certains affixes, la diversité des affixes pouvant illustrer un même sens, l'absence de relation biunivoque entre un mot et un de ses dérivés (les sens des unités lexicales étant répartis différemment), l'existence de « faux dérivés », et enfin des écarts de sens trop importants entre un mot et son dérivé. Il conclut en mentionnant que si la connaissance de règles morphologiques peut aider l'apprenant dans des tâches réceptives comme la lecture, elles sont de peu d'utilité en production.

Les erreurs classées AFFIXE INADÉQUAT mettent en évidence une connaissance approximative des modifications de sens apportées par différents préfixes et suffixes. La forme de l'affixe est connue, les processus de dérivation sont connus, mais leur sens n'est pas maîtrisé. Ces erreurs reflètent des erreurs formelles qui ont des conséquences sur le sens véhiculé par le mot choisi. Elles mettent donc en évidence un rapport forme sens. Différentes catégories sont concernées : les noms, les verbes et les adjectifs. En voici quelques exemples. Des verbes : *se **rétablir** quelque part* (s'établir), *pour **apparaître** plus grand* (paraître). Des adjectifs : *la maison **avoisinante*** (voisine), *le sexe **opposant*** (opposé). Des noms : *ses **écritures*** (écrits), *la **gagnante** du temps* (gain). Pour les verbes, les difficultés semblent surtout liées à l'emploi des préfixes [a-] et [re-].

2.3.6. Récapitulatif des notions et liens entre les classes

Les différents sens d'un mot polysémique sont représentés par des unités lexicales liées entre elles par le sens. Ces unités lexicales entrent dans différents schémas de sous-catégorisation représentant différents sens. Lorsque l'apprenant utilise une construction associée à une autre unité lexicale que celle dont il souhaiterait transmettre le sens, alors il commet une erreur de CONSTRUCTION INCOMPATIBLE.

À un mot polysémique peut correspondre un concept, les unités lexicales associées appartenant éventuellement à différents champs sémantiques. Si l'apprenant utilise un mot entrant dans le même champ sémantique que le mot cible, mais s'en différenciant par une composante de sens, il commettra une erreur de CONCEPT INADÉQUAT.

Les mots présentent des contraintes dans leurs cooccurrences. Ces contraintes sont d'au moins deux types : un mot demande un collocatif ou les relations sémantiques entre les mots connaissent des restrictions de sélection. Si l'apprenant ne respecte pas le premier type de contrainte, il commettra une erreur de COLLOCATION ATTENDUE, tandis que s'il ne respecte pas le second type de contrainte, il commettra une erreur de RELATION INCOMPATIBLE.

Les mots ayant une racine commune peuvent être regroupés en familles de mots se différenciant par des affixes. Si l'apprenant maîtrise mal les conséquences sémantiques de l'utilisation des affixes, il risque de commettre une erreur d'AFFIXE INADÉQUAT.

Nous pouvons regrouper les classes d'erreurs en fonction de si elles concernent le mot lui-même ou le lien du mot avec d'autres mots dans la phrase.

Certaines erreurs sont « internes » au mot, dans le sens où le choix du mot ne dépend pas d'autres mots présents dans la phrase. Il s'agit d'AFFIXE INADÉQUAT (formation de mots, morphologie en lien avec le sens), de CONSTRUCTION INCOMPATIBLE (régime en lien avec le sens) et de CONCEPT INADÉQUAT (composantes de sens). Ce sont réellement des erreurs de sens, le mot choisi n'exprimant pas le sens recherché par l'apprenant, même si le mot choisi peut exprimer un sens assez proche de celui du mot attendu. Le mot choisi par l'apprenant et le mot attendu sont en relation paradigmatique : mots de la même famille pour AFFIXE INADÉQUAT, mots pouvant être quasi-synonymes dans d'autres contextes, ou proches sémantiquement pour CONCEPT INADÉQUAT. Cependant, pour CONSTRUCTION INCOMPATIBLE, le mot choisi et le mot attendu sont le même mot, mais des unités lexicales différentes.

D'autres erreurs sont de nature « contextuelle », dans le sens où le mot choisi dépend d'un autre mot présent dans la phrase. Il s'agit de RELATION INCOMPATIBLE (lien entre un prédicat et son argument) et COLLOCATION ATTENDUE (lien entre un mot-base et son collocatif). Le plus souvent, le mot choisi et le mot attendu sont des quasi-

synonymes, mais le lien entre le mot choisi et le mot du contexte ne fonctionne pas. Des contraintes de nature syntagmatique rendent peu acceptable l'utilisation du mot choisi.

Il est possible de voir les classes d'erreurs « internes » au mot ou « contextuelles » sous l'angle d'un ensemble de contraintes dans le processus de choix lexical. En effet, si l'on devait fournir une procédure à l'apprenant pour l'aider dans son choix lexical, il faudrait tenir compte des contraintes liées aux autres mots présents (collocations et restrictions de sélection) pour procéder au choix purement sémantique.

Du point de vue de l'apprenant, le choix lexical s'effectue éventuellement parmi plusieurs mots concurrents. Des mots s'opposant sur certains traits au sein d'un champ sémantique, différents collocatifs possibles pour un mot-base, des mots quasi-synonymes présentant différentes contraintes de restriction de sélection, des mots se différenciant par des affixes, différentes constructions pour un même mot. Tous ces aspects interviennent en parallèle dans le choix lexical. Nous comprenons alors l'importance du développement des réseaux lexicaux chez les apprenants de niveau intermédiaire-avancé. Le choix lexical s'opère entre des ensembles de mots de sens proche soumis à différentes contraintes, conceptuelles et syntagmatiques.

3. VERS UNE MÉTHODOLOGIE DE L'ENSEIGNEMENT DU CHOIX LEXICAL

L'analyse du corpus a fait émerger différents types d'erreurs de choix lexical sur le plan sémantique. Ces classes d'erreurs, vues sous un angle notionnel, nous ont permis de

mieux comprendre ce qui était en jeu dans le choix lexical et comment ces types d'erreurs s'articulaient entre eux. Nous nous interrogeons maintenant sur la manière d'aider les apprenants dans cette tâche ardue de choix lexical. Après avoir rappelé les limites d'un apprentissage implicite, sans enseignement, nous proposons une démarche en trois étapes (3.1.). Nous allons présenter et justifier chacune de ces étapes, pour ensuite fournir des exemples d'activités pouvant être mises en œuvre dans le cadre de chacune des étapes (sections 3.2. à 3.4.).

3.1. Présentation générale de la séquence didactique

Dans cette section, nous nous positionnons en faveur d'un apprentissage explicite (3.1.1.) et en réseau (3.2.2.) de la sémantique lexicale et présentons les trois étapes de la séquence didactique que nous avons développée (3.1.3.) afin d'apporter des éléments de solution au problème de choix lexical.

3.1.1. Pour un apprentissage explicite et en contexte du sens lexical

L'apprentissage du vocabulaire, au niveau intermédiaire-avancé, s'effectue en partie de manière implicite, principalement grâce à la lecture. Nous avons déjà vu, à la section 1.4., que les connaissances lexicales se développaient avec le temps, mais pas d'une manière satisfaisante par rapport aux besoins des apprenants (Schmitt, 1998 ; Haastруп et Henriksen, 2000), ce qui se traduit par des difficultés dans des tâches d'écriture

(Maignashca, 1984). Dans l'apprentissage du vocabulaire, des apprentissages de type implicite ou explicite se complètent. Nous développons ici un peu plus ces deux aspects.

L'apprentissage implicite du vocabulaire est souvent associé à la lecture où, par conséquent, un contexte est associé aux mots. Il s'oppose à l'apprentissage explicite, qui est souvent, mais pas toujours, associé à la mémorisation de listes de mots pour lesquels une traduction est éventuellement donnée, et qui s'opère donc hors contexte. Ainsi, pour ce qui est de l'apprentissage du vocabulaire, différentes approches sont possibles. Dans une synthèse de travaux sur l'apprentissage du vocabulaire, Coady (1997) présente les différentes approches sous la forme d'un continuum avec, à un bout, l'apprentissage uniquement en contexte grâce à la lecture et à l'autre bout, l'apprentissage par des activités en classe mêlant tout type d'activités de vocabulaire, communicatives ou non. Entre ces deux extrêmes, on trouve tout d'abord l'apprentissage en contexte accompagné de l'enseignement de stratégies d'apprentissage pour aider l'apprenant dans sa gestion du contexte (au niveau intermédiaire-avancé). Toujours sur ce continuum, on trouve également l'enseignement direct des mots considérés comme les plus importants, c'est-à-dire les mots les plus fréquents (au niveau élémentaire), associé à l'apprentissage en contexte sous forme de lecture de livres simplifiés par exemple.

Il est certain que l'apprentissage implicite du vocabulaire présente un intérêt du fait de la présence d'un contexte. En effet, suite à un apprentissage par des listes de mots, les apprenants peuvent rencontrer des difficultés lors de la réutilisation du vocabulaire, qui

nécessite une mise en contexte. Le contexte pourrait donc jouer un rôle important dans l'acquisition des sens lexicaux. Dans ce sens, Jiang (2000) préconise des apprentissages contextuels, par la lecture par exemple. Selon lui, ceux-ci seraient préférables à des apprentissages de mots avec leur traduction hors contexte, car ils ne nécessiteraient pas que l'apprenant s'appuie sur la L1. Cela favoriserait l'atteinte du stade dans lequel les formes de la L2 sont bien associées à des concepts de la L2, et non à des concepts de la L1, c'est-à-dire du stade 3 défini en 1.4.

Cependant, l'apprentissage à partir du contexte seul connaît des limites. Comme le fait remarquer Laufer (2005), c'est plus une stratégie de lecture que d'apprentissage. Sökmen (1997) souligne également que l'apprentissage implicite permet éventuellement la compréhension si l'apprenant devine le sens du mot, mais pas nécessairement la mémorisation, puisque ce n'est pas l'objectif de la tâche. Ainsi, d'après elle, l'apprentissage explicite du vocabulaire devrait être mené en parallèle avec l'apprentissage implicite (en devinant le sens du mot dans des activités de lecture par exemple). Par ailleurs, à un niveau avancé, les sens non maîtrisés apparaissent peu fréquemment (Arnaud et Savignon, 1997) et on peut donc difficilement se baser sur une stratégie d'apprentissage basée sur l'exposition à l'intrant. Enfin, de nombreuses rencontres avec un mot sont nécessaires à son intégration à long terme dans le vocabulaire, ce qui nécessite des quantités de lecture considérables, aspect peu réaliste dans la plupart des contextes (Laufer, 2005). De tous ces éléments il découle qu'un apprentissage explicite est nécessaire, et que celui-ci doit s'effectuer avec des mots en contexte.

