

Interférences intralinguales lors de l'identification lexicale

Olga Théophanous

Université Paul Valéry-Montpellier III

Dans cet article nous nous proposons de rendre compte d'une étude qui examine l'impact de la L1 et du niveau de compétence en français L2 sur les confusions lexicales intralinguales, définies en tant qu'erreurs imputables à la similitude formelle et/ou la proximité sémantique entre deux ou plusieurs mots de la L2. Quatre-cent quatre-vingts anglophones et hellénophones de niveaux débutant, intermédiaire et avancé ont participé à une tâche de traduction de mots français écrits, de fréquences variées, vers leur L1. La comparaison des deux groupes linguistiques a démontré que les hellénophones ont une plus grande tendance que les anglophones à commettre des erreurs de confusion, et ce indépendamment du niveau de compétence en français. La distance L1-L2 semble ainsi avoir un impact sur ces erreurs *a priori* intralinguales et, plus particulièrement, sur la capacité des apprenants à distinguer entre formes lexicales similaires en français L2. Les confusions lexicales seront discutées du point de vue du lexique mental et des stratégies mises en œuvre par les apprenants lors de l'identification des mots français à l'écrit. Des implications pédagogiques pour l'enseignement du vocabulaire seront envisagées en guise de conclusion.

This paper describes the results of an empirical study about a specific type of error in vocabulary acquisition in French as a second language, that is, intralingual lexical confusion due to formal and/or semantic similarity between target words and other French words known by the learners. Four hundred and eighty English- and Greek-speaking learners (at beginner, intermediate and advanced levels of proficiency in French) were given an L2 to L1 translation task of words with various frequencies. A comparison of the two linguistic groups revealed that the Greeks were more prone to commit errors due to intralinguistic lexical confusion regardless of their level of proficiency in French. Interlinguistic distance seems to have had an impact on these *a priori* intralingual errors and more specifically on the learners' ability to distinguish between similar lexical forms in French. Lexical confusion will be discussed from the point of view of the mental lexicon and the strategies that the learners use during lexical identification of written words. The paper will conclude with pedagogical suggestions.

Adresse pour correspondance : Olga Théophanous, Université Paul Valéry-Montpellier III, Département des Sciences du Langage, Route de Mende, 34199 Montpellier Cedex 5, France.

Courriel : olga.theophanous@univ-montp3.fr.

Introduction

Traditionnellement, en langue seconde/étrangère (L2), on distingue deux grandes catégories d'erreurs qui ont généralement été traitées séparément :

- a. les erreurs interlinguales, attribuées exclusivement aux interférences de la langue maternelle (L1) (ou d'une autre langue connue par l'apprenant) dans l'apprentissage de la langue cible, et étudiées par l'analyse contrastive ;
- b. les erreurs intralinguales, imputables aux caractéristiques intrinsèques de la L2, susceptibles de toucher également les enfants qui apprennent leur L1, et dont les études de l'interlangue font leur objet de prédilection.

Sur le plan lexical, et en ce qui a trait aux erreurs interlinguales, la source la plus évidente d'erreurs réside dans les faux amis ; un grand nombre d'études contrastives existe sur ce sujet. Avec les études de l'interlangue l'accent sera mis sur le développement grammatical, principalement la syntaxe, laissant ainsi de côté le développement lexical de la L2 ; cette tendance va d'ailleurs de pair avec l'ambiance générale d'une époque où le désintéressement pour les études de vocabulaire était notoire (cf. Levenston, 1979; Meara, 1980).

Cela étant dit, certains auteurs mentionnent des erreurs de confusion chez les apprenants en L2 imputables à la similarité formelle et/ou la proximité sémantique entre deux ou plusieurs mots de la L2 (Warren, 1982; Zimmerman, 1987; Zughoul, 1991). En compréhension plus particulièrement, Bensoussan (1992) à partir d'erreurs lexicales produites dans des traductions L2-L1 par des apprenants d'anglais L2 ayant comme L1 soit l'hébreu, soit l'arabe, observe que 14 % des erreurs s'expliquent par des substitutions fâcheuses entre formes lexicales similaires. Les sujets testés ont par exemple traduit dans leur L1 « purest » pour *poorest*, « mate » pour *meet*, etc. Le sens lexical peut également être source de confusion. Warren (1982) relève à partir des compositions de ses élèves des « confusions conceptuelles » entre deux mots dont les sens sont voisins. Comme le dit Warren, ces erreurs persistent, surtout quand il s'agit de distinguer des mots sémantiquement proches qui, de plus, se ressemblent phonétiquement (ex. *raise-rise, lie-lay*)¹.

Les confusions sont des erreurs lexicales intralinguales qui prennent source dans la L2 elle-même. Il est pourtant légitime de se demander si la L1 de l'apprenant a une influence sur ces erreurs. Alsulaimani (1990, cité par Ryan et Meara, 1991), dans un test d'identification des mots anglais, observe que les apprenants arabophones d'anglais L2 confondent des mots à suites consonantiques similaires ou identiques. Ils lisent alors à haute voix *basket* pour *biscuit*, *custard* pour *castle*, *stopped* pour *stupid*, *presented* pour *president*, etc. Ces erreurs montrent que les arabophones se fient à la suite consonantique du mot sans prêter attention aux voyelles. Cette manière de faire semble être dictée

par une habitude de lecture en L1. Nous savons en effet que, contrairement aux langues indo-européennes, en arabe le décodage consiste en la reconnaissance de mots déjà connus, identifiables essentiellement par leur suite consonantique.

