La description de la compétence lexicale en langue étrangère : perspectives méthodologiques

Sylviane Granger et Guy Monfort

Édition électronique

URL : <https://journals.openedition.org/aile/4890>

DOI : 10.4000/aile.4890

ISSN : 1778-7432

Éditeur

Association Encrages

Édition imprimée

Date de publication : 1 janvier 1994

Pagination : 55-75

ISSN : 1243-969X

Référence électronique

Sylviane Granger et Guy Monfort, « La description de la compétence lexicale en langue étrangère : perspectives méthodologiques », *Acquisition et interaction en langue étrangère* [En ligne], 3 | 1994, mis en ligne le 20 février 2012, consulté le 14 avril 2022. URL : <http://journals.openedition.org/aile/4890> ; DOI : <https://doi.org/10.4000/aile.4890>

Ce document a été généré automatiquement le 14 avril 2022.

© Tous droits réservés

La description de la compétence lexicale en langue étrangère : perspectives méthodologiques

Sylviane Granger et Guy Monfort

NOTE DE L'AUTEUR

Cet article s'appuie sur une recherche subventionnée par le Fonds National de la Recherche Scientifique et le Fonds de Développement Scientifique de l'Université Catholique de Louvain.

Nos collègues et amis Jean Klein, Jean-Marie Pierret et Stephanie Tyson ont accepté de relire une première version de notre article et de nous faire profiter de leurs suggestions. Qu'ils en soient ici remerciés.

1. Introduction

- 1 En 1984, dans un article qui fera date dans les études sur l'apprentissage du lexique, Meara lançait un cri d'alarme qui a été largement entendu. Le lexique, que Meara présentait, à juste titre, comme l'aspect le plus négligé de l'interlangue de l'apprenant, a depuis fait l'objet d'une série impressionnante d'articles et d'ouvrages. D'autre part, d'un point de vue méthodologique, de nombreux linguistes ont progressivement délaissé l'analyse d'erreurs - approche qui, de l'avis de Meara, est purement « taxonomique » et manque totalement de pouvoir explicatif - pour adopter une orientation psycholinguistique.
- 2 Le présent article se propose de démontrer que l'analyse d'erreurs, si elle repose sur une base empirique suffisante et des catégories bien définies, est un outil de description très utile, dont les résultats sont exploitables tant dans le cadre d'une analyse psycholinguistique que dans le domaine de la didactique des langues. Si nul ne pense

plus aujourd'hui à se cantonner à la seule analyse des erreurs, celle-ci constitue néanmoins une technique d'analyse qui, associée à d'autres approches, permet de mieux appréhender le lexique de l'apprenant. Ringbom (1987: 70-71) reconnaît à la fois les limites - « *error analysis is not sufficient on its own* » - et les côtés positifs de l'analyse d'erreurs lorsqu'il écrit qu'elle représente « *as good a starting point as any for investigating the processes of L2-learning (...), especially if it is combined with other types of investigation, such as frequency counts, contrastive analysis, studies of inferencing procedures and reaction time tests* ».

- 3 En fait, si les analyses d'erreurs se sont révélées décevantes, c'est principalement à cause du flou méthodologique qui les caractérise. En effet, l'hétérogénéité des données et des catégories d'erreurs empêche toute comparaison ou exploitation valable des résultats. Cependant, tout comme Taylor (1986: 145), nous pensons que « *these criticisms need not spell the death of error analysis. They can be met* ».

2. Données et procédés d'analyse

- 4 Se proposant de dresser le bilan des recherches sur l'acquisition des langues secondes durant ces vingt dernières années, Véronique (1992: 7) constate combien il serait difficile de rendre compte de l'ensemble des travaux : « La confrontation et la synthèse des résultats sont rendues difficiles par la diversité des publics observés et des méthodologies utilisées ». Dans l'avant-propos de son livre sur l'acquisition des langues étrangères, Klein (1989) fait le même constat : « Il est vrai qu'il existe de très nombreuses recherches empiriques, programmes ambitieux ou études individuelles, mais leurs résultats sont disparates, souvent difficiles à comparer entre eux, et l'image qui s'en dégage est relativement floue ».
- 5 La recherche sur le rôle - largement controversé - du transfert dans le processus d'apprentissage a souffert des différentes faiblesses méthodologiques que sont l'étroitesse de la base empirique (Singleton, 1990 : 1), le manque de précision dans la description des données (Nickel, 1989 : 298 et Selinker, 1992 : 241-243), et la nature trop subjective des catégories employées (Nickel, *ibid.* : 296 et Sheen, 1988 : 44).
- 6 Pour pallier la plupart de ces lacunes, il est essentiel de définir un cadre méthodologique rigoureux, dont Abbott soulignait déjà l'importance en 1980 : « *It is difficult to avoid the conclusion that without the rigour of an agreed analytical instrument, researchers will tend to find in their corpuses ample evidence of what they **expect** to find* » (p. 121)¹.

3. Conception large du lexique

- 7 Bien que toutes les analyses d'erreurs comportent une catégorie intitulée « erreurs lexicales », rares sont les linguistes qui se donnent la peine de définir ce terme. Lorsqu'on tente de découvrir sur quelle base ils opposent le lexique à la grammaire, on se rend compte que certains linguistes opposent visiblement ce qui est régulier au domaine des irrégularités, mais que la plupart se fondent sur l'opposition ensemble ouvert/ensemble fermé. Ainsi, par exemple, les prépositions, qui constituent un ensemble fermé, sont presque unanimement classées dans le domaine grammatical, alors qu'elles ont un comportement éminemment idiosyncrasique.