L'apprentissage explicite du vocabulaire est d'autant plus nécessaire que l'objectif visé est l'acquisition du sens lexical et que les apprenants sont de niveau intermédiaire-avancé. Ellis (1997) distingue les processus sous-jacents à l'acquisition de la forme de ceux de l'acquisition du sens. Il explique que les processus menant à l'acquisition de la forme d'un mot ou d'une séquence sont implicites et favorisés par la répétition, tandis que ceux qui mènent à l'acquisition du sens et des relations conceptuelles doivent être explicites et sont facilités par un enseignement. Ainsi, Nation (2001) recommande l'utilisation d'activités favorisant l'implicite et la répétition pour l'apprentissage de la forme et l'utilisation d'activités faisant intervenir des connaissances explicites, ainsi que la profondeur du traitement, pour l'apprentissage du sens. L'apprentissage du sens nécessiterait des stratégies d'apprentissage plus explicites et serait donc favorisé par un enseignement explicite des aspects sémantiques du lexique. Comme, par ailleurs, au niveau intermédiaire-avancé, l'apprentissage du vocabulaire passe principalement par l'acquisition de nouvelles relations forme-sens en même temps qu'il porte sur la sémantisation et l'organisation du lexique, il est particulièrement important qu'il se déroule avec une composante explicite.

3.1.2. Pour le développement de réseaux d'association

Nation et Newton (1997) établissent une distinction entre l'apprentissage du vocabulaire et l'accès lexical en production¹². Le premier s'effectue principalement par des activités favorisant la mémorisation. Le second se travaille par deux types d'activités. Les activités de développement de réseaux et d'associations entre les mots, à la fois syntagmatiques (activités de collocation, cartes sémantiques, dictées) et paradigmatisées (activités de mise en relation, de classification) ont pour objectif d'augmenter le nombre de voies d'accès à un mot. Les activités d'aisance à communiquer ont pour objectif de développer la rapidité de l'accès lexical. La séquence didactique que nous proposons relève du premier aspect, le développement des réseaux d'association, en y ajoutant un travail sur les mots eux-mêmes, que ce soit à travers leur polysémie ou leur morphologie. En effet, nous considérons que les différentes unités lexicales d'un mot polysémique sont reliées sémantiquement et qu'il peut être utile de prendre conscience de ces liens. Nous ne travaillerons donc pas sur l'aisance à communiquer, qui concerne davantage l'oral, ni sur l'apprentissage du vocabulaire par des activités de mémorisation, mais sur le développement des réseaux lexicaux et conceptuels.

Les dangers de l'apprentissage de synonymes (Martin, 1984 ; Nation et Newton, 1997) ou de mots appartenant à un même champ sémantique (Waring, 1997 ; Finkbeiner et Nicol, 2003), à savoir une mémorisation moins bonne à cause d'interférences

12 En 1.1., nous avons défini le choix lexical comme le résultat de l'accès lexical en production.

sémantiques, ont été relevés par plusieurs auteurs. Toutefois, ces mises en garde s'appliquent à l'apprentissage de mots nouveaux, et non au raffinement de la connaissance lexicale. Or, au niveau intermédiaire-avancé, nous pouvons supposer qu'une partie des liens formels sont connus et que les efforts d'apprentissage peuvent porter sur les concepts et le développement des liens entre les unités lexicales. L'apprentissage de la forme et l'acquisition des premiers liens forme-sens s'effectuent principalement au niveau débutant, même si l'acquisition de ces liens continue tout au long de l'apprentissage. Dans les lexiques d'apprenants de niveau intermédiaire-avancé, les liens sémantiques se développent, ce qui différencie ces lexiques de ceux d'apprenants de niveau débutant, où les liens sont en grande partie formels (Schmitt, 1998 ; Nation, 2001).

3.1.3. Les trois étapes de la séquence didactique pour le choix lexical

La séquence didactique que nous proposons établit une progression en trois étapes. La première étape consiste à faire prendre conscience à l'apprenant de ce qu'implique le choix lexical. Le choix lexical fait en effet intervenir différentes dimensions le faisant naviguer dans un espace de décisions à prendre. Nous présentons cette étape en 3.2. La deuxième étape vise à mettre en évidence des problèmes de choix lexical rencontrés par l'apprenant dans une tâche communicative de production écrite et à l'amener à les résoudre. L'apprenant sera outillé conceptuellement et méthodologiquement, la résolution de problème se déroulant dans un espace « aidé », « balisé », constructif. La

démarche proposée lui servira dans le cadre d'une autocorrection, mais pourrait se transposer à la production également. Nous explicitons cette étape en 3.3. La troisième et dernière étape vise l'intégration des nouvelles connaissances lexicales par la construction de réseaux sémantiques. Un corrélat en est l'enrichissement du vocabulaire. Nous exposons cette étape en 3.4. Pour chacune de ces trois étapes, nous précisons (i) les objectifs, (ii) les justifications théoriques sous-jacentes, (iii) les moyens utilisés et dans certains cas des exemples d'activités pouvant être utilisées dans ce cadre.

3.2. Faire voir l'« espace de décision »

3.2.1. Présentation

L'objectif de cette première étape de la séquence est de faire voir aux apprenants l'ensemble des paramètres intervenant dans le choix lexical, celui-ci les entraînant dans un espace de décisions à prendre où les mots sont soumis à différentes contraintes. Des notions leur seront présentées : ce sont les notions qui permettent de comprendre les classes d'erreurs et qui ont été décrites en 2.3. Ces notions sont la polysémie, les schémas de sous-catégorisation, les composantes de sens, les collocations, les restrictions de sélection, les concepts, les champs sémantiques, la synonymie et les familles de mots. Les apprenants pourraient ainsi être outillés d'un point de vue notionnel.

L'étape « espace de décision » trouve une justification dans la notion de conscience linguistique (« awareness »). En particulier, il est important que les apprenants prennent conscience du fait que les associations mot-concept et la polysémie ne sont pas nécessairement équivalents en L1 et en L2, et que les mots de la L2 sont soumis à des contraintes contextuelles de diverses natures spécifiques à la L2. La prise de conscience de ces phénomènes pourrait amener les apprenants à les remarquer plus facilement lors d'activités de lecture par exemple, et ainsi à enrichir leur réseau lexical, implicitement ou suite à des recherches. Cela pourrait se répercuter positivement dans des activités de production. En outre, cette prise de conscience prépare l'apprenant aux deux étapes suivantes de la méthodologie.

Plusieurs auteurs abondent dans le sens d'une stratégie de prise de conscience en ce qui concerne le lexique. D'après Chiu (2009), la transition de l'apprenant vers un stade d'acquisition où les formes de la L2 sont associées aux concepts de la L2, et non de la L1, pourrait être facilitée, entre autres, par la prise de conscience du fait que le concept sous-jacent au mot de la L2 ne correspond pas à celui de sa traduction dans la L1. Lewis (1997), dans son approche lexicale, préconise, plutôt qu'un enseignement, des activités de prise de conscience en classe, de nature réceptive. Ces activités sont implicitement, ou explicitement, basées sur différentes notions comme celles de collocation, de lien paradigmatique entre les mots, de polysémie.

Cette première étape de la méthodologie donne lieu à une séquence d'activités lexicales permettant d'aboutir (i) à la mise en valeur et à la compréhension des différentes notions et (ii) à la compréhension des liens entre ces notions. Le « matériel lexical » sera tiré ou adapté du corpus, en reprenant les types d'erreurs les plus fréquents et les champs sémantiques qui posent le plus de difficultés. Par la même occasion, cela permet d'enrichir le lexique de l'apprenant.

3.2.2. Activités pour l'espace de décision

Une première série d'activités a pour objectif de faire comprendre aux apprenants les paramètres qui interviennent dans le choix lexical. Il s'agit d'une découverte intuitive, suivie de la formulation d'une « règle » lexicale.

- un mot dépend de ses cooccurrents (activité 1.1)
- un prédicat dépend de ses arguments (activité 1.2)
- des affixes différents modifient le sens différemment (activité 1.3)
- la polysémie (les concepts) affecte l'interchangeabilité des mots au sein d'un champ sémantique (activité 1.4)
- le sens varie selon la construction syntaxique (structure d'arguments) (activité 1.5)

Il s'agit ici d'activités visant une prise de conscience, et non la production. L'objectif n'est pas d'amener les apprenants à utiliser ou à enrichir leur vocabulaire, mais plutôt de leur faire saisir les contraintes apportées par les différentes notions en les faisant réfléchir

sur les exemples proposés. Les activités reposent ainsi sur des exemples d'énoncés à partir desquels les apprenants sont amenés à prendre des décisions et à argumenter leurs choix. Cela pourrait les conduire à verbaliser des fragments de « règles » lexicales. Les exemples peuvent exploiter les erreurs rapportées dans l'annexe 1. Notons toutefois que, même si le but visé n'est pas l'enrichissement du vocabulaire, ces activités pourraient y contribuer indirectement.

Nous présentons ci-dessous des formats d'activité, que l'enseignant devrait compléter. Par exemple, il faut remplacer le mot [contexte] par une phrase réelle.

Activité 1.1. : Sélectionnez le mot approprié.

Exemples

[Contexte] ----- l'attention (attire, prend, emporte)

[Contexte] ----- l'attention (capte, autres cooccurrents possibles mais un seul va fonctionner dans ce contexte, problème de sens)

Ajouter 3 ou 4 autres exemples de contextes pour des mots base, chacun associé à un collocatif possible et deux autres mots synonymes du collocatif mais non attestés avec le mot base.

Par groupe de 2 à 4 étudiants, discutez de vos choix et justifiez-les.

Pendant la mise en commun, l'enseignant présentera le principe des collocations, des mots qui vont ensemble.

Activité 1.2. : Sélectionnez le mot approprié.

[Contexte] ----- (âgé, vieux, ancien)

Ajouter 3 ou 4 autres exemples de contextes comprenant des « arguments » pour lesquels on fera contraster des prédicats, dont l'un au moins ne fonctionne pas. Dans les différents exemples, on pourra choisir des prédicats de différentes catégories grammaticales (adjectif, nom, verbe).

Par groupe de 2 à 4 étudiants, discutez de vos choix et justifiez-les.

Pendant la mise en commun, l'enseignant présentera le principe des restrictions de sélection.

Activité 1.3 : Expliquez les différences de sens.

[Contexte] --- changeant vs [Contexte] --- changeable

Ajouter 4 ou 5 autres exemples comprenant des mots de différentes catégories grammaticales et différents affixes. Chaque exemple permettra d'opposer les sens de deux affixes différents pour une même racine.