Laufer (1991), dans une importante étude sur des mots similaires (“synforms”) en anglais (ex. *cute–acute*, *sensible–sensitive*), observe que, d’une façon générale, les apprenants ayant comme L1 une langue sémitique ont tendance à confondre plus que les apprenants ayant comme L1 une langue germanique et ces derniers ont à leur tour tendance à confondre plus que les apprenants ayant comme L1 une langue romane. La probabilité de la confusion varie donc selon la distance L1–L2. Des indices opposés, selon lesquels les confusibles poseraient plus de problèmes aux apprenants dont la L1 est proche de la L2, existent aussi. Teemant (1988, cité par Hatch et Brown, 1995, pp. 379–380), dans une étude d’analyse d’erreurs lexicales, constate que les apprenants hispanophones et lusophones d’anglais L2 font plus de confusions formelles que les apprenants ayant comme L1 le japonais. Les confusibles seraient donc plus problématiques dans le cas des apprenants dont la L1 utilise le même alphabet que la L2, plutôt que dans le cas opposé. Par ailleurs, Laufer fait remarquer que le résultat selon lequel plus la L1 de l’apprenant est éloignée de la L2, plus le nombre de confusions augmente, ne s’applique pas à toutes les catégories de synformes qu’elle a testées. Ainsi, les apprenants d’une L1 sémitique font moins d’erreurs que les apprenants d’une L1 germanique ou romane entre synformes qui se différencient par une seule consonne, et ce, que ce soit par substitution (ex. *price–prize*) ou par addition/omission (ex. *simulate–stimulate*). Les conclusions sur l’influence de la L1 dans les erreurs de confusion en L2 étant plutôt fragmentaires et en partie contradictoires, la question initiale reste posée. L’influence du niveau de compétence en L2 dans les erreurs de confusion est une autre piste d’investigation.

Nous avons ainsi entrepris une étude dans le but de répondre aux questions suivantes :

1. Les apprenants dont la L1 est éloignée de la L2 commettent-ils plus d’erreurs de confusion imputables à la similitude formelle des mots de la L2 ?
2. Les apprenants dont la L1 est éloignée de la L2 commettent-ils plus d’erreurs de confusion imputables à la similitude sémantique des mots de la L2 ?
3. Les erreurs de confusion diminuent-elles lorsque les apprenants deviennent plus compétents dans la maîtrise de la L2 ?

Méthode

Participants

Quatre-cent quatre-vingts apprenants anglophones et hellénophones ont participé à une tâche de traduction de mots français vers leur L1, l'anglais ou le grec. Ils étudiaient le français en milieu scolaire traditionnel de non-immersion en Grande-Bretagne et en Grèce et ils étaient de niveaux de compétence variés : débutant, intermédiaire, avancé. Ont été considérés comme débutants les apprenants qui avaient à leur actif environ 80 heures de français, intermédiaires ceux qui en avaient environ 160 et avancés ceux qui en avaient environ 240. Le tableau 1 résume la distribution de l'échantillon par L1 et par niveau.

Tableau 1 : Distribution de l'échantillon par L1 et niveau

Niveau	Anglophones	Hellénophones	Total
Avancés	62	73	135
Intermédiaires	68	61	129
Débutants	117	99	216
Total	247	233	480

Test

Le *Français fondamental* (1^{er} et 2^e degrés) a servi de base lexicale pour la sélection des mots du test. Bien que de bases lexicales plus récentes existent (cf. Goodfellow, Jones et Lamy 2002; Verlinde et Selva, 2001), le *Français fondamental* présente à nos yeux les qualités suivantes :

- du point de vue scientifique, il s'agit d'un ensemble fiable et reconnu,
- du point de vue didactique, il reste, même s'il est en partie vieilli, un outil essentiel dont on se sert toujours,
- du point de vue pratique, il s'agit d'un ensemble lexical facilement maniable (3 000 mots).

Afin de ne pas favoriser l'un ou l'autre groupe linguistique, nous avons éliminé dans la liste de départ les mots amis ou congénères (cf. français *train*, anglais *train* ; fr. *entrer*, angl. *entrance*). Les faux amis (cf. fr. *chat*, angl. *chat* ; fr. *plein*, angl. *plain*) ont été, dans la mesure du possible, également éliminés. Enfin, pour des raisons évidentes, les mots qui risquaient de poser des problèmes de traduction dans l'une ou l'autre des langues maternelles des sujets testés ont aussi été écartés. Il n'est par exemple pas facile de traduire en anglais *charcutier*, *rentrée*, *bonhomme* ou *préciosité*. Une fois ces restrictions appliquées, 100 mots ont été sélectionnés sur une base de 1 sur 10. Les mots choisis ont été regroupés en trois groupes selon leur fréquence lexicale établie dans le *Français fondamental*. Cela dit, lors de l'analyse des données, 17 mots ont encore dû être éliminés. Nos analyses ont ainsi porté sur les 83 mots suivants :

a. mots « très fréquents » :

<i>casser</i>	<i>cour</i>	<i>jaune</i>	<i>midi</i>	<i>œil</i>	<i>pleurer</i>
<i>chaussure</i>	<i>dépenser</i>	<i>jeudi</i>	<i>mois</i>	<i>ouvrier</i>	<i>rencontrer</i>
<i>chemin</i>	<i>heureux</i>	<i>meuble</i>	<i>montrer</i>	<i>perdre</i>	<i>tuer</i>

b. mots « fréquents » :

<i>allumette</i>	<i>cuir</i>	<i>feu</i>	<i>lourd</i>	<i>peigner</i>	<i>sommeil</i>
<i>âne</i>	<i>doigt</i>	<i>foncé</i>	<i>maigrir</i>	<i>pendre</i>	<i>tousser</i>
<i>boisson</i>	<i>enterrer</i>	<i>jouet</i>	<i>nuage</i>	<i>plancher</i>	<i>ventre</i>
<i>bouche</i>	<i>épingle</i>	<i>laid</i>	<i>orage</i>	<i>printemps</i>	<i>viande</i>
<i>chèvre</i>	<i>essuyer</i>	<i>légume</i>	<i>paix</i>	<i>roue</i>	<i>visage</i>
<i>cracher</i>	<i>éteindre</i>				

c. mots « moins fréquents » :

<i>agneau</i>	<i>évier</i>	<i>menton</i>	<i>pois</i>	<i>scier</i>	<i>souligner</i>
<i>bâtir</i>	<i>fourrure</i>	<i>mignon</i>	<i>poumon</i>	<i>secouer</i>	<i>souris</i>
<i>cadeau</i>	<i>gaspiller</i>	<i>nid</i>	<i>renseignement</i>	<i>semelle</i>	<i>tâche</i>
<i>colline</i>	<i>grimper</i>	<i>parapluie</i>	<i>rideau</i>	<i>serrure</i>	<i>tiroir</i>
<i>déménager</i>	<i>lisse</i>	<i>piéton</i>	<i>sapin</i>	<i>soie</i>	<i>tuyau</i>
<i>emprunter</i>	<i>mendiant</i>	<i>poignet</i>			