- 8 L'embarras des linguistes face à l'opposition lexical/grammatical ressort clairement de plusieurs commentaires. Ainsi, Færch et al. (1984 : 109-110) clôturent leur chapitre sur l'analyse d'erreurs par la réserve suivante : « ...and finally that many errors, broadly categorized as "predicate", "noun phrase" or "adverbial phrase", are borderline errors between grammar and idiom usage ». Ils ajoutent : « Certain categories of function words, like prepositions and conjunctions, have not been covered so far in the analysis of grammatical errors. Like idioms they belong to the border area between grammar and vocabulary ». Lennon (1991a : 40), quant à lui, classe les erreurs de prépositions parmi les erreurs grammaticales, mais il reconnaît que, si on adopte une conception plus large du lexique, on peut les considérer comme lexicales.
- 9 De tout ceci, il ressort qu'il est indispensable de définir ce que l'on entend par lexique avant de se lancer dans une analyse des erreurs lexicales. La question se pose alors de savoir quelle conception du lexique adopter. Rey (1991 : 183) apporte une réponse claire à cette question : « En fait, le seul concept opératoire du lexique est pragmatique. En lexicographie, en pédagogie des langues, en traduction, il ne peut s'agir que d'un lexique utile pour un certain projet, et non pas d'une entité théorique définie *a priori* ».
- 10 Comme nous nous situons dans une perspective d'apprentissage des langues, il s'agit de savoir ce que connaître un mot signifie pour l'apprenant. Richards (1976) définit la compétence lexicale de manière large en y incluant, outre une connaissance de type logico-sémantique (le ou les sens et les associations de sens), de nombreux autres facteurs tels que, par exemple, collocabilité, registre et comportement syntaxique. C'est ce type de conception large du lexique que nous nous proposons d'adopter.
- 11 Nous classerons comme lexicales toutes les erreurs qui sont imputables à une mauvaise connaissance des propriétés inhérentes au mot, que celles-ci soient d'ordre logico-sémantique (le sens dénotatif), stylistique (le registre), collocationnel (la combinatoire lexicale) ou lexico-grammatical (les propriétés morpho-syntaxiques du mot). Il est clair en effet que, comme le souligne Carter (1987 : 67), « *knowing a word involves a complex of factors and (...) learning words should, as far as possible, take account of syntactic, semantic, stylistic and collocational dimensions (...)* ». Cet élargissement du lexique a pour conséquence que de nombreuses erreurs qui sont traditionnellement classées comme grammaticales sont ici considérées comme lexicales. Il en est ainsi des erreurs suivantes : **informations* au lieu de *information* ; **responsible of sb* au lieu de *responsible for sb* ; **avoid to do sth* au lieu de *avoid doing sth*. Pour employer correctement le substantif anglais *information*, il est indispensable de savoir qu'il est non dénombrable, de même qu'un bon emploi de l'adjectif *responsible* accompagné d'un complément présuppose que l'on sache de quelle préposition il peut être suivi. Dans le troisième cas, c'est la complémentation du verbe - gérondif ou infinitif - qui est en cause.
- 12 Notre approche du lexique de l'apprenant se fonde sur un élargissement de la notion de sémantique lexicale qui, selon Rey (1991 : 190), « ne concerne pas seulement (...) la "sémantique" *stricto sensu*, mais aussi des éléments syntactiques (d'où la pertinence du concept de grammaire-lexique) ». Cette démarche rejoint les recherches actuelles de linguistes tels que Gross (1991) pour le français et Sinclair (1991) pour l'anglais, qui mettent tous deux en évidence l'interrelation entre grammaire et lexique. C'est une démarche similaire qui est également à la base des manuels d'apprentissage du vocabulaire les plus récents².

4. Classification des erreurs lexicales

4.1. Principes méthodologiques

- 13 Le flou méthodologique ne concerne pas que l'opposition grammaire/lexique. La comparabilité des résultats est également rendue difficile par la diversité de classement des erreurs lexicales elles-mêmes. Quelques exemples suffiront pour démontrer cette hétérogénéité. De manière générale, on peut distinguer trois principes de classement : (1) une description linguistique des erreurs ; (2) une analyse des sources d'erreurs ; (3) un amalgame des deux.
- 14 Parmi les partisans du premier type de classement, certains - comme Færch *et al.* (*ibid.* : 286) ou Lennon (1991a : 31-39)³ - classent les erreurs lexicales selon la catégorie grammaticale à laquelle le mot appartient : erreurs de nom, erreurs de verbe, etc. D'autres, tels que Zimmermann (1986), ont recours à la sémantique lexicale et distinguent des erreurs de sens (synonymie, inclusion, etc.), de collocation, de registre, de formation des mots, etc. Si le classement par catégorie grammaticale peut mettre en évidence des phénomènes intéressants (tels que la prépondérance des erreurs de préposition, par exemple), il présente le net désavantage d'être très éloigné des pratiques didactiques qui, comme nous l'avons indiqué plus haut, se basent sur les différents types de « connaissance lexicale ». Nous adopterons donc une catégorisation basée sur la sémantique lexicale.
- 15 Duskova (1969) et Ringbom (1987) représentent la deuxième approche : ils distinguent, par exemple, les erreurs lexicales dues à une confusion entre mots de forme et/ou de sens similaires, les erreurs dues à une divergence lexicale (un vocable de L1 correspond à plusieurs vocables en L2), etc. Dans le troisième cas, les principes de classement sont hybrides, c'est-à-dire qu'ils associent des éléments de description et d'explication. Ghadessy (1980 : 103), par exemple, distingue, à côté des erreurs lexicales, les erreurs dues à l'influence du persan. Il est clair que le second type d'erreurs est d'un ordre totalement différent du premier, qu'il recouvre d'ailleurs partiellement (voir aussi Zughouli, 1991).
- 16 Dans notre analyse, nous distinguerons clairement description linguistique et explication psycholinguistique des erreurs. Une telle distinction a été prônée par plusieurs linguistes, mais ceux-ci n'ont malheureusement pas toujours été suivis, loin s'en faut. En 1980, déjà, Abbott affirmait : « *The aim of any EA is to provide a "psychological explanation" - a reliable account of the causes of errors. But before this can be attempted it is necessary to order the chaotic data by a preliminary process of description and linguistic categorisation* » (p.124). Zimmermann (1986 : 32) va dans le même sens lorsqu'il écrit : « *I am making the analytic distinction between linguistic **description** and psychological **explanation** (...). In doing so, I hope to do better justice to the **subject matter** of learning by disregarding the learning **process*** ».
- 17 Les deux principes méthodologiques que nous suggérons d'adopter pour la catégorisation des erreurs lexicales peuvent se résumer de la façon suivante :
1. effectuer l'analyse en deux étapes distinctes : une analyse purement descriptive, basée sur la sémantique lexicale, suivie d'une analyse à orientation psycholinguistique, destinée à déterminer la source des erreurs ;
 2. adopter des catégories clairement définies, qui se recouvrent le moins possible et sont suffisamment larges pour réduire le risque de classements arbitraires.