Par groupe de 2 à 4 étudiants, discutez de vos explications.

Pendant la mise en commun, l'enseignant sensibilisera les apprenants à l'influence des affixes sur le sens (nuances).

Activité 1.4 : Parmi cette liste de mots, lesquels sont synonymes ?

Dresser une liste de quelques paires de synonymes, sur deux colonnes.

Parmi ces phrases, dites lesquelles sont acceptables. Si elles ne le sont pas, expliquez pourquoi.

Faire plusieurs listes de 3 phrases, où A et B sont synonymes : [Contexte] :

A et B sont interchangeables, [Contexte] : B fonctionne mais pas A.

Par groupe de 2 à 4 étudiants, discutez de vos explications.

Pendant la mise en commun, l'enseignant sensibilisera les apprenants à la polysémie, aux champs sémantiques et à la différenciation de mots par composantes de sens.

Activité 1.5 : Expliquez les différences de sens.

[Contexte] --- compter faire vs [Contexte] --- compter sur

Ajouter 3 ou 4 autres exemples où plusieurs constructions d'un même verbe mènent à des sens différents.

Par groupe de 2 à 4 étudiants, discutez de vos explications.

Pendant la mise en commun, l'enseignant sensibilisera les apprenants à la polysémie et à l'influence de la construction sur le sens.

Une deuxième série d'activités a pour objectif de proposer une séquence, sous forme de liste de vérifications à effectuer, pour le choix lexical. Nous avons vu en 2.3.6. que certaines classes d'erreurs étaient internes au mot tandis que d'autres étaient en lien avec le contexte immédiat. Ces deux aspects se retrouvent dans la distinction Concept vs Contexte présentée dans l'activité 2 ci-dessous.

Activité 2

Questions à se poser lors d'une tâche de choix lexical, sous forme d'un ensemble de conditions parallèles.

- Concept
 - A-t-on le bon concept ? Si non, chercher un synonyme ou un mot de sens proche.
 - Le mot comprend-il des affixes ? Si oui, vérifier que c'est bien le bon, qu'il donne le bon sens.
 - Est-ce que la construction choisie reflète bien le sens que l'apprenant souhaite véhiculer ?
- Contexte
 - Y a-t-il un mot-base qui appellerait un collocatif ? Si oui, choisir le bon collocatif.
 - Y a-t-il une relation argument-prédictat qui imposerait des restrictions de

sélection ? Si oui, choisir le bon prédicat (ou argument).

3.3. Autocorrection : mode d'emploi

Une fois l'apprenant familiarisé avec les notions intervenant dans le choix lexical ainsi qu'avec l'« espace de décision » au sein duquel il peut naviguer, il lui est possible de « passer à l'action ». Dans cette section, nous présentons une méthode basée sur l'autocorrection (3.3.1.), ainsi que le matériel didactique élaboré dans cet objectif (3.3.2).

3.3.1. Présentation

Cette deuxième étape de la séquence didactique a un double objectif. Le premier est, à partir d'erreurs lexicales dans des productions écrites d'apprenants, de fournir un diagnostic le plus précis possible afin d'orienter l'apprenant vers une correction possible, de lui donner du « feedback » sur sa production, en soulignant et en codifiant des cas de choix lexical problématiques. Le second est de donner les moyens à l'apprenant de corriger ses erreurs lexicales, étant donné qu'il n'existe pas de « règles » lexicales comme il existe des « règles » grammaticales. À ce sujet, Zhou (2009) rapporte que les apprenants signalent des difficultés à estimer si le choix lexical qu'ils ont effectué est correct ou non, et qu'ils se sentent dépendants de l'aide d'un professeur ou d'un locuteur

natif. Cela nous conduit à vouloir placer l'apprenant en situation d'analyse, d'ajustement, de prise de décision, et à l'outiller méthodologiquement pour qu'il puisse développer des savoir-faire dans le cadre de l'autocorrection qui soient transposables dans le cadre de la production.

L'idée théorique sous-jacente est que la précision lexicale peut être atteinte par une série d'approximations successives, de restructuration des concepts et des réseaux. Cette réorganisation du lexique est, selon Sonayia (1991), stimulée entre autres par l'exposition à la langue et la correction. Ici, il s'agit plutôt d'autocorrection suite à une signalisation d'erreur. Comme le travail sur le sens est préférablement explicite (Ellis, 1997 ; Nation, 2001), la mise en évidence d'erreurs, leur catégorisation et un travail subséquent de l'apprenant pour résoudre le problème de correction lexicale devraient être conducteurs d'apprentissage. En effet, l'un des axes retenus par Sökmen (1997) pour faciliter la mémorisation du vocabulaire était d'avoir un niveau de traitement de l'information en profondeur. Toutefois pour que la tâche d'autocorrection mène à de l'apprentissage, il faut que le problème soit posé de manière à ce que l'apprenant soit en mesure de le résoudre, qu'il ait les outils cognitifs nécessaires. L'apprenant entre donc dans un espace de résolution de problème où il pourra mettre en application sa compréhension de l'espace de décision présenté précédemment. Par ailleurs, Hemchua et Schmitt (2006) soulignent que l'utilisation d'une taxonomie d'erreurs en classe, donc de codes, permettrait de développer la prise de conscience des erreurs lexicales chez les apprenants.

Cette deuxième étape de la méthodologie prendra place dans le cadre d'une activité de production écrite. Des codes d'erreurs avec leur définition et des exemples seront fournis aux apprenants. Chaque code correspondra à une catégorie d'erreur issue de l'analyse du corpus et présentée dans la section 2.3. Ces codes référeront aux notions présentées à la première étape de la méthodologie, en 3.2. Ils seront appliqués par l'enseignant à la production écrite de l'apprenant qui devra ensuite proposer une correction de ses erreurs et les soumettre à l'enseignant qui validera ou sinon suggérera une correction adéquate. Ce travail de l'apprenant est individuel.

Afin que l'apprenant soit en mesure de comprendre comment, concrètement, corriger son erreur, des activités pédagogiques seront proposées. Elles visent à ce que l'apprenant se familiarise avec la démarche d'autocorrection, en le guidant sur plusieurs types d'auto-correction, au moins un par code. Ce seront donc des cas d'autocorrection à résoudre avec guidage, afin de bien poser l'espace de résolution de problème en s'appuyant sur les notions vues en 3.2. L'apprenant sera amené à utiliser des outils, par exemple un dictionnaire de collocations (ou un corpus) pour les codes COLLOCATION ATTENDUE, à faire des recherches dans les dictionnaires et les corpus pour les CONCEPT INADÉQUAT, etc. Cela nécessite d'être guidé puisque les apprenants peuvent rencontrer des difficultés à utiliser convenablement des dictionnaires (Coady, 1997).

La méthode d'autocorrection pour les productions écrites est déjà utilisée dans le cours dont est extrait le corpus, de même que dans l'enseignement collégial et certaines

universités. Toutefois, comme les codes utilisés concernent toutes les erreurs langagières, et pas seulement les erreurs lexicales, les codes liés aux erreurs lexicales sont en nombre assez limité et ne permettent pas toujours à l'apprenant de savoir comment corriger son erreur. Nous proposons des codes d'erreurs lexicales détaillés qui aiguillent l'apprenant dans la recherche d'une solution. Ce sont les codes correspondant aux catégories d'erreurs présentées en 2.3. De plus, nous proposons une procédure de résolution de problèmes d'autocorrection. Associés à la première étape de la méthodologie qui présente l'espace de décision, ils devraient avoir une valeur explicative de l'erreur, orientant l'apprenant dans sa recherche d'une solution. Par ailleurs, nous pouvons noter que le travail sur les codes permet de travailler sur les réseaux. En effet, le travail sur les codes peut amener à la compréhension de la différence sémantique entre deux mots, le mot choisi et le mot attendu (relations paradigmiques) et au rapport entre des mots sur l'axe syntagmatique (avec le code COLLOCATION ATTENDUE par exemple).

3.3.2. Activités pour l'espace de résolution de problème

Tout d'abord, il s'agit de définir « pédagogiquement » les codes et de les illustrer par des exemples parlants, comme dans le tableau 4 ci-dessous.

CATÉGORIE	DÉFINITION ET EXEMPLES
RELATION INCOMPATIBLE	<p>L'association argument-prédicat ne fonctionne pas pour des raisons sémantiques, de restrictions de sélection.</p> <p>Exemples : <i>une maison âgée</i> (<i>âgé</i> s'applique à un être humain), <i>le teint de ses cheveux</i> (le <i>teint</i> est associé au visage)</p>
CONCEPT INADÉQUAT	<p>Il existe des éléments de sens communs au mot utilisé et au mot attendu, qui sont des quasi-synonymes se différenciant par certaines composantes de sens.</p> <p>Exemples : <i>mettre fin à un rapport</i> (relation/amitié), <i>Il est difficile de trancher dans cette industrie</i> (se démarquer)</p>
AFFIXE INADÉQUAT	<p>La forme utilisée se différencie de la forme correcte par un changement d'affixe. Le sens est ainsi modifié, même si un élément de sens est commun aux deux formes. Les mots appartiennent à la même famille.</p> <p>Exemples : <i>Elle se renfermait dans sa chambre</i> (s'enfermait), <i>des sentiments changeables</i> (changeants)</p>
CONSTRUCTION INCOMPATIBLE	<p>Le mot est polysémique, différentes constructions étant associées à différents sens. La construction dans laquelle le mot est employé ne fonctionne pas avec ce sens.</p> <p>Exemples : <i>elle n'a pas réalisé le camion</i> (réalisé qu'il y avait un camion, remarqué/vu le camion), <i>filer les ciels gris</i> (fuir)</p>

COLLOCATION ATTENDUE	<p>Il existe avec un mot du contexte une collocation que l'on s'attendrait à voir utiliser. Le sens du mot choisi est cependant correct.</p> <p>Exemples : apporter de la compassion envers quelqu'un (faire preuve de), emporter l'attention de quelqu'un (attirer), faire des décisions (prendre)</p>
-----------------------------	--

Tableau 4 : Présentation des codes : définitions et exemples

Une fois les codes présentés, il est nécessaire de proposer une méthode pour corriger les erreurs. L'objectif de cette série d'activités est de guider la correction en indiquant les questions à se poser, en mentionnant les outils à utiliser. Nous en donnons deux exemples, à partir des codes COLLOCATION ATTENDUE et RELATION INCOMPATIBLE.