Procédure

Dans le but de confirmer que la L1 des élèves était bien l'anglais ou le grec et pour savoir si d'autres langues étrangères étaient connues par eux, un questionnaire biographique rédigé en anglais ou en grec a précédé la complétion du test. Les tests ont été administrés pendant l'horaire normal des classes. Chaque élève a reçu les mots testés présentés dans une liste imprimée, sur deux colonnes et dans un ordre non alphabétique. La consigne imprimée dans la L1 des apprenants était la suivante : « Traduisez les mots français que vous connaissez dans la liste suivante ». Il n'y avait pas de temps limite mais la complétion de la tâche demandée n'a en aucun cas pris plus de 30 minutes. Les résultats des tests administrés ont été saisis à l'aide d'une feuille de calcul Excel, convertie par la suite en tableau SPSS en vue des traitements statistiques. Les trois premières colonnes de la feuille de calcul ont été réservées à l'identification numérique de l'étudiant (de 1 à 480), à sa L1 (1 : anglais, 2 : grec) et à son niveau (1 : avancé, 2 : intermédiaire, 3 : débutant). À partir de la quatrième colonne, les colonnes contiennent les réponses fournies pour chaque mot testé correspondant. Les types de réponses sont au nombre de quatre : a) erronée, b) bonne, c) aléatoire (c'est-à-dire impossible à interpréter ou à déchiffrer), d) omission. Chaque réponse erronée a été codée selon une suite numérique à trois chiffres. Le premier chiffre indique le type d'erreur commise, à savoir : 1) une confusion formelle, 2) une confusion sémantique, 3) une confusion sémantico-formelle, 4) un faux ami. Le deuxième chiffre de la

suite numérique caractérise un type particulier de réponse erronée à l'intérieur d'une même catégorie d'erreur.

Parmi les confusions formelles, nous avons distingué celles qui mettent en jeu a) des homophones (ex. « poids » pour *pois*); b) des parophones (ex. « semaine » pour *chemin*)². Parmi les confusions sémantiques, nous avons distingué celles qui mettent en jeu a) des co-hyponymes (ex. « velours » pour *soie*); b) des antonymes (ex. « grossir » pour *maigrir*); c) des parasyonymes (ex. « lavabo » pour *évier*). Certaines confusions sémantiques ont été qualifiées de métonymiques (ex. « embrasser » pour *bouche*) et d'autres d'hyponymiques (ex. « index » pour *doigt*). Parmi les confusions sémantico-formelles, nous avons distingué a) celles que nous avons appelées « directes » mettant en jeu deux mots qui se ressemblent sémantiquement et formellement (ex. « chaussette » pour *chaussure*); b) celles que nous avons appelées « indirectes » ou « doubles » impliquant un troisième mot (ex. « pull » pour *chemin* via *chemise*). Malgré nos précautions lors de la sélection des mots du test, certaines réponses erronées relèvent d'une fausse amitié avec les mots testés. Les faux amis ont été codés mais ne font pas partie de nos analyses subséquentes. Le tableau 2 résume la codification des réponses erronées³.

Tableau 2 : Codification des réponses erronées

1.	Confusions formelles
1.1	Homophones (ex. « poids » pour <i>pois</i>)
1.2	Parophones (ex. « semaine » pour <i>chemin</i>)
2.	Confusions sémantiques
2.1	Co-hyponymiques (ex. « velours » pour <i>soie</i>)
2.2	Antonymiques (ex. « grossir » pour <i>maigrir</i>)
2.3	Parasyonymiques (ex. « lavabo » pour <i>évier</i>)
2.4	Métonymiques (ex. « embrasser » pour <i>bouche</i>)
2.5	Hyponymiques (ex. « index » pour <i>doigt</i>)
3.	Confusions sémantico-formelles
3.1	Simple ou directes (ex. « chaussette » pour <i>chaussure</i>)
3.2	Doubles ou indirectes (ex. « pull » pour <i>chemin</i> via <i>chemise</i>)
4.	Faux amis
4.1	L1–L2 (ex. <i>gasp</i> (anglais) pour <i>gaspiller</i>)
4.2	Synonyme du faux ami L2–L1 (ex. <i>chat</i> (anglais) pour <i>gaspiller</i> via <i>gossip</i> (anglais))
4.3	L3–L2 (ex. « poche » pour <i>tâche</i> via <i>tasca</i> (italien) ou <i>Tasche</i> (allemand))

Enfin, le troisième chiffre de la suite numérique indique l'erreur proprement dite dont il a été question chaque fois pour tel ou tel mot correspondant. Pour des informations plus détaillées sur la méthodologie, voir Théophanous (2001a).

Résultats

Nous avons utilisé deux tests statistiques, à savoir le test du «t» et l'analyse de variance (ANOVA) pour savoir s'il y avait une différence significative entre les L1 et les niveaux. Conformément à ces deux tests la comparaison a été effectuée entre les moyennes obtenues. Ainsi à l'aide du test du t nous avons comparé les moyennes obtenues pour l'anglais et le grec. L'analyse de variance a été utilisée pour la comparaison des moyennes obtenues pour les trois niveaux à l'étude.

La réponse à la première question de recherche a été affirmative. Ainsi la tendance à commettre des erreurs de confusion en français L2 est plus grande chez les hellénophones que chez les anglophones. Le test du t a démontré qu'il y avait une différence significative entre les L1. En effet, les hellénophones ont fait dans l'ensemble plus de confusions lexicales que les anglophones et plus de confusions formelles en particulier ($\bar{x} = 11,3$ et $8,7$ respectivement), le test du t de $6,7$ étant significatif à $p < .001$. Ce résultat est valide même si on isole le niveau de compétence: les hellénophones ont fait en moyenne plus de confusions en général et plus de confusions formelles en particulier et ce, à tous les niveaux (cf. tableau 3). Par contre, la réponse à la deuxième question — à savoir si la L1 a un impact sur les confusions sémantiques — a été négative, le test du t de $0,3$ n'étant pas significatif.