4.2. Etape descriptive

4.2.1. Remarque préliminaire

- 18 L'étape descriptive vise à détecter et à classer les erreurs lexicales des apprenants. On entend par erreur « a linguistic form or combination of forms which, in the same context and under similar conditions of production would, in all likelihood, not be produced by the speakers' native speaker counterparts » (Lennon, 1991b: 182). Il s'agit donc d'une approche « externe » qui adopte résolument le point de vue du locuteur natif. Cette démarche permet de mettre en évidence les maladresses lexicales (« unhappy lexical choices », Martin, 1984 : 130) qui distinguent nettement l'apprenant de niveau avancé du locuteur natif. Dans son article consacré à la compétence lexicale en français langue étrangère, Sautermeister (1989 : 123) a montré que ces différences se situent à plusieurs niveaux : différences de signification, confusions entre les différents registres, erreurs de type collocationnel, par exemple. Pour l'anglais langue étrangère, Martin (1984: 130) a relevé des erreurs lexicales à quatre niveaux distincts: « the four dissonances that interfere with encoding by advanced learners are stylistic (**dunk** the chicken pieces in the beaten egg mixture); syntactic (the author **purports** that tobacco is harmful); collocational (I used to be a **large** smoker); and semantic (I was badly **damaged** in the accident) ». Sur le plan pédagogique, le but ultime de cette approche est de délimiter des pistes de recherche pour la conception de méthodes d'enseignement du vocabulaire et l'élaboration d'outils de référence à l'intention des apprenants de niveau avancé.

4.2.2. Vue d'ensemble de la classification

- 19 Le schéma ci-dessous donne une vue d'ensemble de la catégorisation des erreurs lexicales :

Schéma 1

- 20 La catégorisation que nous proposons est visiblement simple. C'est là une décision délibérée, qui découle de notre volonté de mettre en œuvre le second principe méthodologique (voir 4.1.). Il va de soi que chacune des catégories que nous distinguons

est susceptible d'être subdivisée. Toutefois, plus on affine l'analyse, plus on augmente le risque de classements arbitraires et donc d'aboutir à des données non comparables. Nous préférons donc nous en tenir, dans un premier temps, à des catégories assez larges⁴.

4.2.3. Erreurs lexico-grammaticales

- 21 Les erreurs lexico-grammaticales constituent des violations des propriétés morpho-syntaxiques inhérentes au lexème employé, telles que le genre ou l'opposition continu/discontinu pour les substantifs, les formes verbales et les pluriels irréguliers, le type de construction (transitive, intransitive, pronominale) pour les verbes, et de manière générale, toute la problématique de la combinatoire syntaxique (régime des noms, des verbes, des adjectifs et des adverbes). Elles se distinguent par leur caractère individuel des erreurs proprement grammaticales qui sont, quant à elles, de l'ordre du régulier, du généralisable (règles d'accord, emploi des temps, des modes, des pronoms, des articles, etc.).
- 22 Les exemples suivants, extraits de nos corpus de FLE et d'ALE, illustrent les types d'erreurs les plus fréquents.

(1)	le livre sterling {la}
(2)	sa puissance militaire et son pouvoir améliorent sensiblement {s'améliorent}
(3)	Le système de gouvernement et d'administration se fonctionnera un peu comme à l'URSS {fonctionnera}
(4)	...dont les priorités sont différentes que celles de la France {de}
(5)	un grand nombre de gens craint perdre leurs propres coutumes {craint de perdre}
(6)	the Europe of 1992 aims at progresses in the fields of... {progress}
(7)	their parents, girlfriends or wifes {wives}
(8)	our old Europe will rejuvenate {be rejuvenated}
(9)	leaders cannot agree with the capital of Europe {on}
(10)	To live in the same nation means to share the same language {sharing}
(11)	respect of the rights of man {for}

- 23 Les dictionnaires d'apprentissage devraient contenir suffisamment de renseignements sur le comportement grammatical des mots pour permettre aux non-natifs d'éviter de telles erreurs. Un effort remarquable a été réalisé pour la langue anglaise, qui dispose de plusieurs « *learner's dictionaries* » de grande qualité. Malheureusement, le français accuse un sérieux retard dans ce domaine (cf. Corbin, 1991 : 165).