Exemple d'activité/procédure : code COLLOCATION ATTENDUE

- Dans le contexte, par rapport au mot souligné comme erroné, quel est le mot base ?
- Rechercher dans un dictionnaire de cooccurrences les collocatifs possibles pour ce mot-base.
- Parmi les collocatifs possibles, lequel ou lesquels rend(ent) compte du sens recherché ?
- Vous pouvez aussi vérifier si la collocation trouvée se retrouve à l'aide un concordancier.

Exemple d'activité/procédure : code RELATION INCOMPATIBLE

- Dans le contexte, par rapport au mot souligné comme erroné, quels sont les mots qui lui sont liés sémantiquement ? Dont il peut être le prédicat ?
Dont il peut être l'argument ?
- Avec quel mot du contexte le mot erroné est-il incompatible ?
- Rechercher la définition du prédicat dans un dictionnaire unilingue ; y a-t-il des indications sur les caractéristiques de ses arguments sous forme de traits (par ex. humain) ou d'autres spécifications ? Est-il possible d'inférer certaines informations sur les arguments à partir des exemples donnés ?
- Quels mots connaissez-vous qui véhiculent le même sens que le mot erroné, mais qui auraient d'autres restrictions de sélection ? Si vous n'en connaissez pas d'autres, aidez-vous d'un dictionnaire de synonymes.
- Vérifier la définition du mot choisi dans un dictionnaire unilingue. Le mot que vous avez sélectionné est-il compatible avec ses arguments dans votre contexte ?

3.4. Appréhension de la polysémie et construction de réseaux

À travers les deux étapes décrites dans les sections précédentes (3.2. et 3.3.), l'apprenant peut comprendre la « problématique » du choix lexical et s'entraîner tout en étant guidé. Dans cette section, nous présentons les objectifs de la troisième étape de la séquence didactique ainsi que les moyens envisageables pour travailler sur les réseaux sémantiques et la polysémie.

3.4.1. Présentation

Les objectifs de cette troisième étape de la séquence didactique sont (i) que les apprenants puissent intégrer les nouvelles connaissances acquises par l'autocorrection de façon durable dans leur vocabulaire et qu'ils développent ainsi leurs réseaux lexicaux et (ii) qu'ils développent une autonomie dans l'enrichissement de leur vocabulaire. Ce dernier objectif se base sur le constat d'une impossibilité d'explicitier tout le vocabulaire, en salle de classe ou à distance, mis à part dans un dictionnaire pédagogique.

L'idée théorique sous-jacente est que le choix lexical peut être facilité une fois que les réseaux lexicaux de l'apprenant (similitudes et différenciation entre les mots, relations syntagmatiques et paradigmatisques, etc.) se sont enrichis (Saussure, 1916/1995 ; Meara 1996 ; Nation et Newton, 1997 ; Henriksen, 1999 ; Haastrup et Henriksen, 2000). En 1.4., nous avons vu que deux éléments en particulier peuvent contraindre le choix lexical

chez les apprenants : la non-correspondance, ou non-équivalence, des concepts en L1 et en L2, alors que les apprenants associent des concepts de la L1 à des formes de la L2, et une taille et une organisation moindres des réseaux lexicaux de la L2 chez les apprenants. Il est donc important de permettre aux apprenants de travailler sur la restructuration des concepts afin de dévoiler les concepts de la L2, ainsi que sur le renforcement des liens paradigmatiques et syntagmatiques entre les mots. Tout ce qui peut contribuer à développer ces deux aspects sera particulièrement utile. Cela passe principalement par un travail sur la polysémie en L2, la distinction de mots de sens proches et plus généralement sur les liens entre les mots de la L2. Les notions de concepts et de réseaux lexicaux sont d'ailleurs étroitement entrelacés. Certains auteurs associent la justesse de la représentation sémantique chez l'apprenant à la maîtrise des frontières des concepts (Ijaz, 1986) ou à la maîtrise des relations (ou distinctions) entre des mots de sens proche (Sonayia, 1991). L'identification des relations entre les mots permet en partie de clarifier les frontières des concepts, et vice versa. L'identification d'un dénominateur commun aux différents sens d'un mot polysémique doit aussi aider.

Nous allons maintenant examiner comment il est possible, par un travail didactique sur les réseaux et la polysémie de faciliter le choix lexical. Nous présentons tout d'abord des approches qui s'attaquent principalement au phénomène de la polysémie, pour ensuite nous attacher à des approches travaillant sur les réseaux lexicaux.

Connaissant l'influence de la langue maternelle sur la définition des concepts en langue seconde, comment est-il possible d'enseigner les nuances de sens ? Différents moyens

de travailler sur les concepts, basés sur des modèles théoriques distincts, sont proposés. Ils consistent principalement en des activités de réception et sont donc à exploiter pour développer la conscience linguistique de l'apprenant. Les deux études présentées ci-dessous mettent en évidence l'intérêt de connecter les différents sens d'un mot polysémique pour la compréhension, la rétention et l'usage. Elles portent toutefois sur l'apprentissage de mots nouveaux, en contexte, et non sur l'approfondissement de la connaissance sémantique de mots déjà connus, qui prévaut dans ce mémoire.

Verspoor et Lowie (2003) ont pour thèse que donner à l'apprenant le sens principal, « core meaning », d'un mot nouveau permet de mieux deviner un sens périphérique de ce mot polysémique rencontré dans un contexte par la suite et de mieux le mémoriser. Des liens existent entre le sens principal et les sens périphériques, puis entre ces sens périphériques et d'autres sens davantage périphériques. Ceci est particulièrement important quand les sens centraux et périphériques ne sont pas distribués de la même manière que dans la langue maternelle ; cela rejoint Ijaz (1986) et Jiang (2000). Verspoor et Lowie (2003) suggèrent de présenter aux apprenants comme stratégie d'apprentissage la recherche du sens central (« core ») du mot, quel que soit le sens de l'unité lexicale rencontrée par l'apprenant, ainsi que la recherche des liens entre le sens central et les autres sens. Cela est sous-tendu par l'importance qu'il y a à faire comprendre qu'il n'y a pas un sens unique associé à une forme. Les auteurs encouragent également à la présentation des différents sens des mots polysémiques de façon à ce que les liens entre ces différents sens, d'une part, et ces sens et le sens central, d'autre part, soient découverts plus facilement. Cela permet à l'apprenant

d'établir des liens plus élaborés entre les sens et d'enrichir le réseau lexical dont il dispose.

Avec un objectif similaire à celui de Verspoor et Lowie (2003), Adachi (2003) se réfère à la notion de schéma en linguistique cognitive. Elle s'intéresse à la fois aux extensions, c'est-à-dire aux différents sens du mot, et à ce qui le distingue d'un mot de sens proche. Son étude porte sur les conditions de l'acquisition de la signification, des sens, de mots nouveaux, lorsqu'ils sont rencontrés pour la première fois par l'apprenant. L'auteure montre que si on a présenté aux apprenants des phrases contenant des sens différents du mot polysémique dans des contextes variés correspondant à des types de situations distincts qui présentent des contrastes et qui mettent en évidence des éléments de sens distinctifs par rapport à un mot de sens proche, alors les apprenants peuvent inférer le sens métaphorique ou abstrait plus facilement, et utiliser le mot dans des contextes nouveaux avec justement ces sens métaphorique ou abstrait plus facilement. Cette manière de présenter le vocabulaire permet entre autres aux apprenants de « remarquer » que les contours des concepts ne sont pas similaires en L1 et en L2, ainsi que d'avoir une intuition sur le sens du mot. Il serait souhaitable qu'ils consultent différents sens dans un dictionnaire.

Ce qui se dégage des deux études précédentes est principalement, en compréhension, la nécessité d'une mise en contexte des mots à apprendre, d'une mise en évidence du sens principal, ou central, lors de l'apprentissage du mot, ainsi que d'une mise en évidence des différents sens et des liens entre eux, dans ce qu'ils ont de commun et de

contrastant. La prise de conscience de la notion de polysémie en réception, dans la mesure où elle permettrait d'élaborer les concepts, pourrait contribuer à une meilleure mise en réseau des mots rencontrés par la suite, ce qui devrait se traduire par un meilleur choix lexical en production. Les deux études précédentes vont également dans le sens d'élaborer des stratégies d'utilisation de dictionnaires.

Les réseaux lexicaux sont souvent associés à la compétence du choix lexical (par ex. Richards, 1976 ; Nation, 2001). Nous avons également vu que les apprenants devaient continuer à construire leurs réseaux lexicaux. Sur quel type de réseau, de champ sémantique, de liens entre les mots est-il intéressant de travailler pour favoriser le choix lexical, et comment ? Comme nous l'avons vu précédemment (Martin, 1984 ; Nation et Newton, 1997 ; Waring, 1997 ; Finkbeiner et Nicol, 2003), il y a un risque d'interférence sémantique à apprendre des synonymes en même temps, de même que des mots reliés sémantiquement. C'est seulement quand ces mots sont déjà connus des apprenants qu'il devient utile et aidant d'appliquer des stratégies de construction de réseaux.

Sbéa-Jarbue (1998) présente le cas de mots polysémiques dans la L1 qui correspondent à de nombreux mots différents dans la L2, chacun demandant un contexte d'utilisation particulier. Les apprenants éprouvent de la difficulté à choisir le mot qui conviendrait au contexte. Pour remédier à cela, Sbéa-Jarbue (1998) préconise l'apprentissage du vocabulaire avec mise en contexte, distinction des usages des mots de sens proche et présentation des principaux collocatifs.

Maiguashca (1984) explique la difficulté rencontrée par les apprenants à sélectionner le mot adéquat en production par la méconnaissance du lexique en tant que système, la méconnaissance des relations entre les mots. Elle se réfère en cela à Saussure ainsi qu'à des linguistes ayant développé la notion de champ sémantique (notion introduite par Jost Trier, un linguiste allemand, en 1931). L'approche qu'elle propose pour remédier à cet état de fait s'appuie sur la notion de champ sémantique qui, d'après elle, présente l'avantage de faire ressortir la structure du lexique. Elle présente une panoplie d'exercices qui permettent d'explorer des champs sémantiques. En particulier, les exercices exploitent différentes relations existant entre les mots au sein d'un même champ sémantique. Il peut s'agir de la gradation de l'intensité, de l'antonymie, des liens morphologiques, de différences de connotation, de style. Elle propose également des explorations plus ou moins guidées de champs sémantiques. Les exercices qu'elle suggère sont le plus souvent connectés à des contextes d'utilisation du vocabulaire dans des phrases.