Tableau 3 : Comparaison des L1 pour les confusions formelles

	\bar{x}	s	t
			6,7***
Anglophones	8,7	4,3	
Hellénophones	11,3	4,5	
Avancés			5,4***
Anglophones	5,6	3,0	
Hellénophones	8,6	3,3	
Intermédiaires			3,8***
Anglophones	9,9	3,7	
Hellénophones	12,5	4,5	
Débutants			4,0***
Anglophones	9,6	4,6	
Hellénophones	12,1	4,5	

*** $p < .001$

En ce qui trait à la troisième question de recherche, l'analyse de variance (ANOVA) a démontré qu'il y avait des différences significatives entre les niveaux de compétence, les résultats ($F = 37,8$; $F = 41,3$) étant significatifs

à $p < .001$ (cf. tableau 4). La moyenne des confusions lexicales en général et des confusions formelles en particulier est significativement inférieure chez les avancés ($\bar{x} = 11,1$ et $\bar{x} = 7,2$ respectivement) et supérieure chez les intermédiaires ($\bar{x} = 16,4$ et $\bar{x} = 11,4$ respectivement).

Tableau 4 : Comparaison des niveaux de compétence

	\bar{x}	s	F	Comparaisons post hoc	
Confusions lexicales					
			37,8***	Intermédiaires	Débutants
Avancés	11,1	4,3		***	***
Intermédiaires	16,4	5,3			***
Débutants	14,4	5,7			
Confusions formelles					
			41,3***	Intermédiaires	Débutants
Avancés	7,2	3,4		***	***
Intermédiaires	11,4	4,4			ns
Débutants	10,7	4,6			

*** $p < .001$; ns = non significatif

La tendance chez les intermédiaires à commettre plus d'erreurs de confusion reste valide même si on garde la L1 stable, les résultats des analyses de variance ($F = 20,2$; $F = 18,7$) étant significatifs à $p < .001$ (cf. tableau 5).

Tableau 5 : Comparaison des niveaux de compétence par L1

	\bar{x}	s	F	Comparaisons post hoc	
Confusions lexicales					
Anglophones			20,2***	Intermédiaires	Débutants
Avancés	9,4	4,0		***	***
Intermédiaires	14,9	5,2			ns
Débutants	13,5	5,6			
Hellénophones			18,7***	Intermédiaires	Débutants
Avancés	12,6	4,0		***	***
Intermédiaires	17,5	5,2			**
Débutants	15,5	5,8			

*** $p < .001$, ** $p < .01$; ns = non significatif

Des analyses supplémentaires ont démontré qu'il y avait des différences significatives entre les niveaux de compétence en tenant compte de la fréquence lexicale des mots testés. Les résultats montrent que pour les mots « très fréquents » ($F = 141$), la moyenne des confusions lexicales est supérieure ($p < .001$)

chez les débutants ($\bar{x} = 4,2$). Pour les mots « fréquents » ($F = 37,1$) et pour ceux « moins fréquents » ($F = 49,8$), par contre, elle est supérieure ($p < .001$) chez les intermédiaires ($\bar{x} = 7,2$ et $\bar{x} = 5,5$) (cf. tableau 6).

Tableau 6 : Comparaison des niveaux selon la fréquence lexicale

	\bar{x}	s	F	Comparaisons post hoc	
Confusions lexicales					
Mots très fréquents			141***	Intermédiaires	Débutants
Avancés	1,2	1,1		***	***
Intermédiaires	3,2	1,4			***
Débutants	4,2	1,9			
Mots fréquents			37,1***	Intermédiaires	Débutants
Avancés	4,3	2,3		***	***
Intermédiaires	7,2	3,1			*
Débutants	6,2	3,2			
Mots moins fréquents			49,8***	Intermédiaires	Débutants
Avancés	4,8	2,5		*	***
Intermédiaires	5,5	2,6			***
Débutants	3,1	2,1			

*** $p < .001$, ** $p < .01$, * $p < .05$

Discussion

L'interférence des mots formellement semblables lors de l'identification lexicale en français L2 peut être expliquée par l'hypothèse de l'analyse contrastive et ce, soit sous sa forme traditionnelle de transfert direct L1–L2, soit sous sa forme élargie de transfert indirect L1–L2.

Depuis Lado (1957) et selon le point de vue traditionnel de l'analyse contrastive, nous savons que la difficulté d'apprentissage est directement liée à la différence entre L1 et L2. Conformément à ce point de vue, les apprenants hellénophones sont moins bien équipés que les apprenants anglophones pour réaliser des transferts de connaissance L1–L2 à cause de l'alphabet grec qui, d'une manière générale, obstrue le transfert des formes. Ainsi, lors d'une tâche d'identification lexicale en français l'apprenant hellénophone semble plus perturbé que son homologue anglophone à cause de l'étrangeté que représente pour lui le français sur le plan graphique. En ce qui concerne le transfert sur le plan phonétique, nous avons remarqué, plus particulièrement, que l'opposition /s/–/ʃ/ étant inexistante en grec, les hellénophones confondent systématiquement, contrairement aux anglophones, *tasse–tâche*, *semaine–chemin*, *cacher–casser*, *choix–soie*, *toucher–tousser*, *chapeau–sapin*⁴. Les hellénophones remplacent ainsi un élément difficile de la L2, le phonème /ʃ/, par l'élément de

leur L1 qui lui est le plus proche, le phonème /s/. En ce qui a trait au transfert sur le plan lexical, l'existence des congénères aussi bien que celle des faux amis épargne nombre de confusions aux anglophones comme *toucher-tousser*, *forcé-foncé*, *placer-plancher*, *fiancé-foncé*, *cache-casser* qui sont, par contre, fréquentes chez les hellénophones. L'existence en anglais de *touch*, *force*, *place*, *fiancé*, *cash* semble agir comme garde-fou en protégeant ainsi l'apprenant anglophone contre ces confusions. Ce résultat mériterait, nous semble-t-il, des analyses subséquentes.

