

# La compétence lexicale en production orale

HILTON Heather

Laboratoire LLS, Université de Savoie

---

Quels sont les marqueurs discursifs spécifiques pour différents niveaux de compétence en langue orale ? De nombreux travaux en acquisition et en didactique des L2, inscrits dans un cadre théorique chomskyen de la « compétence » linguistique, ont pris pour objet essentiel d'étude l'ordre d'acquisition de morphèmes et de structures morpho-syntaxiques dans diverses langues et se sont beaucoup moins penchés sur une éventuelle progression dans la compétence lexicale des apprenants. Le corpus PAROLE, constitué à l'Université de Savoie, est conçu pour comparer les productions orales d'apprenants de différents niveaux de compétence, dans différentes L2. Nous limitant, pour cette étude, à des apprenants francophones de l'anglais L2, nous présenterons les caractéristiques lexicales de leurs productions à deux niveaux de maîtrise : un niveau élémentaire (palier européen A1-A2) et un niveau avancé (C1-C2). Différentes mesures de la compétence lexicale seront rapidement présentées. Avec Meara (1996) et Ellis (2000), nous postulons que certains aspects de la compétence lexicale sont des indicateurs fiables du niveau de maîtrise de la L2 : la diversité lexicale indique l'étendue des connaissances mobilisées (et mobilisables) dans la production ; divers phénomènes de disfluente dans la parole (pauses, hésitations, reprises, répétitions ; Griffiths 1991) révèlent l'aisance avec laquelle on peut accéder à ce réseau mnésique.

---

Après une « période faste » dans les années 1950 et 1960 (Arnaud 1985, p. 72), et une période d'oubli relatif (entre 1985 et 2000), la question de l'acquisition lexicale revient au goût du jour – dans le domaine de l'acquisition du français langue maternelle, surtout, mais également en didactique des L2<sup>1</sup>. Il était temps, car en psycholinguistique de la L1 on sait depuis plus de soixante ans que la compétence lexicale est un facteur déterminant de la performance verbale (Smith 1941, cité par Nagy & Herman 1987, p. 23 ; pour une étude récente à grande échelle, voir Hart & Risley 1995). Dans le domaine de l'acquisition et la didactique des L2, nous n'avons pas suffisamment examiné le rapport entre les connaissances et compétences lexicales des apprenants, et leur niveau de maîtrise et d'aisance communicationnelles.

Nous proposons donc dans cette étude de mener une partie de cette investigation, en examinant les caractéristiques lexicales des productions orales de deux groupes d'apprenants de niveau de compétence différent. Le premier groupe de dix sujets francophones a, après 7,2 années d'étude de l'anglais L2 en milieu scolaire, un niveau élémentaire en compréhension de l'oral (A1 ou A2, selon l'échelle du *CECRL*<sup>2</sup>) ; pour éviter de leur coller l'étiquette disgracieuse de « faibles », nous parlerons du groupe *disfluent*, le taux d'hésitation dans leurs productions atteignant

---

<sup>1</sup> Par exemple, Calaque & Grossman 2000 ; Grossman *et al.* 2005 ; de Robien 2006. En acquisition des L2 en dehors de la France, la recherche en acquisition lexicale est une véritable industrie, avec des centaines de publications par an. Pour avoir une idée du volume d'ouvrages produits, voir *The Vocabulary Research Group Archives* de l'University of Wales Swansea, <http://www.swan.ac.uk/cals/calsres/varga/index.htm>.

<sup>2</sup> *Cadre européen commun de référence pour les langues* ; les sujets ont passé les tests de compréhension en ligne proposés par le consortium européen DIALANG : <http://www.dialang.org>.

en moyenne 50,2% du temps total de prise de parole. En net contraste, notre deuxième groupe '*fluent*' est composé de dix apprenants – parlant couramment l'anglais (avec un taux moyen d'hésitation de 27,8%), et possédant un niveau avancé (C1 ou C2) en compréhension de l'oral. Quatre sujets dans ce groupe sont des francophones préparant le concours d'entrée au corps professoral du secondaire en anglais, et six sont germanophones (linguistes se trouvant à l'Université de Savoie dans le cadre d'un échange européen).