Selon Channell (1981), dans l'enseignement du vocabulaire, ce sur quoi doit être mis l'accent est de savoir comment un mot est lié à d'autres mots de sens proche et quels sont les autres mots avec lesquels il est utilisé, en plus de connaître son sens de base. Pour cela, elle propose d'emprunter les notions de champ sémantique et d'analyse componentielle (approche fondée par Jerrold J. Katz en 1972 et s'appuyant sur la notion de traits sémantiques) à la linguistique théorique. Dans son application didactique, elle présente le vocabulaire aux apprenants sous forme de diagrammes. Par exemple, pour illustrer les nuances différenciant des mots de sens proche, elle présente un tableau à

double entrée correspondant à un champ sémantique, avec d'un côté une série de mots de sens proche, de l'autre des traits sémantiques, partagés ou non. Les mots ont tous un trait en commun, et se distinguent selon les autres traits. Elle travaille par ailleurs les collocations, les présentant également sous forme de tableaux par exemple. Cette manière de présenter se distingue de la consultation de dictionnaires et lui permet d'atteindre ses deux objectifs d'enseignement. Des exercices permettent de s'entraîner à l'utilisation des mots étudiés. Ce sont des exercices d'explicitation de similitudes et différences entre deux mots, de choix entre différents collocatifs possibles. Les résultats de ce travail sont que les apprenants ont dans l'ensemble trouvé cette méthode utile mais difficile. Un commentaire d'apprenant porte sur la difficulté à mémoriser les éléments travaillés. On peut penser que s'ils construisaient les tableaux eux-mêmes, ils mémoriseraient peut-être les mots plus facilement.

Jullian (2000) présente une activité qu'elle a menée auprès d'apprenants de niveau intermédiaire-avancé dans un cours de production lexicale. L'objectif de l'activité était double : développer la conscience linguistique des apprenants en ce qui concerne la sémantique lexicale, enrichir leur vocabulaire productif en leur permettant d'être plus précis dans le choix lexical. Sa méthode vise à « créer le concept » (p. 38) correspondant à un mot de la L2, les concepts en question n'existant pas nécessairement dans la L1. Pour cela, elle travaille sur une série lexicale, c'est-à-dire une série de mots qui ont tous quelque chose en commun, mais en même temps quelques nuances qui les différencient. Les exercices permettent de rechercher ce qui différencie les mots sur l'axe paradigmatique et les collocations sur l'axe syntagmatique.

L'activité se déroule en six étapes : rassembler des ensembles lexicaux, établir une liste d'éléments distincteurs au sein d'un ensemble lexical, classifier les mots de l'ensemble selon ces distincteurs, identifier les contextes dans lesquels ils pourraient être utilisés, utiliser les mots dans une phrase en illustrant le sens, faire ses propres recherches lexicales. Cette méthode présente aussi l'avantage de rendre les apprenants plus autonomes dans leur apprentissage du vocabulaire.

Le travail sur les réseaux sémantiques peut donc s'effectuer de différentes façons : par un travail sur les champs sémantiques (Maiguashca, 1984), par l'utilisation de diagrammes basés sur les notions de champ sémantique et d'analyse componentielle, avec des exercices d'application (Channel, 1981), ou enfin par un travail sur des séries lexicales afin de « créer le concept » (Jullian, 2000). D'autres auteurs ont également fait des propositions dans le même sens : Martin (1984), Machalias (1991), Lawal (1996). Nous ne développons pas ici les nombreux travaux effectués sur les collocations.

Cette troisième étape de la méthodologie se présente sous la forme d'activités variées : activités de type « lexicographique », activités de mise en contexte phrastique, activités de regroupement, de rapprochement et de différenciation, etc. Le travail pourra se faire collectivement, par groupes d'apprenants avec mise en commun dans la classe, mais aussi individuellement. Comme nous l'avons vu dans cette section, des exercices (ou types d'exercices) existent déjà. Nous pouvons donc adapter des types d'activités existantes aux besoins des apprenants.

3.4.2. Sélection du vocabulaire et activités

Cette étape implique une sélection du vocabulaire. Dans le travail sur les réseaux lexicaux, comment choisir le vocabulaire à travailler ? La sélection du vocabulaire à enseigner peut se fonder sur au moins trois critères, la fréquence (par ex. Nation, 2001), la difficulté (Sökmen, 1997 ; Laufer, 1990, 1997) et l'utilité (Bogaards, 1994). Si dans un type de texte, les erreurs semblent porter sur des séries de mots récurrentes, le vocabulaire pourrait être considéré comme utile à la production de ce type de texte.

Pour chaque type de texte, nous pouvons dégager des mots appartenant à certains champs sémantiques spécifiques, sur lesquels portent de nombreuses erreurs. L'étude du corpus permet dans une certaine mesure de sélectionner du vocabulaire à enseigner, éventuellement les champs sémantiques qui posent le plus de difficultés par rapport à un type de texte donné. Pour délimiter des champs, nous avons procédé de manière ascendante, en formant des catégories de sens à partir des erreurs du corpus, pour chaque type de texte. Les erreurs renseignent sur les mots ou expressions problématiques, donc difficiles, dans chaque champ sémantique.

Nous présentons ci-dessous, dans le tableau 5, les champs majeurs que nous avons dégagés, en signalant leur degré de présence dans les différents types de textes, en suivant la convention suivante : ++ (très représenté), + (représenté), - (absent). Nous remarquons tout d'abord que plusieurs champs sémantiques se retrouvent dans tous les types de textes, par exemple, le champ de la description. Cela semble logique

puisqu'aussi bien les textes narratifs et informatifs que descriptifs incluent des parties descriptives, et donc contiennent du vocabulaire spécifique à la description. Toutefois, certains champs ont une plus grande affinité avec un type de texte donné. Par exemple, le champ des liens conceptuels est davantage présent dans les textes informatifs, ce qui, à nouveau, n'a rien de surprenant. De nombreux autres champs ont une représentation minimale dans différents types de texte. Cela est dû au fait que le sujet des compositions était ouvert et que par conséquent les thèmes choisis par les apprenants étaient très variés (droit, santé, économie, sport...). Nous ne les reprendrons pas ici.

Champ sémantique	Texte descriptif	Texte narratif	Texte informatif
Description	++	++	+
Observation	++	++	+
Émotions et sentiments	+	++	-
Mental et pensée	+	++	-
Lien conceptuel	-	+	++
Énonciation et parole	-	++	-
Action	+	++	+
Processus	-	+	+
Cheminement dans la vie	++	++	++
Vie personnelle et familiale	++	+	-
Éducation et carrière	++	-	++
Temps	+	+	+

Tableau 5 : Répartition des champs sémantiques dans différents types de textes

Comme certains champs sémantiques sont très représentés dans plusieurs types de textes, il serait intéressant de les traiter la première fois qu'ils sont rencontrés par rapport à la planification du cours, mais également avec le type de texte pour lequel le lien sémantique est le plus fort. Par exemple, même si le champ de la description est présent dans tous les types de textes, on s'attend à le voir traiter pour la production de textes descriptifs. Il pourrait toutefois être approfondi dans les cours suivants.

Nous avons regroupé les erreurs extraites des différents types de textes dans des champs sémantiques unifiés, dont nous présentons la liste complète dans l'annexe 2. Le tableau 5 permet de retrouver dans quels types de textes les divers champs sémantiques sont apparus¹³.

Voici un exemple de champ sémantique que nous pourrions retenir comme source de vocabulaire pour la présentation des activités.

Cheminement dans la vie / Cheminement personnel (à présenter dans l'étude des textes descriptifs)

- Collocatifs de : adversité (connaître), chemin (prendre un autre, suivre, tracer, trouver son propre), décision (prendre), défi (s'attaquer à, lancer),

13 Il serait éventuellement intéressant de compléter ces champs sémantiques par du vocabulaire leur appartenant, indépendamment du corpus.

effort (concentrer, faire porter), engagement (prendre), expérience (acquérir, revivre, renouveler), futur (construire, bâtir, préparer), idée (renoncer à, abandonner), inspiration (importante, incroyable), intérêt (S porter sur), occasion (avoir, saisir, se saisir de, profiter de ; en or, rêvée, inespérée), passion (s'adonner à, se consacrer à), rêve (réaliser), rythme (effréné), succès (valoir, apporter)

- Verbes : obtenir/cultiver/récolter, faire appel à/se tourner vers, amener à/conduire à/obliger (succès)
- Noms : opportunité/occasion/option, décision/choix, futur/avenir, voie/appel, réalisations/parcours, accomplissement/réussite

Il est possible de travailler à partir de ce champ en sensibilisant les apprenants aux collocations demandées par les unités lexicales du champ, en détaillant les connaissances liées aux unités lexicales (restrictions de sélection, structure d'arguments, modification de sens par des affixes) et enfin en différenciant sémantiquement certains mots les uns des autres.

Les principaux types d'erreurs dégagés du corpus étant relativement fréquents, on peut considérer qu'ils correspondent à des difficultés rencontrées par les apprenants. Il est donc intéressant de les faire travailler sur ces difficultés. Duin et Graves (1987), dans une étude portant sur des enfants et la langue maternelle, établissent un lien entre

enseignement du vocabulaire et écriture : le pré-enseignement de vocabulaire enrichit les compositions. Un enseignement intensif du vocabulaire avant l'écriture peut donc être proposé.

Une autre possibilité est de travailler avec le vocabulaire produit par les apprenants dans un cours donné. Une fois l'autocorrection rendue au professeur, un travail de mise en commun au niveau de la classe peut être effectué. Il est possible d'utiliser des champs sémantiques qui se révèlent suite aux erreurs corrigées – mise en commun des erreurs et corrections possible au niveau de la classe (par un outil en ligne tel que Moodle par exemple) et choix des champs sémantiques à retenir par le professeur. Ensuite, il est possible de faire un travail de structuration à partir de ces champs : spécification de concepts et travail sur les réseaux lexicaux en fournissant un outillage conceptuel aux apprenants.

CONCLUSION ET PERSPECTIVES

Ce mémoire portait sur le choix lexical en français langue seconde. Nous avons vu que le développement de la compétence de choix lexical passe par l'affinement des concepts de la L2 et le développement des réseaux lexicaux. Ces deux aspects sont d'ailleurs interdépendants puisque l'affinement d'un concept amène souvent à différencier le mot de mots de sens proche, et donc à enrichir un réseau. Certaines erreurs sont considérées « internes » au mot puisqu'il s'agit d'erreurs liées à une mauvaise

représentation conceptuelle de la forme choisie, tandis que d'autres sont considérées « externes » au mot dans le sens où elles résultent d'une inadéquation du mot avec son contexte immédiat. Dans le premier cas, le mot choisi et le mot attendu entrent dans une relation paradigmatique, alors que dans le deuxième cas, le mot choisi et son contexte entrent dans une relation syntagmatique. Ces deux types de relations contribuent à enrichir les réseaux lexicaux. Dans un objectif de développement de la compétence de choix lexical, l'apprentissage du vocabulaire doit s'effectuer sur deux plans : celui du mot et celui du réseau lexical.