L'explication qui tient compte du transfert indirect L1-L2 est à examiner aussi. Ici il convient d'apporter un éclaircissement. Alors que par transfert direct L1-L2 on entend généralement la transposition d'un élément de la L1 dans la L2, le transfert indirect rend compte de la transposition dans la L2 d'une attitude ou d'une habitude dictée par la L1. Pour ne parler que des mots monomorphémiques (comme le type des mots que nous avons testés dans l'ensemble) ceux que l'on peut raisonnablement appeler « confusibles formels » se limitent, en grec, à quelques paires lexicales généralement bien connues. On distingue ainsi a) certains parophones qui se différencient par l'accent tonique (ex. γέρος [ˈjɛros] 'vieux', γρός [jeˈros] 'fort'); b) certains homophones qui se différencient à l'écrit par un graphème (ex. κλίμα [klima] 'climat', κλήμα [klima], 'vigne') (cf. Triandafyllidis, 1963). En anglais, par contre, le nombre de confusibles formels semble beaucoup plus important comme en témoignent les compilations dans différents dictionnaires conçus spécifiquement pour les locuteurs natifs (cf. Room, 1979; Carpenter, 1993). De plus, ces confusibles peuvent avoir les formes les plus diverses, qu'ils soient homophones (*storey-story*, *already-all ready*, *slay-sleigh*, *principle-principal*) ou parophones (*live-leave*, *canal-channel*, *civic-civil*, *detract-distract*, *envelop-envelope*, *cave-cavern*, *lay-lie*, *quiet-quite*, *raise-rise*, *suit-suite*). De ce fait, nous pouvons supposer que les anglophones, habitués dans leur L1 à faire attention aux formes lexicales afin d'éviter d'éventuelles confusions, transfèrent cette habitude lors une tâche d'identification lexicale en L2 et obtiennent ainsi de meilleurs résultats. D'autre part, des confusions comme *boucher* pour *bouche*, *lycée* pour *lisse*, *meublé* pour *meuble* traduisent chez les hellénophones le transfert indirect d'une habitude de décodage en grec où les voyelles graphiques en fin de mots sont toujours prononcées.

La tendance des hellénophones à commettre plus de confusions formelles peut également être expliquée par ce que Corder (1979) décrit comme la distance entre L1-L2 et « l'étendue de la tâche d'apprentissage ». Selon Corder, tous les apprenants suivent, indépendamment de leur L1, la même séquence de développement quand ils apprennent une L2. Cependant là où la L2 présente des similitudes avec la L1 de l'apprenant le progrès sera plus rapide. Selon cette théorie, les anglophones surmontent leurs difficultés de fixation des formes lexicales en français L2 plus rapidement que les hellénophones. Ces

derniers ont besoin de plus de temps pour accomplir les mêmes tâches avec succès. Comme le dit Corder (1979, p. 33), l'effet facilitateur de la L1 dans l'apprentissage d'une L2 ne se manifeste pas dans le transfert des éléments de la L1 vers l'interlangue de l'apprenant mais dans la rapidité avec laquelle ces éléments seront découverts par celui-ci dans la langue qu'il est en train d'apprendre. Quelle que soit l'explication que l'on préfère, le point de vue de l'analyse contrastive ou celui de Corder, le fait est que la distance entre L1-L2 affecte la tâche de l'apprenant et dans notre cas la capacité de fixer correctement la forme des mots en L2. Ce résultat corrobore celui avancé par Laufer (1991). Il reste pourtant, du moins en partie, liée à la spécificité des mots testés. Ultérieurement il serait intéressant de tenir également compte de mots composés et de mesurer le rôle de la préfixation et de la suffixation dans les confusions formelles.

En ce qui concerne le niveau de compétence en L2 et conformément à nos résultats, la tendance à commettre des erreurs de confusion, et plus particulièrement des erreurs formelles, est moindre chez les apprenants de niveau avancé que chez ceux de niveau intermédiaire et débutant, et ce même si on isole la L1. Ce résultat, significatif à $p < .001$, est tout à fait normal. Ainsi, étant donné que l'acquisition du vocabulaire est un processus graduel, il est raisonnable de penser que plus l'apprenant devient compétent en L2, moins il commet d'erreurs de confusion. La tendance des apprenants de niveau intermédiaire à commettre plus d'erreurs de confusion que les apprenants de niveau débutant peut paraître surprenante étant donné leur maîtrise supérieure de la L2. Nous pouvons toutefois penser que si un certain niveau de connaissances lexicales est la condition nécessaire pour commettre des erreurs de confusion — car, comment confondre deux mots entre eux si on ne connaît ni l'un ni l'autre ? — les intermédiaires courent forcément plus de risques que les débutants. Par ailleurs, nous avons observé que les intermédiaires ont tendance ($p < .001$) à commettre plus d'erreurs de confusion sur des mots moins fréquents contrairement aux débutants qui, eux, ont tendance à commettre plus d'erreurs de confusion sur des mots très fréquents (cf. tableau 6). Il s'ensuit que le niveau intermédiaire est un moment plus critique par rapport aux premiers stades d'apprentissage d'une L2 où peu de mots sont connus.

Erreurs de confusion et reconnaissance des mots

On peut se demander dans quelle mesure la reconnaissance des mots est compatible avec la question des confusions lexicales, étant donné que ces dernières sont des erreurs. Cependant, étant donné que certains traits formels et/ou sémantiques du mot cible sont préservés dans l'erreur, on peut penser que les confusions sont des reconnaissances lexicales relativement réussies. Par reconnaissance lexicale, on entend généralement en psycholinguistique a) la locali-

sation de la forme lexicale dans le lexique mental⁵ ou « identification lexicale » ; b) l'accès aux différents types d'informations (sémantiques et syntaxiques) associées à cette forme lexicale ou « accès lexical ». Généralement on suppose que l'identification lexicale précède l'accès lexical mais le cas contraire est aussi envisageable (cf. Frauenfelder, 1991, p. 10). À partir de nos données, nous tenterons de montrer à quelle étape (« identification lexicale » ou « accès lexical ») se situe l'erreur selon le type de confusion lexicale impliqué.