### **Le corpus PAROLE**

Les productions que nous allons analyser ici proviennent du Corpus PAROLE (PARAllèle Oral en Langue Étrangère), actuellement en cours de transcription à l'Université de Savoie. Pour constituer ce corpus, les participants ont effectué cinq tâches courtes de production orale : la description de trois séquences vidéo (deux résumés formulés après le visionnement et un commentaire en cours de visionnement), une tâche les contraignant à produire des phrases complexes (tirée de la batterie CELF, Semels *et al.* 1994), et un simple récit autobiographique. Les sujets étaient reçus individuellement pour les tâches de production ; le protocole relatif à la collecte de données insistait sur une intervention minimaliste de l'interlocuteur et les productions sont donc assez proches du monologue. Nous avons opté pour cette limitation artificielle de l'interactivité en situation de face-à-face, car pour les besoins de comparaison d'un corpus parallèle nous voulions générer des productions comparables, de taille raisonnable, relativement dépourvues des complexités interactionnelles qui rendent la transcription et le calcul des temps de parole difficiles. L'objectif global du projet étant l'identification et la comparaison de ce qui caractérise différents niveaux de compétence productive dans les L2 du projet, nous nous penchons non pas sur les mécanismes interactionnels de la compétence, mais plutôt sur les caractéristiques linguistiques (phonologiques, lexiques, morphologiques, syntaxiques) qui semblent déterminer l'aisance de la production orale.

En plus des tâches de production, nous avons demandé à nos sujets<sup>3</sup> de compléter une batterie de tests et de remplir deux questionnaires, afin d'obtenir des renseignements sur certaines variables individuelles. Un questionnaire sur le profil linguistique de chaque participant nous permet de déterminer le nombre d'années d'étude de la L2, le temps passé (éventuellement) dans un pays où la L2 est parlée, la note au Baccalauréat dans la langue, etc. Un questionnaire de motivation (adapté de Gardner 2004) fournit un score qui reflète l'engouement de chaque apprenant pour la langue. En anglais, les sujets ont également passé six tests : trois tirés des instruments diagnostiques mis en ligne sur le site européen de DIALANG (le test de compréhension de l'oral déjà cité, ainsi qu'un test de grammaire et de vocabulaire), un test estimant l'étendue du lexique en anglais (Hever sans date), un test d'aptitude à l'analyse grammaticale (Meara *et al.* 2001) et un test de répétition de pseudomots, administré d'abord en L1 français<sup>4</sup> (Casalis 2003), ensuite en anglais (adapté de Gathercole & Baddeley 1996). Ce dernier, essentiellement un test de la capacité phonologique en mémoire à court terme, fournit une mesure qui s'est révélée pertinente dans l'aptitude à la mémorisation de nouveaux mots en L2 (Gathercole & Thorn 1998 ; Papagno *et al.* 1991).

---

<sup>3</sup> Chaque participant fut rémunéré (au SMIC) pour les trois heures qu'il a consacrées au projet PAROLE.

<sup>4</sup> Les sujets germanophones n'ont pas passé ce test en L1.

## Les analyses

Pour ces premières analyses des productions regroupées dans le corpus PAROLE, nous nous limiterons aux deux tâches déjà complètement transcrites : les deux résumés de séquence vidéo. Les séquences sélectionnées comme supports aux productions l'ont été pour différentes raisons : la première séquence pour les défis linguistiques qu'il soulève : l'objet saillant de la séquence (un réfrigérateur), d'abord patient d'une action, devient l'agent d'une catastrophe domestique (soulevé par une grue, le frigo glisse de ses cordes, et tombe sur une voiture stationnée dans la rue en dessous). La deuxième séquence provoque un défi au niveau de l'organisation discursive, car cette publicité télévisée met en scène un homme, d'abord pendant son enfance, et ensuite, en tant qu'adulte, vivant des événements au présent. Les résumés produits par nos sujets suite au visionnement de ces deux séquences constituent un mini-corpus d'environ 4700 mots, avec deux à trois minutes de production par sujet. Nos enregistrements numériques sont transcrits à l'aide des outils proposés par le *Child Language Data Exchange System* (MacWhinney & Spektor 2003-2007), selon les conventions de transcription CHAT. Le logiciel CLAN dispose de fonctionnalités intéressantes pour la transcription d'enregistrements numériques : en « mode sonique » (simple option d'affichage, obtenue en cliquant dans un menu déroulant), le transcripteur voit, en bas de l'écran, la courbe acoustique du fichier son. Il peut facilement poser des balises temporelles dans le fichier, ainsi reliant chaque énoncé de la transcription au segment sonore qui y correspond (figure 1). L'affichage de la durée de tout segment sélectionné dans la courbe acoustique nous a permis de mesurer (en millisecondes) la longueur de toutes les hésitations produites par les sujets.