4.2.4. Lexème « inexistant » en L2⁵

- 24 Nous classerons sous cette rubrique les lexèmes qui n'existent pas en tant que tels dans la langue cible. Les cas les plus fréquents sont les créations lexicales et les emprunts. Des exemples tels que (12) et (15), que Ringbom (1987) appelle « *hybrids* » ou « *blends* », constituent un amalgame de ces deux procédés.

(12)	un climat dans lequel les rivautés entre ces deux pays furent énormes {rivalités}
(13)	la création d'un Europarliament supranational {europarlement}
(14)	Les conservatives britanniques {conservateurs}
(15)	European citizens are being sensibilized to Europe {becoming more aware of}
(16)	... look mistrustingly at Europe {mistrustfully}
(17)	an unnegligible opportunity {significant}

- 25 Il est important de souligner que les erreurs d'orthographe proprement dites n'entrent pas dans cette catégorie. Elles constituent une catégorie à part, distincte des erreurs grammaticales et des erreurs lexicales⁶.

4.2.5. Erreurs logico-sémantiques

- 26 Nous rangerons dans la présente catégorie les erreurs qui sont dues à une méconnaissance du sens dénotatif du mot⁷.

(18)	...la Constitution allemande prévient l'envoi d'un contingent {prévoit}
(19)	... si les institutions réjouissaient de plus de pouvoir {jouissaient}
(20)	elle est positive dans son engagement {ferme, résolue}
(21)	des issues idéologiques {questions, problèmes}
(22)	...concrete realizations {achievements}
(23)	Every nation should bring its own part to the construction of Europe {contribution}
(24)	the late dispute between France and Belgium {recent}
(25)	Their population is as important as the rest of Europe {large}

- 27 Zimmermann (1986) subdivise cette catégorie en plusieurs sous-catégories : fautes d'associations de sens (« *sense relation errors* » : hyperonyme/hyponyme ; co-hyponymes ; hétéronymes ; partie/tout ; etc.), fautes de champ sémantique (« *field errors* »), fautes de traits distinctifs (« *feature errors* »). Il reconnaît cependant lui-même

que « *this typology raises several problems of delimitation and overlapping of categories* » (p. 36). Ainsi, par exemple, il admet que de nombreuses erreurs d'associations de sens pourraient très bien être considérées comme des erreurs de traits distinctifs. Fidèles au second principe méthodologique (cf. 4.1.), nous ne subdiviserons pas systématiquement cette catégorie.

4.2.6. Erreurs stylistiques⁸

- 28 Il s'agit ici de confusions entre les différents registres : familier, soutenu, littéraire, technique, archaïque, etc. Cette catégorie contient également les cas de variété géographique. De nombreuses études (Halliday, McIntosh et Strevens, 1964 ; Cass, 1983 ; Laufer, 1990) ont souligné l'importance de ce type d'erreurs. Laufer (1990 : 293 et 302), par exemple, présente les « *register restrictions* » comme un des facteurs qui font que « tel mot nouveau dans une langue étrangère s'apprend plus facilement que tel autre ».

(26)	Ce serait affreux si la femme rejetait tout à fait le lien privilégié qu'elle a avec la nature {regrettable}
(27)	Il y a ceux qui ont des sous {de l'argent}
(28)	Europeans politicians will have a bone to pick with each other on various issues {have many issues to resolve}
(29)	I hope the leaders will find great solutions to the problems. {positive}

4.2.7. Erreurs syntagmatiques

- 29 Alors que les deux catégories précédentes sont d'ordre paradigmatique, les erreurs classées sous cette rubrique relèvent de la syntagmatique du mot. Il s'agit d'une catégorie très large, qui recouvre tous les cas de cooccurrence lexicale restreinte - de la collocation à la locution figée - ainsi que les paraphrases.
- 30 A l'instar de Van Roey (1990: 48), nous employons le terme « collocation » dans le sens restreint de collocation lexicale fixée par l'usage: « *Collocation is the term we shall now use both for the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than that of its "synonyms" because of constraints which are not on the level of syntax or conceptual meaning but on that of usage, as also for the word combinations which represent this phenomenon* ». Sautermeister (1989 : 124) souligne la difficulté que présentent ces groupements pour les locuteurs non natifs : « Par rapport à la locution dont les unités sont difficilement séparables (**faire fi de**), chaque membre de la collocation conserve son autonomie (**faire du sport**). Aussi la collocation est-elle difficile à distinguer du groupement libre pour un étudiant étranger et elle donne lieu à des interférences dans la mesure où chaque langue possède ses propres collocations ». Le terme « paraphrase » est, quant à lui, employé dans le sens que lui confère Zimmermann (1988 : 300) : « *The term **lexical paraphrase** is restricted to phrase-level expressions normally replacing shorter forms, i.e. simplex or complex words* » (voir exemples 34-35 et 41-42).