Dans les confusions formelles, l'erreur se situerait à la première étape de la reconnaissance : celle de l'identification lexicale. Devant *cadeau*, les formes *gâteau*, *chapeau* ou *manteau*, déjà existantes dans le lexique mental de l'apprenant, peuvent interférer lors de l'identification grapho-phonémique de *cadeau* (si ce dernier n'est pas correctement stocké dans le lexique mental). L'interférence de *gâteau*, *chapeau* ou *manteau* risque d'avoir lieu même si *cadeau* est complètement inconnu et donc inexistant dans le lexique mental ; ici aussi l'erreur se situerait à l'étape de l'identification lexicale mais dans ce cas elle serait uniquement due au stockage déficient de *gâteau*, *chapeau* ou *manteau*.

Dans les confusions sémantiques, l'erreur se situerait à la deuxième étape : celle de l'accès lexical. Devant *dépenser*, que l'apprenant a correctement décodé en tant que D+É+P+E+N+S+E+R (étape d'identification lexicale), il commet une erreur lors de l'accès lexical en associant à cette forme l'information sémantique « vendre » qui appartient à la forme *vendre*. Contrairement aux confusions formelles, les confusions sémantiques indiquent logiquement que les mots cibles font partie du lexique mental (ils ne sont donc pas vus pour la première fois), sauf qu'ils sont mal stockés du point de vue sémantique.

Dans les confusions sémantico-formelles directes (*prêter* pour *emprunter*), l'erreur se situerait soit à la première étape (on retombe ainsi dans le cas des confusions formelles, si l'identification lexicale a lieu avant l'accès lexical), soit à la deuxième étape (on retombe ainsi dans les confusions sémantiques, si l'accès lexical a lieu avant l'identification lexicale). Enfin, dans les confusions sémantico-formelles indirectes mettant en jeu un troisième mot (ex. *pull* pour *chemin* via *chemise*), l'erreur se situerait successivement aux deux niveaux. Ainsi il y aurait d'abord erreur à l'étape de l'identification lexicale (*chemise* pour *chemin*) et ensuite il y aurait erreur à l'étape de l'accès lexical (« pull » pour « chemise »). Comme dans le cas des confusions formelles, le mot cible (*chemin*) peut ou non faire partie du lexique mental. Si oui, il est vraisemblablement mal stocké ; si non, c'est le mot avec lequel on l'a identifié qui est mal stocké.

Confusions formelles et stratégies d'identification lexicale

Il ne sera question ici que de la première étape de la reconnaissance lexicale, à savoir l'identification lexicale. Il existe deux façons principales d'identification lexicale à partir d'un stimulus écrit : a) la route visuelle directe (ou route lexicale) ; b) la route phonologique (ou route non lexicale) (cf. Kolinsky *et al.*, 1991). Devant un mot écrit, l'apprenant a donc le choix a) de passer directement à son identification à partir de ses indices visuels ; b) de passer d'abord par une médiation (ou représentation) phonologique de ses indices visuels. Diverses stratégies d'identification lexicale se dégagent à partir de nos données.

Tout d'abord les apprenants semblent s'appuyer sur les parties initiales du mot. Ainsi, dans 75 % des cas l'élément initial (graphème et/ou phonème) est commun au mot cible et au mot erroné. Ce pourcentage atteint 80 % si l'on ajoute les cas dans lesquels mot cible et mot erroné sont de phonème initial voisin (ex. /p/-/b/ « poison » pour *boisson*, /g/-/k/ « gâteau » pour *cadeau*, /ʃ/-/s/ « choix » pour *soie*). Une autre stratégie consiste à s'appuyer sur la désinence du mot. Nous avons ainsi remarqué qu'une terminaison faussement transparente peut amener l'apprenant à appliquer l'approche analytique⁶ et à penser, à tort, que le mot est, par exemple, un verbe, comme dans le cas de *cuir* (traduit par « cuire »). Les apprenants semblent également se fier à la longueur des mots. Nous avons ainsi vu que dans 69 % des cas le nombre de syllabes est le même entre le mot cible et le mot erroné. Ce pourcentage augmente si l'on tient compte du fait que parmi les confusions formelles produites, mot erroné et mot cible, bien qu'ils ne partagent pas le même nombre de syllabes, sont visuellement de la même longueur, étant donné certaines redondances orthographiques du français (*bâtir-battre*, *foncé-force*, *peigner-peindre*, *piéton-pointe*). Une stratégie heuristique importante semble être la rime sonore : mot erroné et mot cible riment ainsi littéralement comme dans *alumette-lunette* [alyɛt]-[lyɛt], *semelle-jumelle* [səmɛl]-[ʒymɛl] dans 64 % de nos confusions formelles. Nous avons distingué plus particulièrement les rimes pauvres (*foie-joie*), les rimes suffisantes (*image-visage*) et les rimes riches (*semelle-jumelle*). Les apprenants font ainsi appel à la représentation phonologique qu'ils ont d'un mot donné. Mais la stratégie la plus importante de reconnaissance lexicale est la rime visuelle. Nous avons observé en effet que dans le mot erroné ce sont généralement les lettres plutôt que les sons du mot cible qui se répètent (*âne-aîné* [A-N-E], *allumette-alimentation* [A-L-M-E-T]). Les apprenants s'appuient ainsi sur l'aspect visuel des mots afin de les reconnaître. Cela est compatible avec les habitudes d'identification lexicale de ces apprenants dans leur L1, l'anglais ou le grec. Nous émettons ainsi l'hypothèse que les lecteurs de ces langues, qui sont complexes du point de vue des correspondances grapho-phonémiques, apprennent à s'appuyer sur les indices visuels (route directe d'accès lexical) plutôt que sur la représentation

phonologique (route indirecte d'accès lexical). L'appui sur l'aspect visuel des mots lors de l'identification lexicale peut alors être interprété comme un transfert de la L1 vers la L2, le français, qui est une langue d'orthographe lexicale complexe, elle aussi, et pour cela le passage par la route visuelle directe semble privilégié (cf. aussi Suarez et Meara, 1989). Selon une autre explication la rime visuelle serait plutôt une stratégie compensatoire à laquelle les apprenants ont recours afin de reconnaître un mot donné, à défaut de pouvoir se le représenter phonologiquement de façon rapide et automatique (cf. Singleton, 1994).