Nous avons proposé des codes de correction d'erreurs sémantiques de choix lexical obtenus à partir d'une analyse d'erreurs d'un corpus de production écrite d'apprenants de français langue seconde. Ces codes pourront être apposés sur des copies d'apprenants pour signaler des erreurs de choix lexical. Les catégories d'erreurs qui sont à l'origine de la définition de ces codes ont été caractérisées à l'aide de notions en linguistique. Cette analyse nous a permis de proposer des activités de prise de conscience des phénomènes en jeu dans le choix lexical, ainsi qu'une « méthodologie » d'autocorrection, c'est-à-dire une liste de questions que l'apprenant peut se poser pour résoudre un problème de choix lexical lorsqu'il se trouve confronté à un code d'erreur sur sa copie. Les codes d'erreurs sémantiques peuvent alors être considérés comme des outils diagnostiques. Ils fournissent des pistes à l'apprenant sur la manière de résoudre le problème de choix lexical. Par ailleurs, le travail sur les codes, dans le sens où il part d'une forme choisie pour arriver si possible à une forme attendue, met en lien des mots sur un axe paradigmatique, de même que syntagmatique lorsque les codes

indiquent des erreurs de collocation ou de relation incompatible. Il contribue donc également à l'enrichissement des réseaux lexicaux.

La question de l'enseignement du vocabulaire pose inévitablement la question de la sélection du vocabulaire à enseigner. Nous avons répondu à cette question implicite de deux, voire trois manières. La première consiste à se baser sur les erreurs issues du corpus et à considérer que celles-ci représentent des difficultés particulières rencontrées par les apprenants, en même temps que du vocabulaire utile à la rédaction de certains types de textes puisque justement les mots employés d'une manière erronée ont été utilisés. Nous avons ainsi mis en évidence un certain nombre de champs sémantiques qui posent difficulté et qu'il semble utile d'enseigner au niveau intermédiaire-avancé. La seconde consiste à faire travailler les apprenants justement sur les erreurs qu'ils commettent. C'est ce qui se passe dans la démarche d'autocorrection. Une troisième manière serait, par des activités de prise de conscience, de sensibiliser les apprenants à l'observation plus attentive des propriétés lexicales lorsqu'ils rencontreront un mot par la suite. Une plus grande attention et une plus grande propension à remarquer pourraient être des facteurs d'enrichissement des connaissances lexicales. De ce point de vue, le développement de la conscience linguistique et des stratégies de choix ou de consultation de dictionnaire pourraient constituer une réponse au problème que pose l'étendue du vocabulaire à acquérir comparativement aux moyens limités dont disposent les apprenants et les enseignants. Cela permettrait d'espérer que les apprenants deviennent plus autonomes dans leur apprentissage et effectuent des choix de plus en plus sûrs.

Ce travail permet d'envisager plusieurs perspectives, en particulier sur les dictionnaires d'apprenants et les stratégies de choix lexical. Comme nous l'avons mentionné, les moyens permettant d'enseigner beaucoup de vocabulaire, aussi bien quantitativement que qualitativement, sont assez limités. Il pourrait donc être intéressant de réfléchir à des spécifications pour un dictionnaire d'apprenant qui facilite à la fois la compréhension et la production, en particulier le choix lexical. Un tel dictionnaire devrait fournir des informations précises concernant (i) la polysémie, les sens centraux et les relations entre les différents sens d'un mot et (ii) les réseaux lexicaux, en particulier les traits qui différencient des mots de sens proches et des informations concernant les restrictions de sélection et les cooccurrents. Comme un dictionnaire de ce type comporterait beaucoup d'informations sur les liens entre les mots, un format électronique avec une navigation aisée à travers des réseaux serait probablement le plus adapté.

Une autre perspective intéressante serait d'enrichir la compréhension du choix lexical et des causes des erreurs des apprenants par des enquêtes sur les stratégies de choix lexical des apprenants. Il ne s'agirait pas des stratégies d'apprentissage du vocabulaire, mais vraiment de ce qu'ils font quand ils écrivent, à la fois quand ils sont conscients qu'il leur manque un mot (Utilisent-ils un dictionnaire ? Unilingue ? Bilingue ? Autre ?) et quand ils utilisent un mot, qui en fait est erroné, sans qu'ils en aient conscience. Cela permettrait de proposer des stratégies de choix lexical plus adaptées.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Adachi, T. (2003). Accurate learning of word usage: Differentiating semantically similar words. *Foreign Language Annals*, 36(2), 267-278.
- Arnaud, J.L., & Savignon, S. (1997). Rare words, complex lexical units and the advanced learner. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*, (pp. 157-173). Cambridge: Cambridge University Press.
- Beheydt, L. (1987). The semantization of vocabulary in foreign language learning. *System*, Vol. 15(1), 55-67.
- Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*, Paris: Crédif-Hatier/Didier.
- Bolly, C. (2005). Séquences (semi) figées et constructions récurrentes avec le verbe à haute fréquence prendre en FL1 et FL2. *CILL*, 31(2), 149-167.
- Bossé-Andrieu, J., & Mareschal, G. (1999). Paramètres de délimitation des collocations du français courant. *Revue canadienne de linguistique appliquée*, 2(1-2), 21-31.
- Bouffard, P. (2003). *FRAN 306. Communication écrite. Recueil de notes*. 2^e édition Université Concordia, Montréal.
- Channell, J. (1981). Applying Semantic Theory to Vocabulary Teaching. *ELT Journal*, 35(2), 115-122.
- Chiu, C.-H. (2009). ESL learners' semantic awareness of English words. *Language Awareness*, 18(3-4), 294-309.
- Coady, J. (1997). L2 vocabulary acquisition. A synthesis of the research. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*, (pp. 273-290). Cambridge: Cambridge University Press.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- De Bot, K. (1992). A bilingual production model: Levelt's 'Speaking' model adapted. *Applied Linguistics*, 13(1), 1-24.
- Duin, A. H., & Graves, M. F. (1987). Intensive vocabulary instruction as a prewriting technique. *Reading research quarterly*, 22(3), 311-330.

- Ellis, N. C. (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*, (pp. 122-139). Cambridge: Cambridge University Press.
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24, 369-383.
- Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20(3), 465-480.
- Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10(2), 221-239.
- Hamel, M.-J., & Milićević J. (2007). Analyse d'erreurs lexicales d'apprenants du FLS: démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage. *Revue canadienne de linguistique appliquée*, 10(1), 25-46.
- Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect*, 21(3), 3-25.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *SSLA*, 21, 303-317.
- Hill et Graves (1987). Intensive vocabulary instruction as a prewriting technique. *Reading Research Quarterly*, 22(3), 311-330.
- Hu Hsueh-chao, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign language*, 13(1), 403-430.
- Ijaz, H. I. (1986). Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning*, 36(4), 401-451.
- James, C. (1998). *Errors in language learning and use. Exploring Error Analysis*. London: Longman.
- Jiang, N. (2000). Lexical Representation and Development in a Second Language. *Applied Linguistics*; 21(1), 47-77.
- Jiang, N. (2004). Semantic Transfer and Its Implications for Vocabulary Teaching in a Second Language. *The Modern Language Journal*, 88(3), 416-432.
- Jullian, P. (2000). Creating word-meaning awareness. *ELT Journal*, 54(1), 37-46.

- Laufer, Batia. (1989). What percentage of lexis is essential for comprehension. In C. Lauren & M. Nordman (Eds), *From Humans Thinking to Thinking Machines*, (pp. 316-323). Clevedon: Multilingual Matters.
- Laufer, B. (1990). Ease and Difficulty in Vocabulary Learning: Some Teaching Implications. *Foreign Language Annals*, 23(2), 147-155.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In P.J.L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132).
- Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*, (pp. 140-155). Cambridge: Cambridge University Press.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223-250.
- Lawal, A. (1996). Implications and applications of semantic field theory and componential analysis in vocabulary teaching. *Glottodidactica*, 24, 63-72.
- Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition*, 42, 1-22.
- Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*, (pp. 255-270). Cambridge: Cambridge University Press.
- Lewis, M. (2000). *Teaching collocation. Further developments in the lexical approach*. Hove: Language Teaching Publications.
- Liu, E. T. K., & Shaw, Ph. M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL*, 39, 171-194.
- Machalias, R. (1991). Semantic networks in vocabulary teaching and their application in the foreign language classroom. *Babel*, 26(3), 19-24.
- Maiguashca, R. U. (1984). Semantic Fields: Towards a Methodology for Teaching Vocabulary in the Second-Language Classroom. *The Canadian Modern Language Review*, 40(2), 274-297.

- Martin, M. (1984). Advanced Vocabulary Teaching: The Problem of Synonyms. *The Modern Language Journal*, 68, 130-137.
- Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, 11, 28-38.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance & Competence in Second Language Acquisition* (pp. 35-53). Cambridge: Cambridge University Press.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A.P. Cowie (Ed.), *Phraseology. Theory, analysis, and application* (pp. 23-53). Oxford: Clarendon Press.
- Mel'čuk, I. A., Clas A., & Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Miller G. A., & Fellbaum, Ch. (1991). Semantic networks of English. *Cognition*, 41, 197-229.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*, (pp. 238-254). Cambridge: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2007). The path from learner corpus analysis to language pedagogy: some neglected issues. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 305-315). Amsterdam: Editions Rodopi.
- Nicolae, Ioana (2003). *Words have grammar : investigating semantic and syntactic aspects of word knowledge for verbs*. Mémoire de maîtrise. Concordia University.
- Qian, D. D. (1999). Assessing the Role of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension. *The Canadian Modern Language Review*, 56(2), 282-307.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined?. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-217). Amsterdam: John Benjamins.

- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.
- Salem, I. (2007). The lexico-grammatical continuum viewed through student error. *ELT Journal*, 61(3), 211-219.
- Saussure, F. de (1916/1995). *Cours de linguistique générale*, Paris: Éditions Payot.
- Sbéa-Jarbue, S. (1998). L'influence des polysèmes de la langue maternelle sur l'apprentissage du vocabulaire en langue étrangère. *IRAL*, 36(2), 162-171.
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48(2), 281-317.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Singleton, D. (2000). *Language and the lexicon: an introduction*. London: Edward Arnold.
- Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*, (pp. 237-257). Cambridge: Cambridge University Press.
- Sonaiya, R. (1991). Vocabulary acquisition as a process of continuous lexical disambiguation. *IRAL: International Review of Applied Linguistics in Language Teaching*, 29(4), 273-284.
- Tréville, M.-C., & Duquette, L. (1996). *Enseigner le vocabulaire en classe de langue*. Paris: Hachette.
- Verspoor, M., & Lowie, W. (2003). Making sense of polysemous words. *Language Learning*, 53(3), 547-586.
- Waring, R. (1997). The negative effects of learning words in semantic sets: a replication. *System*, 25(2), 261-274.
- Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: Collins.
- Zhou, A. A. (2009). What adult ESL learners say about improving grammar and vocabulary in their writing for academic purposes. *Language Awareness*, 18(1), 31-46.

ANNEXE 1 : Exemples tirés du corpus représentant quelques classes

CONCEPT INADÉQUAT

Exemple	Mot utilisé	Mot attendu	Différence de sens
<i>étayer la porte avec une poutre</i>	étayer	renforcer caler barrer	direction
<i>distinguer sa richesse</i>	distinguer	percevoir	comparaison, discrimination
<i>ses yeux estiment que</i>	estimer	démontrer ?	Point de vue
<i>sa qualité travailleuse s'expose aux mains</i>	s'exposer	s'observer	point de vue
<i>jeter les hommes</i>	jeter	projeter	se débarrasser
<i>éclipser le public de l'horreur</i>	éclipser	protéger soustraire ?	[construction]
<i>enlever son meilleur enquêteur</i>	enlever	remplacer	
<i>ne plus trouver ses amis</i>	trouver	voir	suppose : chercher
<i>quitter quelqu'un seul chez lui</i>	quitter	laisser	implique : définitif
<i>obtenir un verre d'eau à la cuisine</i>	obtenir	prendre	suppose : de quelqu'un, effort ?
<i>commencer à se côtoyer</i>	se côtoyer	se fréquenter	volonté, relation plus forte
<i>collecter ses bibelots préférés</i>	collecter	rassembler	de différents propriétaires
<i>tout en dénigrant Paul</i>	dénigrer	dédaigner	verbalement
<i>nerveux et en secouant</i>	secouer	trembler	force de l'extérieur vs action de l'intérieur
<i>son vagabondage a été brusquement arrêté</i>	arrêter	interrompre	peut recommencer
<i>ses yeux indiquaient sa jeunesse</i>	indiquer	révéler	point de vue

<i>le moment indésirable s'approximait</i>	approximer	approcher	mathématiques
<i>Claire met son manteau et commence à sortir</i>	commencer	se préparer	aspect
<i>claquer la boîte</i>	claquer	refermer brusquement	collocatif de porte
<i>les deux arrivèrent à un rocher géant</i>	arriver	s'arrêter	destination finale
<i>il dégager/vider tout son change à côté de la caisse</i>	dégager, vider	déposer	
<i>apercevoir ce grand homme</i>	apercevoir	observer	attention
<i>s'intégrer aux évolutions technologiques</i>	s'intégrer	s'adapter	conformité à l'environnement vs changement
<i>ces arts martiaux innovés par l'homme</i>	innover	inventer	[construction]
<i>aviser quelqu'un</i>	aviser	conseiller	prévention
<i>arranger un environnement</i>	arranger	aménager	
<i>avoir l'opportunité</i>	opportunité	occasion	
<i>de nouvelles opportunités</i>	opportunité	possibilité	
<i>une décision concernant son futur</i>	futur	avenir	
<i>mettre fin à un rapport</i>	rapport	amitié	
<i>donner un coup d'œil dans son passé</i>	coup d'œil	aperçu	
<i>les mémoires de sa vie</i>	mémoire	souvenir	
<i>être l'opposante</i>	opposant	rebelle	contre quelque chose en particulier, défini
<i>outils de nettoyage</i>	outil	matériel	
<i>l'addition d'une caméra</i>	addition	ajout	mathématiques
<i>la prochaine destination</i>	prochain	suisant	repérage temporel
<i>ne rien trouver de méfiant</i>	méfiant	inquiétant	[construction]
<i>leurs choses étaient absentes</i>	absent	avoir disparu	présence avant
<i>rigide et effrayée</i>	rigide	figée, raide ?	

<i>comme ils faisaient banalement à chaque matinée</i>	banalement	habituellement	commun
--	------------	----------------	--------

COLLOCATION ATTENDUE

Exemple	Mot base	Collocatif existant	Mot utilisé
<i>emporter l'attention</i>	attention	attirer	emporter
<i>prendre l'attention</i>	attention	capter	prendre
<i>fortifier la vérité</i>	vérité	renforcer, rétablir	fortifier
<i>faire un engagement</i>	engagement	prendre	faire
<i>prendre une maladie</i>	maladie	attraper	prendre
<i>donner un regard</i>	regard	lancer	donner
<i>créer une réunion</i>	réunion	organiser	créer
<i>risquer la stabilité de sa famille</i>	stabilité	compromettre, menacer	risquer
<i>abréger/brimer les droits civils</i>	droit	bafouer	abréger, brimer
<i>un amour n'est pas resté</i>	amour	durer	rester
<i>la tradition se maintient</i>	tradition	se perpétuer	se maintenir
<i>la rumeur se diffuse</i>	rumeur	se répandre, se propager	se diffuser
<i>une vaste générosité</i>	générosité	grand, immense	vaste
<i>son titre céleste</i>	titre	prestigieux	céleste
<i>ses traits visuels</i>	trait	physique	visuel
<i>une douleur insurmontable</i>	douleur	insupportable, insoutenable, intolérable	insurmontable
<i>une histoire ancestrale</i>	histoire	ancien	ancestral
<i>un bonheur énorme</i>	bonheur	extraordinaire, formidable, immense	énorme
<i>pleurer excessivement</i>	pleurer	à chaudes larmes	excessivement
<i>gagner sa vie aisément</i>	gagner sa vie	bien	aisément

RELATION INCOMPATIBLE

Erreurs provenant de l'incompatibilité sémantique du prédicat utilisé avec l'argument

Exemple	Prédicat	Argument	Incompatibilité du prédicat
<i>période d'agrandissement</i>	agrandissement	enfant	humain (croissance)
<i>le poète personnifié</i>	personnifié	poète	contradiction (archétype de)
<i>descendants répandus dans plusieurs pays</i>	répandu	descendant	humain vs matière (éparpillé)
<i>maison âgée de soixante ans</i>	âgé	maison	humain (vieille)
<i>évoquer une présence</i>	évoquer	présence	contradiction (révéler)
<i>le teint de ses cheveux</i>	teint	cheveu	visage (teinte, couleur)
<i>se doter d'une silhouette</i>	se doter de	silhouette	contradiction (conserver)
<i>écouler ses terres</i>	écouler	terre	marché (vendre)
<i>pied sablonneux</i>	sablonneux	pied	matière (recouvert de sable)
<i>dévaluer ses sentiments</i>	dévaluer	sentiment	monnaie (dévaloriser)
<i>des lèvres roux</i>	roux	lèvre	cheveu (rouge)
<i>ses yeux me regardent</i>	regarder	yeux	humain (se fixer sur, s'arrêter sur)
<i>les gouttes ont grandi</i>	grandir	goutte	humain (grossir)
<i>les gouttes sont coulées</i>	couler	goutte	matière/comptable (tombé, glissé)
<i>le tempérament bilingue de la ville</i>	tempérament	ville	humain (caractère)
<i>engourdir la douleur</i>	engourdir	douleur	partie du corps (atténuer)
<i>déléguer l'équipe</i>	déléguer	équipe	(travail) ?
<i>transférer ses pensées sur papier</i>	transférer	pensée	deux destinataires (transposer)
<i>aider son investigation</i>	aider	investigation	humain (faciliter)
<i>les moments contents</i>	content	moment	humain (heureux)
<i>l'obscurité empirait</i>	empirer	obscurité	situation

<i>être témoin d'une voix</i>	témoin	voix	événement (entendre)
<i>la cage a sursauté</i>	sursauter	cage	humain (bouger)
<i>dire son rêve étrange</i>	dire	rêve	propos (raconter)
<i>un ours est alors deux fois plus probable d'attaquer</i>	probable	ours	événement, phénomène (susceptible)
<i>être extatique</i>	extatique	quelqu'un	état, sentiment, comportement (en extase)
<i>cette gourmandise est son apogée quotidienne</i>	apogée	quelqu'un	non humain (l'apogée de sa journée)
<i>Jack, l'un des éléments de la bande</i>	élément	bande	ensemble (membre)
<i>jusqu'à la fin du corridor</i>	fin	corridor	(fond, bout)
<i>un gros chien blond</i>	blond	chien	humain (au pelage clair)
<i>des contenus âgés</i>	âgé	contenu	humain (ancien)
<i>manger un petit-déjeuner</i>	manger	petit-déjeuner	plat, aliment vs repas (prendre)
<i>les gens de l'entourage des arts martiaux</i>	entourage	art martial	humain (milieu, domaine)
<i>assister à une école primaire</i>	assister	école	événement (fréquenter)
<i>faire appel aux cigarettes</i>	faire appel à	cigarette	humain (se tourner vers)
<i>nier des lunettes</i>	nier	lunettes	propos (refuser)
<i>donner raison à son interdiction</i>	donner raison à	interdiction	humain (justifier)
<i>mettre un terme à son adversaire</i>	mettre un terme à	adversaire	humain (achever)
<i>finir son adversaire</i>	finir	adversaire	humain (achever, battre, éliminer, défaire)

Erreurs liées à la violation de contraintes sémantiques sur les arguments

Exemple	Prédicat	Argument	Incompatibilité de l'argument
<i>raccrocher la ligne</i>	raccrocher	ligne	partie/tout (combiné)

<i>raccrocher le téléphone</i>	raccrocher	téléphone	partie/tout (combiné)
<i>le poète personnifié</i>	personnifié	poète	contradiction (poésie)
<i>les personnes internationales</i>	international	personne	hyperonyme (étudiant)