(30)	La voie vers le marché unique a été entamée par l'Acte unique en 1986 {ouverte}
(31)	L'Europe continuera à agrandir les bénéfices {augmenter}
(32)	...une collaboration proche avec les autres pays {étroite}
(33)	la volonté (...) est, de toute probabilité, ...{selon}
(34)	les hommes scientifiques {scientifiques}
(35)	la Grande-Bretagne (...) est en coopération avec la France {coopère}
(36)	Efforts are done in this direction {made}
(37)	Other countries are too busy with their inner problems {domestic}
(38)	a common external policy {foreign}
(39)	But, on the other side , we will form a new nation {hand}
(40)	...if they feel such indifference from the part of their parents {on the part of}
(41)	...eliminating the existing barriers of trade {trade barriers}
(42)	For instance, in the field of education, Belgium has a system that is different from that of France {the Belgian and French systems are different}

4.3. Etape explicative

- 31 Dès l'avènement de l'analyse d'erreurs, les linguistes se sont intéressés aux sources d'erreurs. Il faut bien constater aujourd'hui que les résultats des diverses analyses sont peu concluantes, l'influence de la langue maternelle étant jugée insignifiante par certains, prépondérante par d'autres. Ces contradictions sont principalement dues au fait que la plupart des linguistes n'invoquent que deux sources d'erreurs possibles - interlinguale et intralinguale - alors qu'une proportion importante des erreurs permettent les deux interprétations.
- 32 Pour pallier cette situation, nous proposons d'adopter la tripartition prônée, entre autres, par Sheen (1988) et Singleton (1990), qui distinguent les catégories suivantes : (1) les erreurs intralinguales ou développementales ; (2) les erreurs interlinguales, subdivisées en deux sous-catégories : (a) les erreurs qui ne s'expliquent que par le transfert ; (b) les erreurs pour lesquelles le transfert ne constitue qu'une explication parmi d'autres.
- 33 Nous utiliserons trois des exemples cités plus haut pour illustrer cette tripartition. L'exemple (3) appartient clairement à la première catégorie : l'emploi de **se fonctionner** s'explique par analogie avec d'autres formes pronominales propres à L2, telles que **se passer**, **se dérouler**, etc. Dans des cas d'emprunt tels que (13), l'explication interlinguale est la seule possible. Quant à l'exemple (10), il illustre la catégorie (2b) : l'emploi de l'infinitif peut s'expliquer soit de manière interlinguale, le verbe **signifier**

étant suivi de l'infinitif, soit de manière intralinguale, par analogie avec des verbes tels que *hope*, *expect*, etc.

- 34 Une analyse plus approfondie de la source des erreurs permet de distinguer les erreurs dues à une confusion de type formel de celles qui ont une origine sémantique ou qui - comme c'est le cas pour les « faux amis » - font appel tant à la forme qu'au sens. Les résultats de ce type d'analyse devraient permettre de mieux appréhender le lexique mental de l'apprenant et de voir dans quelle mesure il diffère de celui du locuteur natif.
- 35 Dans ce cadre, un corpus de compositions libres constitue une base empirique intéressante. Par sa nature même, il représente un terrain d'investigation relativement neutre par rapport à d'autres types de données, telles que les résultats de tests d'associations de mots (Meara, 1984), de C-tests (Singleton et Little, 1991) ou de traductions (Zimmermann, 1987). Il est clair, par exemple, que dans un « *cloze test* » les réponses des apprenants auront plus tendance à être sémantiquement motivées, alors que dans une traduction, l'influence de la forme du texte de départ risque de se faire sentir. La composition est de ce point de vue un exercice moins « orienté » (bien que nous soyons conscients qu'un grand nombre d'étudiants, même avancés, pensent dans leur langue maternelle et donc traduisent en même temps qu'ils composent).
- 36 D'autre part, les corpus rassemblés à l'Université de Louvain représentant un niveau avancé d'interlangue, nous pourrions vérifier si, comme le suggèrent Singleton et Little (1991), l'apprenant, tout comme le jeune enfant qui apprend sa langue maternelle, suit un processus évolutif, passant de stratégies plus formelles au départ à des liens de plus en plus sémantiques par la suite. Nos recherches préliminaires dans ce domaine nous permettent d'affirmer, à l'instar de Zimmermann (1987 : 65), que la part du formel reste importante à un niveau avancé : « *Even seemingly advanced learners can rely on L1-based form-orientation as a strategy of lexical search to an unforeseen extent.* » Des recherches plus approfondies seront nécessaires pour démontrer la part respective des processus formels et sémantiques.

5. Apport de l'informatique

- 37 L'analyse d'erreurs a bien entendu ses limites, qui ont d'ailleurs été abondamment décrites dans la littérature. Ainsi, par exemple, elle ne permet pas de mettre en évidence la richesse et la densité lexicales, pas plus que les cas de surreprésentation ou de sous-représentation de mots ou de groupes de mots. Etant donné que la description de tels phénomènes repose sur l'exploitation de données quantitatives, l'apport de l'informatique se révèle ici inestimable. A cet égard, il n'est pas inutile de rappeler ce que, selon Muller (1991 : 13), la lexicologie peut attendre de l'informatique : « D'une part, sur chaque élément du lexique, on va accumuler un nombre considérable de contextes, donc de données sémantiques et syntaxiques (...). D'autre part, on va obtenir une description quantitative de ce qu'est le vocabulaire d'un texte (...) ». Ce qui est vrai pour la lexicologie en général l'est également pour le lexique de l'apprenant.
- 38 Nos recherches dans le cadre du projet ICLE nous ont permis de préciser le rôle de l'informatique dans le traitement des données d'interlangue.
- 39 Dans un premier temps, l'ordinateur permet d'obtenir rapidement des listes de fréquences. Une comparaison des fréquences « brutes » des mots et expressions employés par les apprenants avec celles des locuteurs natifs livre des résultats très

intéressants et parfois surprenants. Ainsi, une étude portant sur les connecteurs logiques (cf. Granger & Tyson, 1993) montre que les apprenants francophones et les locuteurs natifs emploient ces éléments du discours à une fréquence équivalente. Ce résultat contredit des affirmations, telles que celle exprimée par Crewe (1990 : 324), selon lesquelles « *Not only are they [connectors] used with erroneous meanings (...) but they are over-used* ». On aurait d'autant plus été en droit de s'attendre à un suremploi des connecteurs chez les francophones que les études contrastives de l'anglais et du français soulignent de manière unanime la préférence du français pour une explicitation des articulations du discours. Notre étude montre que cette prédilection ne vaut que pour certains types de connecteurs et qu'elle est largement compensée par un sous-emploi caractérisé d'autres catégories de connecteurs.