Dans les limites de cette étude ces questionnements restent ouverts. Nous pouvons cependant penser que les stratégies de reconnaissance lexicale de bas niveau (identification de la forme) auxquelles ont recours les apprenants anglophones et hellénophones de français L2, mettent en jeu aussi bien l'aspect visuel (cf. la stratégie de la rime visuelle) que la représentation phonologique (cf. la stratégie de la rime sonore) des mots en question, avec une préférence pour les indices visuels qui peut être considérée comme une stratégie compensatoire faute d'une représentation phonologique précise. Les cinq stratégies identifiées ci-dessus reflètent les cinq traits formels, à savoir l'élément initial, la catégorie grammaticale, le nombre de syllabes, la voyelle accentuée et la configuration graphique, d'après lesquels les mots seraient stockés dans le lexique mental aussi bien en L1 qu'en L2 (cf. Aitchison, 1994; Arnaud, 1999; Singleton, 2000; Théophanous, 2001b). Nous postulons ainsi un rapport entre le stockage et l'identification lexicale que nous résumons dans le tableau 7.

Tableau 7 : Rapport entre stockage et identification lexicale

Traits formels dans le lexique mental	Stratégies d'identification lexicale
Élément initial	→ Partie initiale du mot
Catégorie grammaticale	→ Partie finale du mot
Nombre de syllabes	→ Longueur du mot
Voyelle accentuée	→ Rime sonore
Configuration graphique	→ Rime visuelle

Implications pédagogiques

Plutôt que de considérer l'acquisition du vocabulaire comme allant de soi, il nous semble indispensable d'insister sur l'intérêt qu'il y a pour l'apprenant de pouvoir reconnaître les mots de la L2 de façon exacte et automatique (repérage de la forme et attribution du sens). Nous considérons nécessaire de familiariser les apprenants avec les sons nouveaux, les suites grapho-phonémiques typiques ou même probables de la L2 et ses régularités orthographiques.

Dès le niveau intermédiaire, l'enseignant devrait mettre les apprenants en garde contre les risques de la confusion lexicale et les aider plus particulièrement à faire la différence entre mots similaires. On peut étudier explicitement l'orthographe et la prononciation d'un mot. D'autres options consistent à visualiser la forme orthographique dans le but de s'en souvenir ou de créer une représentation mentale du son du mot (cf. Schmitt, 1997; Hulstijn, 1997). Comme le propose Bogaards (1994), pour fixer les oppositions phonologiques importantes, on peut donner un mot et demander si le suivant est identique ou différent: *pendre-peindre, chou-sou*. Pour fixer les correspondances entre son et graphie, on peut donner un mot par écrit et faire entendre soit ce mot, soit un autre qui lui est semblable. Insister sur la désambiguïsation des formes lexicales (orales et écrites) est d'autant plus important qu'un mot inconnu peut être confondu avec un mot connu qui lui ressemble.

Il se peut que la tendance née de l'approche communicative ait éloigné les enseignants des exercices systématiques de vocabulaire et les ait plutôt incités à l'enseigner au fur et à mesure des occurrences en pratiquant surtout l'inférence du sens à partir du contexte. Nous considérons qu'un retour aux exercices plus « traditionnels » est pleinement justifié. De plus, la recherche montre que la rencontre fortuite avec les mots est loin d'être suffisante pour qu'il y ait apprentissage (cf. Laufer et Nation, 1995; Paribakht et Wesche, 1998; Horst, 2000). Si des exercices ciblés et un renforcement de l'attention sur les mots sont essentiels pour que quelques traces de ces mots puissent s'inscrire dans la mémoire, cela ne signifie pas qu'il faille insister de façon ponctuelle sur la forme orale et écrite de chaque mot que l'on rencontre en salle de classe. Nous préconisons avant tout l'accent sur l'aspect acoustique et visuel des mots de base de la langue cible et ce, non seulement parce qu'ils sont nécessaires et comme tels ne peuvent pas être laissés pour compte, mais, en plus, parce qu'ils ouvrent la voie vers la maîtrise de la L2 en général. Le reste du temps, l'enseignant saura agir de façon *ad hoc* et saura attirer l'attention sur la forme de tel ou tel mot jugée difficile, inhabituelle ou insolite.

Notes

Je tiens à remercier les trois évaluateurs anonymes de la *Revue Canadienne de Linguistique Appliquée* pour leurs commentaires sur la première version de cet article. Un grand merci également à Monsieur Michel Brabant, statisticien au Centre de ressources pour la recherche académique de l'Université d'Ottawa.

¹ Ces formes sont d'ailleurs souvent délicates à distinguer pour les natifs eux-mêmes : la distinction entre *lie* et *lay* fait ainsi l'objet d'une entrée fort complexe dans Urdang (1988, pp. 200–201) qui écrit : « La confusion causée par ces mots est tellement courante qu'il est probable que ces mots fusionneront à un moment donné dans l'avenir. » Notre traduction.

- ² Entre guillemets est indiquée la traduction de la réponse anglaise ou grecque pour le mot testé qui, lui, figure en italique.
- ³ Nous devons ici un grand merci à Monsieur Gérard Kahn, lexicographe à Oxford University Press et didacticien du FLE, pour son aide lors du classement des erreurs.
- ⁴ Rappelons ici que l'identification écrite des mots n'exclut pas la subvocalisation, cf. Kolinsky *et al.*, 1991 *et infra* « Confusions formelles et stratégies d'identification lexicale ».
- ⁵ En psycholinguistique par « lexique mental » on entend le lieu où les individus emmagasinent leurs connaissances lexicales et où ils vont chercher leurs mots afin de réaliser telle ou telle tâche langagière aussi bien en production qu'en compréhension (Aitchison, 1994).
- ⁶ Par opposition à « approche holistique » où les mots sont reconnus en bloc.