AFFIXE INCOMPATIBLE

Exemple	Mot utilisé	Mot requis	Différence
Verbes			
<i>se forcer à regarder tout droit</i>	se forcer à	s'efforcer de	Préfixe [ø/ef-]
<i>se rétablir quelque part</i>	se rétablir	s'établir	Préfixe [r-/ø]
<i>se renfermer quelque part</i>	se renfermer	s'enfermer	Préfixe [r-/ø]
<i>son succès le mena à déménager</i>	mener	amener	Préfixe [ø/a-]
<i>pour apparaître plus grand</i>	apparaître	paraître	Préfixe [ap-/ø]
<i>un homme apparaît un peu mystérieux</i>	apparaître	paraître	Préfixe [ap-/ø]
<i>la stabilité intellectuelle récompense</i>	récompenser	compenser	Préfixe [ré-/ø]
<i>ce jeune homme se porte avec beaucoup de confiance et d'indépendance</i>	se porter	se comporter	Préfixe [ø/com-]
<i>affirmer mes suspenses sur lui</i>	affirmer	confirmer	Préfixe [af-/con-]
<i>accrocher le combiné avec force</i>	accrocher	raccrocher	Préfixe [ø/r-]
<i>un filet de l'eau a découlé du robinet</i>	découler	couler	Préfixe [dé-/ø]
<i>elle semblait plutôt à une montagne des boîtes</i>	sembler	ressembler	Préfixe [ø/res-]
<i>apercevoir une différence flagrante entre eux deux</i>	apercevoir	percevoir	Préfixe [ap-/ø]
<i>se trouver toute seule dans un parc</i>	se trouver	se retrouver	Préfixe [ø/re-]

<i>ses yeux grandissent en voyant une boîte de chocolat</i>	grandir	s'agrandir	Préfixe [ø/a-]
<i>je ressentais mon cœur ralentir</i>	ressentir	sentir	Préfixe [re-/ø]
<i>elle sent une force qui la retira brusquement vers le haut</i>	retirer	tirer	Préfixe [re-/ø]
<i>nous ne le marquons plus</i>	marquer	remarquer	Préfixe [ø/re-]
<i>le couloir était éclairci</i>	éclaircir	éclairer	Suffixe [-cir/ø]
Adjectifs			
<i>un père courageux et avoué tant à sa famille qu'à son travail</i>	avoué	dévoué	Préfixe [a-/dé-]
<i>la maison avoisinante</i>	avoisinant	voisin	Préfixe [a-/ø] suffixe [-ant/ø]
<i>des sentiments changeables</i>	changeable	changeant	Suffixe [-able/-ant]
<i>être inquiet</i>	inquiété	inquiet	Suffixe [-é/ø]
<i>le sexe opposant</i>	opposant	opposé	Suffixe [-ant/-é]
<i>les visages pâles et incroyants</i>	incroyant	incrédule	Suffixe [-oyant/-édule]
<i>des femmes rondelles</i>	rondelle	rondelette	Suffixe [-lle/-lette]
<i>des filles prétendues</i>	prétendu	prétentieux	Suffixe [-du/-tieux]
<i>les frais sont variés partout au monde</i>	varié	variable	Suffixe [-é/-able]
<i>les légumes contiennent une quantité significative de vitamines, minéraux...</i>	signifiant	significatif	Suffixe [-ant/-catif]
Noms			
<i>plusieurs personnes nécessitent le confort de ces rayons</i>	confort	réconfort	Préfixe [ø/ré-]
<i>l'endommagement de la maison</i>	endommagement	dommage	Préfixe [en-/ø] suffixe [-ment/ø]
<i>ses écritures</i>	écriture	écrit	Suffixe [-ure/ø]

<i>quelques ans plus tard</i>	an	année	Suffixe [ø/-née]
<i>être une vraie inspiratrice</i>	inspiratrice	inspiration	Suffixe [-trice/-tion]
<i>ouvrir un concessionnaire d'automobiles</i>	concessionnaire	concession	Suffixe [-naire/ø]
<i>déclarer sans réservation</i>	réservation	réserve	Suffixe [-ation/-e]
<i>faire des préparations pour sa fille impatientement attendue</i>	préparation	préparatif	Suffixe [-ion/-if]
<i>pleine d'espérance</i>	espérance	espoir	Suffixe [-érance/-oir]
<i>la gagnante du temps</i>	gagnant	gain	Suffixe [-gant/-in]
<i>la signifiante d'avoir une éducation</i>	signifiante	signification	Suffixe [-ante/-cation]
<i>plusieurs avances en technologie</i>	avance	avancée	Suffixe [-e/-ée]

ANNEXE 2 : Exemples de champs sémantiques issus du corpus

Vie personnelle et familiale

- Collocatifs de : temps (préserver), pratiques religieuses (s'adonner à, se livrer à), famille (diriger, fonder), stabilité (menacer, compromettre), ami (gagner, se faire), amitié (durer, lier, unir), amour (durer), sommeil (plonger), maladie (attraper), sommeil (tirer de), rêverie (tirer de), tradition (se perpétuer)
- Adjectifs : natal/maternel
- Verbes : se rétablir/s'établir, épouser/se marier
- Noms : rapport/amitié/relation, mémoires/souvenirs

Éducation et carrière

- Collocatifs de : éducation (recevoir), diplôme (obtenir), connaissance (acquérir, accumuler), industrie (réussir, se démarquer), entreprise (démarrer), projet (participer), réunion (organiser), argent (dépenser), bourse d'études (généreuse), travail (acharné), chômage (important, élevé, massif, considérable), compétence (acquérir), coût (élevé, occasionnel), frais (élevé), consommation (importante)
- Verbes : écouler/vendre, enseigner/apprendre, assister à/fréquenter, apprendre/inculquer (bonnes habitudes), s'intégrer/s'adapter
- Noms : un concessionnaire/une concession

Émotions et sentiments

- Collocatifs de : colère, frustration (manifester, exprimer), agressif (se montrer), compassion (faire preuve, éprouver), crime (inqualifiable, irréparable, abominable, terrible), confusion (jeter dans), sentiment (éprouver), situation (se trouver dans, vivre, être dans), pression (peser sur, s'exercer sur), bonheur (être transporté de, nager, baigner dans), responsabilité (assumer), douleur (insupportable, insoutenable, intolérable), cauchemar (être en proie à, être en plein), bonheur (extraordinaire, formidable, immense)
- Verbes : dévaluer/dévaloriser, se forcer à/s'efforcer de, ressentir/sentir
- Noms : parole/mot à dire/voix, espérance/espoir
- Adjectifs : étonnant/surprenant, content/heureux (moments), extatique/en extase, méfiant/inquiétant, inquiété/inquiet, incroyant/incrédule

Observation

- Collocatifs de : attention (attirer), curiosité (piquer), regard (lancer, poser, porter (point de vue)), yeux (scruter, se figer, s'arrêter, se fixer), attention (capter), montre (indiquer), beauté (admirer)
- Verbes : prévoir/pressentir/anticiper, regarder/fixer/s'arrêter sur, distinguer/mesurer/percevoir, signifier/traduire/faire écho à/représenter, estimer/démontrer, s'exposer/se voir, dévoiler/déchiffrer, paraître/apparaître/se montrer, témoigner/être témoin de, se porter/se comporter, être témoin de/entendre, indiquer/révéler, évoquer/montre, apercevoir/observer, faire attention/découvrir/remarquer, sembler/ressembler, apparaître/paraître/avoir l'air,

apercevoir/percevoir, apparaître/paraître/sembler, témoigner/être témoin de,
marquer/remarquer

- Noms : similarité/points communs/similitude, spécificité/particularité, coup d'œil, aperçu

Description

- Collocatifs de : jeunesse (conférer, redonner), sourire (arborer, afficher), signe (montrer), générosité (grande, immense), sourire (grand, large), caractère (affirmé), yeux (grands), courage/force (exceptionnels, incroyables), trait (physique), sourire (se dessiner, adresser, gratifier de), visage (comptait), larme (coule), rumeur (se répandre, se propager), odeur (dégager), popularité (croître), silhouette (mince, fine, affinée, svelte)
- Noms : agrandissement/croissance, teint/teinte/couleur, force/puissance, caractère/tempérament, fin/bout, postérieur/arrière,
- Adjectifs : âgé/vieux/ancien, roux/rouge, changeable/changeant, entêté/entêtement, blond (pour un chien), âgé/ancien, opposant/rebelle, opposant/opposé, rigide/figé, avoisinant/voisin, varié/variable, signifiant/significatif
- Verbes : révéler/évoquer, conserver/se doter de, répandre/éparpiller/disperser, étayer/renforcer/consolider, jeter/projeter/catapulter, grandir/grossir, couler/tomber, grandir/s'agrandir, infiltrer/figurer dans

Mental et pensée

- Verbes : entamer/penser à, penser/réfléchir, savoir/comprendre, négliger/oublier, savoir/apprendre/entendre/prendre connaissance de, mémoire/souvenir
aider/faciliter, ruiner/changer/modifier (sentiments, opinion), affirmer/confirmer
- Noms : pensée/commentaire, support/soutien

Parole

- Collocatifs de : mot (adresser), phénomène (parler de, expliquer), pensée (réfléter, rendre compte de, transposer)
- Verbes : dire/raconter, crier/ajouter, annoncer/dire/demander, lire/dire/être écrit (message), annoncer/s'exclamer

Lien conceptuel

- Collocatifs de : conséquence (découler), effet (avoir, engendrer, produire), conclusion (tirer), nécessité (souligner, réaffirmer), problème (causer, engendrer), interdiction (déplorer, protester contre), avantage (présenter, procurer, apporter), inquiétude (éveiller, faire naître, inspirer, semer, susciter, causer, soulever, provoquer)
- Verbes : nier/refuser, donner raison à/justifier, dire/signifier/indiquer/vouloir dire, s'ensuivre/résulter/devenir/se transformer en, rapporter/provoquer/avoir pour conséquence, s'ensuivre à/occasionner
- Noms : fin/motif/raison, argument/discussion, signifiante/signification

Action

- Collocatifs de : promenade (faire), achat (faire, réaliser), autobus (prendre, manquer), coup de poing (asséner), achat (effectuer)
- Verbes : s'approximer/approcher, sursauter/bouger, sortir/se répandre, se bouger/se déplacer, jeter/éjecter, claquer/refermer brusquement, ramener/rapprocher/rassembler, arriver/s'arrêter à, prêter/utiliser/emprunter, dégager/vider/déposer, venir/approcher/passé (autobus), supporter/soutenir, remplir/accomplir/effectuer/faire (tourné), se trouver/se retrouver, retirer/tirer/soulever, pratiquer/répéter (pièce), innover/inventer, retourner/revenir, arranger/aménager, aviser/conseiller
- Noms : préparation/préparatif, addition/ajout

Processus

- Verbes : interrompre/arrêter, débiter/commencer/entamer, commencer/se préparer à, se mettre à/commencer à, débiter à/commencer à, avancer/se poursuivre
- Noms : évolution/avancée (technologique)

Temps

- Noms : temps/moment/fois, jour/journée, an/année, soir/soirée, gagnante/gain/gagne [de temps]