- 40 D'autres données chiffrées fournies par l'ordinateur, telles que la proportion vocable/mot (« *type/token ratio* »), permettent d'évaluer la richesse lexicale des apprenants. Cependant, il est important de noter que, si l'on ne dispose pas d'un bon lemmatiseur, l'intérêt de ce type de calcul reste très limité (Kirk, à paraître).
- 41 Les données quantitatives ne sont bien sûr qu'un point de départ pour une analyse qualitative. Les logiciels de concordance tels que TACT ou OCP pour l'anglais⁹, permettent de visualiser chaque occurrence d'un mot ou d'une expression dans son contexte. En ce qui concerne les connecteurs, cette étape a permis de mettre en évidence une différence d'emploi importante entre les anglophones et les francophones : ces derniers placent pratiquement deux fois plus souvent les connecteurs en début de phrase que les anglophones.
- 42 L'emploi des logiciels de concordance permet de mieux appréhender le lexique phraséologique de l'apprenant. TACT, par exemple, permet de rechercher automatiquement dans un corpus toutes les séquences récurrentes de deux mots ou plus. Une comparaison portant sur la nature et la fréquence d'emploi de ces séquences chez les natifs et les non-natifs (S. Granger, en préparation) devrait permettre de préciser le rôle que joue le « préfabriqué » dans la compétence non-native. Cette discussion a pour toile de fond le débat sur l'opposition « *open-choice principle* »/« *idiom-principle* » lancé par Sinclair (1987).
- 43 Toutes les analyses décrites ci-dessus ont été effectuées à partir d'un corpus à l'état « brut » (« *raw corpus* »). Des recherches bien plus complexes encore sont possibles si le corpus de départ est enrichi d'informations grammaticales ou autres. Les logiciels de catégorisation automatique, par exemple, permettent d'associer à chaque mot une étiquette qui indique la catégorie grammaticale (nom, verbe, adjectif, etc.) à laquelle il appartient. Un corpus d'interlangue ainsi annoté présente d'énormes avantages. Si un chercheur souhaite étudier l'emploi des prépositions chez les apprenants et qu'il n'a accès qu'à un corpus à l'état brut, il pourra facilement effectuer une recherche automatique, puisque les prépositions constituent un ensemble fermé. Cependant, cette recherche lui apportera également une part importante de données indésirables, étant donné que la plupart des prépositions anglaises sont également des particules adverbiales (*in the garden/come in*). Par contre, s'il dispose d'un corpus catégorisé (« *tagged corpus* »), il aura immédiatement accès aux seules prépositions.
- 44 Les résultats de l'étape descriptive de l'analyse d'erreurs peuvent, eux aussi, être encodés. Chaque erreur reçoit, outre son étiquette grammaticale, un code d'erreur. Ainsi, l'emploi de **inner problems* au lieu de *domestic problems* (cf. 4.2.6.) est accompagné du code général « lexical » et du code spécifique « collocationnel ». Cet

encodage permet de dresser un catalogue complet des diverses catégories d'erreurs lexicales présentes dans un corpus donné, catalogue qui se révélera très utile pour l'élaboration de nouveaux outils pédagogiques.

- 45 Les diverses pistes que nous avons tracées montrent que l'informatique apporte un éclairage nouveau au lexique de l'apprenant. Tout porte à croire que l'informatisation des données d'interlangue pourrait provoquer dans ce domaine de recherches une révolution tout aussi spectaculaire que celle qui a eu lieu dans le monde de la lexicographie (cf. Granger, à paraître, 1994)¹⁰.

6. Conclusion

- 46 Malgré l'intérêt indéniable des recherches portant sur le « processus » d'apprentissage, très en vogue actuellement, il y a encore une place en linguistique appliquée pour une approche axée sur le « produit ». En particulier, une analyse d'erreurs rigoureuse, enrichie de l'apport de l'informatique, peut fournir, tant aux didacticiens qu'aux psycholinguistes, la base empirique solide qui leur a tant fait défaut jusqu'à présent.

BIBLIOGRAPHIE

ABBOTT, G. (1980). « Towards a more rigorous analysis of foreign language errors », *International Review of Applied Linguistics*, 18(2) : 121-134.

CARTER, R. (1987). *Vocabulary. Applied Linguistics Perspectives, Aspects of English*, London, Unwin Hyman.

CASS, G. (1983). *Stylistic Deficiency in the English of German Advanced Learners in Teacher Education*, Tübingen, Gunter Narr Verlag.

CORBIN, D. (1987). *Morphologie dérivationnelle et structuration du lexique*, *Linguistische Arbeiten*, 193 et 194, Tübingen, Max Niemeyer Verlag.

CORBIN, P. (1991). « La lexicographie moderne : contrepoint », *Travaux de Linguistique*, 23 : 161-169.

CREWE W. (1990). « The illogic of logical connectives », *ELT Journal*, 44/4: 316-325.

DUSKOVA, L. (1969). « On sources of errors in foreign language learning », *International Review of Applied Linguistics*, 7: 11-36.

FÆRCH, C. et al. (1984). *Learner Language and Language Learning*, Clevedon, Multilingual Matters.

GHADESSY, M. (1980). « Implications of error analysis for second/foreign language acquisition », *International Review of Applied Linguistics*, 18(2): 93-104.