Références

- Aitchison, J. 1994. *Words in the Mind : An Introduction to the Mental Lexicon*. Oxford, Blackwell.
- Arnaud, P.J.L. 1999. Target-error ressemblance in French word substitution speech errors and the mental lexicon. *Applied Psycholinguistics* 20, pp. 269–287.
- Bensoussan, M. 1992. Learners' spontaneous translations in an L2 reading comprehension task : Vocabulary knowledge and use of schemata. In P. Arnaud et H. Béjoint (dir.), *Vocabulary and Applied Linguistics*. London, Macmillan Academic and Professional Ltd., pp. 102–112.
- Bogaards, P. 1994. *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris, Crédif, Hatier-Didier.
- Carpenter, E. 1993. *Confusable Words*. London, Harper Collins.
- Corder, S.P. 1979. Language distance and the magnitude of the language learning task. *Studies in Second Language Acquisition* 1/2, pp. 27–36.
- Goodfellow, R., G. Jones et M.N. Lamy 2002. Assessing learners' texts using the lexical frequency profile, *ReCALL*, 14, pp. 129–142. Disponible à : http://www.er.uqam.ca/nobel/r21270/cgi-bin/F_webfreqs/glynn_jones.html.
- Gougenheim, G., P. Rivenc, A. Sauvageot et R. Michéa. 1964. *L'élaboration du français fondamental*. Paris, Didier.
- Frauenfelder, U. 1991. Une introduction aux modèles de reconnaissance des mots parlés. In R. Kolinsky *et al.* (dir.), *La reconnaissance des mots dans les différentes modalités sensorielles : études de psycholinguistique cognitive*. Paris, PUF, pp. 7–36.
- Hatch, E. et C. Brown 1995. *Vocabulary, Semantics and Language Education*. Cambridge, CUP.
- Horst, M. 2000. Incidental vocabulary learning : Patterns of growth. Communication présentée à l'Université d'Ottawa, Institut des langues secondes, 3 mars 2000.
- Hulstijn, J. 1997. Mnemonic methods in foreign language vocabulary learning : Theoretical considerations and practical implications. In J. Coady et T. Huckin (dir.), *Second Language Vocabulary Acquisition : A Rationale for Pedagogy*. Cambridge, CUP, pp. 203–224.

- Kolinsky, R., J. Morais et J. Segui (dir.). 1991. *La reconnaissance des mots dans les différentes modalités sensorielles : études de psycholinguistique cognitive*. Paris, PUF.
- Lado, R. 1957. *Linguistics across Cultures*. Ann Arbor, University of Michigan Press.
- Laufer, B. et P. Nation 1995. Vocabulary size and use : Lexical richness in L2 written production. *Applied Linguistics* vol. 16, n° 3, pp. 307–322.
- Laufer, B. 1991. *Similar Lexical Forms in Interlanguage*. Tübingen, Günter Narr.
- Levenston, E.A. 1979. Second language acquisition : Issues and problems. *Interlanguage Studies Bulletin-Utrecht* vol. 4, n° 2, pp. 147–160.
- Meara, P. 1980. Vocabulary acquisition : A Neglected aspect of language learning. In V. Kinsella (dir.), *Language Teaching Surveys I*. Cambridge, CUP, pp. 221–246.
- Paribakht, T.S. et M. Wesche. 1998. “Incidental” and instructed L2 vocabulary acquisition : Different contexts, common processes. Communication présentée au congrès de l’Association canadienne de linguistique appliquée, Ottawa, Carleton University, 20–23 mai 1998.
- Room, A. 1979. *Room’s Dictionary of Confusibles*. London, Routledge and Kegan Paul.
- Ryan, A. 1997. Learning the orthographical form of L2 vocabulary — A receptive and a productive process. In *Vocabulary : Description, Acquisition, and Pedagogy*. N. Schmitt et M. McCarthy (dir.) Cambridge, CUP, pp. 181–198.
- Ryan, A. et P. Meara. 1991. The case of the invisible vowels : Arabic speakers reading English words. *Reading in a Foreign Language* vol. 7, n° 2, pp. 531–540.
- Schmitt, N. 1997. Vocabulary learning strategies. In N. Schmitt et M. McCarthy (dir.), *Vocabulary : Description, Acquisition, and Pedagogy*. Cambridge, CUP, pp. 199–227.
- Singleton, D. 1994. Le rôle de la forme et du sens dans le lexique mental en L2. In D. Singleton (dir.), *L’acquisition du lexique d’une langue étrangère*. Paris, AILE, pp. 56–69.
- Singleton, D. 2000. Connectivité interlinguistique dans l’acquisition et l’exploration du lexique de la langue seconde. *Le Français dans le monde — Recherches et applications*, juillet 2000, pp. 56–69.
- Suarez, A. et P. Meara. 1989. The effects of irregular orthography on the processing of words in a foreign language. *Reading in a Foreign Language* vol. 6, n° 1, pp. 349–356.
- Théophanous, O. 2001a. Confusions lexicales dans l’acquisition du vocabulaire d’une langue seconde/étrangère. Thèse de doctorat, Université de Montréal.
- Théophanous, O. 2001b. Déviations lexicales de forme et de sens chez les apprenants de français langue étrangère. *Revue française de linguistique appliquée*, vol. 6, n° 1, *Troubles langagiers*, sous la dir. de H. Huot, pp. 107–120.
- Triandafyllidis, M. 1963. *Œuvres complètes : recherches*. Tome I [en grec]. Thessaloniki, Université Aristote/Institut d’Etudes néohelléniques.
- Urdang, L. 1988. *Differences Over 5000 Commonly Confused Words Explained*. London, Bloomsbury.
- Verlinde, S. et T. Selva. 2001. Corpus-based versus intuition-based lexicography : Defining a word list for a French learner’s dictionary. In P. Rayson, A. Wilson, T. McEnery,

- A. Hardie and S. Khoja (dir., *Proceedings of the Corpus Linguistics 2001 Conference, Lancaster University*, pp. 594–598. Disponible à : <http://www.kuleuven.ac.be/grølep/publicat/verlinde.pdf>.
Conference Proceedings: <http://korpus.dsl.dk/c12003/cdrom/c12001/>.
- Warren, B. 1982. Common types of lexical errors among Swedish learners of English. *Moderna Språk* vol. 76, n° 3, pp. 209–228.
- Zimmermann, R. 1987. Form-oriented and content-oriented lexical errors in L2 learners. *International Review of Applied Linguistics in Language Teaching* vol. 25, n° 4, pp. 55–67.
- Zughoul, M.R. 1991. Lexical choice : Towards writing problematic word lists. *International Review of Applied Linguistics in Language Teaching* vol. 29, n° 1, pp. 45–60.