GRANGER, S. (1993). « The International Corpus of Learner English », in J. AARTS, P. DE HAAN et N. OOSTDIJK (eds.). *English Language Corpora: Design, Analysis and Exploitation*, Amsterdam et Atlanta, Rodopi.

- GRANGER S. & S. TYSON. (1993). « Connector usage in native and non-native English essay writing ». Exposé présenté lors de la 10ème Conférence AILA, 9-13 Août 1993, Amsterdam.
- GRANGER S. (en préparation) « Prefabricated patterns in advanced EFL writing: collocations and lexical phrases ».
- GRANGER, S., (à paraître, 1994). « The Learner Corpus: a Revolution in Applied Linguistics », *English Today*.
- GROSS, M. (1991). « Lexique et syntaxe », *Travaux de Linguistique*, 23 : 107-132.
- HALLIDAY, M.A.K., McINTOSH, A. et STREVESEN, P. (1964). *The Linguistic Sciences and Language Teaching*, Longmans' Linguistic Library, London, Longmans.
- HARMER, J. et R. ROSSNER (1991). *More than words. Vocabulary for upper intermediate to advanced students*, London, Longman.
- HARRISON, M. (1990). *Word Perfect. Vocabulary for fluency*, Nelson.
- KIRK, J. (à paraître). « Taking a Byte at Corpus Linguistics », in L. FLOWERDEW et K.K. TONG (eds). *Proceedings of the Joint Seminar on Corpus Linguistics and Lexicology* (18-23 June, 1993), Hong Kong, Hong Kong University of Science and Technology.
- KLEIN, W. (1989). *L'acquisition de langue étrangère*, Paris, Armand Colin.
- LAUFER, B. (1990). « Why are some words more difficult than others? - Some Intralexical Factors that Affect the Learning of Words », *International Review of Applied Linguistics*, 28(4): 293-307.
- LENNON, P. (1991a). « Error and the very advanced learner », *International Review of Applied Linguistics*, 29(1): 31-44.
- LENNON, P. (1991b). « Error: Some Problems of Definition, Identification, and Distinction », *Applied Linguistics*, 12/2: 180-196.
- MARTIN, M. (1984). « Advanced Vocabulary Teaching: The Problem of Synonyms », *The Modern Language Journal*, 68/2: 130-137.
- MEARA, P. (1984). « The study of lexis in interlanguage », in A. DAVIES, C. CRIPER et A.P.R. HOWATT (eds.). *Interlanguage*, Edinburgh, Edinburgh University Press.
- MULLER, C. (1991). « Lexicographie et informatique », *Travaux de Linguistique*, 23 : 11-15.
- NICKEL, G. (1989). « Some controversies in present-day error analysis: "contrastive" vs. "non-contrastive" errors », *International Review of Applied Linguistics*, 27: 293-305.
- REY, A. (1991). « A propos de la sémantique lexicale », *Travaux de Linguistique*, 23 : 179-193.
- RICHARDS, J. (1976). « The role of vocabulary teaching », *TESOL Quarterly*, 10(1): 77-89.
- RINGBOM, H. (1987). *The Role of the First Language in Foreign Language Learning*, Clevedon, Multilingual Matters.
- SAUTERMEISTER, C. (1989). « Pour une meilleure compétence lexicale », *Le Français dans le Monde*, N° spécial (août-septembre) : 122-133.
- SELINKER, L. (1992). *Rediscover Interlanguage*, London and New York, Longman.
- SHEEN, R. (1988). « The importance of negative transfer in the speech of near-bilinguals », in D. NEHLS (ed.). *Interlanguage Studies*, Heidelberg, Julius Groos Verlag: 43-57.
- SINCLAIR, J. (1991). *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.

- SINCLAIR J. (1987). « Collocation: a progress report », in R. STEELE & T. THREADGOLD (eds.). *Language Topics. Essays in honour of Michael Halliday*, Volume II, Amsterdam & Philadelphia, John Benjamins: 319-331.
- SINGLETON, D. (1990). « The Cross-linguistic factor in second language learning », *CLCS Occasional Paper*, 24: 120.
- SINGLETON, D. et LITTLE, D. (1991). « The second language lexicon: some evidence from university-level learners of French and German », *Second Language Research*, 7(1): 61-81.
- TAYLOR, G. (1986). « Errors and Explanations », *Applied Linguistics*, 7(2): 144-166.
- VAN ROEY, J. (1990). *French-English Contrastive Lexicology. An Introduction*, Louvain-la-Neuve, Peeters.
- VERONIQUE, D. (1992). « Recherches sur l'acquisition des langues secondes : un état des lieux et quelques perspectives », *AILE*, 1 : 5-35.
- ZIMMERMANN, R. (1986). « Classification and distribution of lexical errors in the written work of German learners of English », *Papers and Studies in Contrastive Linguistics*, 21: 31-40.
- ZIMMERMANN, R. (1987). « Form-oriented and content-oriented lexical errors in L2 learners », *International Review of Applied Linguistics*, 25(1): 55-67.
- ZIMMERMANN, R. (1988). « Paraphrase errors and word-formation errors in advanced German learners of English », in J. KLEGRAF et D. NEHLS (eds.). *Essays on the English Language and Applied Linguistics on the occasion of G. Nickel's 60th birthday*, Heidelberg, Julius Groos Verlag: 300-314.
- ZUGHOUL, M. (1991). « Lexical choice: towards writing problematic word lists », *International Review of Applied Linguistics*, 29(1): 45-60.

NOTES

1. Les corpus de FLE et d'ALE rassemblés à Louvain se caractérisent par un haut degré d'homogénéité. La variété d'interlangue qu'ils recouvrent peut être décrite de la manière suivante :

- (1) **conditions d'apprentissage** : langue étrangère (et non langue seconde) ;
- (2) **langue maternelle** : français pour les apprenants d'anglais (cf. *infra*) et anglais pour les apprenants de français
- (3) **genre de tâche** : composition libre de type argumentatif ;
- (4) **niveau d'apprentissage** : avancé (étudiants universitaires d'anglais ou de français en fin de cursus).

Toutes ces données - compositions et profils des apprenants - sont informatisées (cf. point 5).

Le corpus d'anglais langue étrangère produit par des apprenants francophones s'inscrit dans une plus vaste base de données intitulée *The International Corpus of Learner English*, qui comprend des productions écrites en anglais par des apprenants de langues maternelles différentes (français, néerlandais, allemand, suédois, chinois). Pour une description complète du corpus et de son exploitation informatique, voir Granger (1993).

2. Citons, par exemple, *Word Perfect* de Harrison (1990) ou *More than words* de Harmer et Rossner (1991). Dans l'introduction de son livre, Harrison précise: « *Each entry in Word Perfect includes the information which students need to be able to use the new items of vocabulary accurately. This information includes: prepositions, verb patterns, countable or uncountable and collocations* ».

3. Le principe de base de la classification de Lennon est la catégorie grammaticale mais, en outre, une distinction est faite entre les formes qui présentent ou non un problème collocationnel. Il

distingue les « *verb choice errors with/without collocational inappropriacy* », les « *noun choice errors with/without collocational inappropriacy* », etc.

4. En ce qui concerne la définition du concept d'erreur lui-même, il faut noter que nous ne faisons pas - initialement du moins - de distinction entre « *mistakes* » et « *errors* ».

5. Nous sommes conscients du caractère relatif de la notion de « mot inexistant ». Corbin (1987) propose plutôt les notions de « mot construit possible non attesté », de « mot non complexe non construit », etc.

6. Par erreur d'orthographe, nous entendons les violations des conventions graphiques, qui n'affectent pas ou peu la prononciation. Les cas les plus fréquents ont trait à l'emploi des majuscules (**chef d'état**), aux accents (**coûtures**), au redoublement de consonnes et de voyelles (**britannique** ; **finallement**) ainsi qu'à une série d'autres graphies erronées (**volontée** ; **aspet**). Il s'agit dans tous les cas d'orthographe dite « d'usage » et non d'orthographe grammaticale (**tu va** ; **chevals**).

7. Font également partie de la catégorie des erreurs logico-sémantiques les erreurs touchant à la connotation du mot, c'est-à-dire à ses valeurs affectives, telles que les nuances péjoratives ou mélioratives qui lui sont associées. Ces erreurs sont relativement rares et, en outre, la ligne de démarcation entre le dénotatif et le connotatif est extrêmement floue, comme le montre, par exemple, le traitement lexicographique de l'adjectif **skinny**. La nuance péjorative de cet adjectif est considérée comme faisant partie intégrante du sens du mot dans le *Longman Dictionary of Contemporary English* (*skinny* : infml, often derog very thin, esp. in a way that is unattractive), alors qu'elle n'est considérée que comme une simple connotation dans l'*Oxford Advanced Learner's Dictionary* (*skinny* : infml usu derog very thin).

8. Les erreurs stylistiques sont ici considérées comme une sous-catégorie des erreurs lexicales, mais il est évident que la variation stylistique dépasse largement le cadre du lexique. En effet, le contexte situationnel influence tous les niveaux de description linguistique. Cass (1983: 102-105) explique qu'il existe une « *multiplicity of levels of language which are potentially adaptable as planes on which style markers of some contextual feature may be located (graphemic, phonemic, morphemic, lexical, etc.)* ». On retrouve notamment des exemples de variantes stylistiques en syntaxe, telles que « *Never in my life have I done such a thing* » et « *I have never done such a thing in my life* », ou « Qu'est-ce qu'il veut dire ? » et « Que veut-il dire ? ».

9. TACT est distribué par J. Bradley, University of Toronto Computing Services, 4 Bancroft Avenue, Toronto, Ontario, Canada, M5S 1A1. OCP est distribué par Oxford University Press, Walton Street, Oxford, OX2 6DD.

10. En outre, l'informatisation des données d'interlangue a l'avantage de rendre les données « brutes » accessibles à l'ensemble de la communauté internationale. Jusqu'alors, seuls les résultats étaient disponibles et nous avons démontré combien ceux-ci étaient sujets à caution.

RÉSUMÉS

Le but principal de l'article est de démontrer que l'analyse d'erreurs reste un outil très utile pour décrire la compétence lexicale de l'apprenant. Elle peut, en effet, fournir - tant aux didacticiens qu'aux psycholinguistes - la base empirique solide qui leur a tant fait défaut jusqu'à présent.

La procédure méthodologique préconisée dans l'article comprend, outre une analyse rigoureuse des erreurs lexicales, une phase d'exploitation semi-automatique, au cours de laquelle les

données lexicales d'interlangue sont soumises à des logiciels de calcul de fréquence et de concordance.

The main purpose of this article is to show that error analysis can still be an extremely useful tool to describe learners' lexical competence, providing psycholinguists and people involved in language teaching research with the strong empirical base that they have needed for so long.

The methodological approach put forward in this article is a combination of lexical error analysis and semi-automatic processing during which lexical interlanguage data are analysed using frequency and concordancing software.

AUTEURS

SYLVIANE GRANGER

Centre d'Études Anglaises, Université Catholique de Louvain

GUY MONFORT

Centre d'Études Anglaises, Université Catholique de Louvain