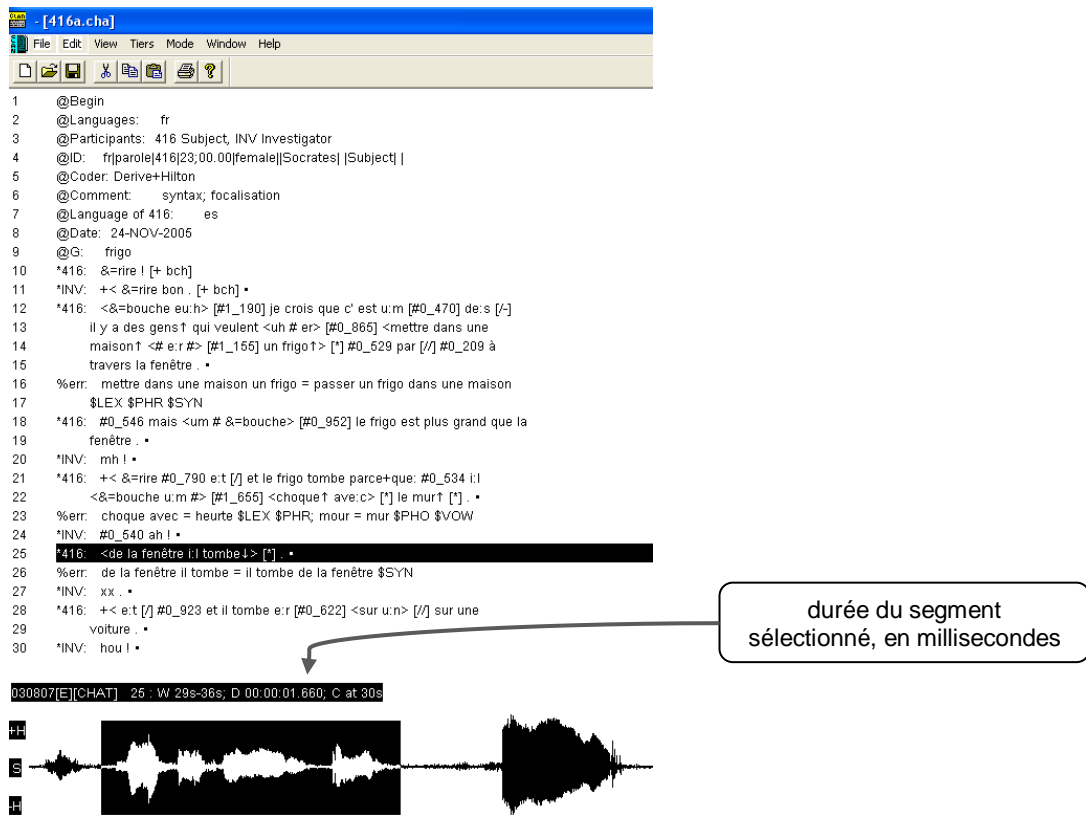


Figure 1 : visualisation (et chronométrage) d'un énoncé dans le « mode sonique » de CLAN

CLAN permet également la réalisation rapide et fiable de différentes analyses pertinentes à une appréciation de la compétence lexicale. On peut, avec le programme *mor*, générer un corpus étiqueté (*tagged*, en anglais), où la fonction syntaxique de chaque mot est automatiquement analysée et affichée ; le programme *post* permet de désambiguer intelligemment cet étiquetage initial (qui peut ensuite être vérifié rapidement par le chercheur). *Mor* génère une ligne « morphologique » pour chaque ligne de la transcription, qui permet d'effectuer des analyses sur les lemmes<sup>5</sup> ou les morphèmes du corpus original. Sur cette ligne, chaque mot de l'énoncé « *the elephant looked disappointed* » sera donc étiqueté : *det/the n/elephant v/look-PAST part/disappoint-PERF*. On imagine aisément l'intérêt d'une telle ligne pour des analyses morpho-syntaxiques ; pour des analyses lexicales, elle permet d'effectuer et de comparer différentes mesures de la richesse lexicale. Le rapport type-occurrence (RTO) compare le nombre de mots produits par le sujet (les occurrences) au nombre de mots différents utilisés (les types) – un rapport de 1 indiquant une extrême richesse lexicale<sup>6</sup>, où aucun mot n'est répété par l'interlocuteur. L'étiquetage du corpus permet de calculer le RTO sur les lemmes – les trois « occurrences » *walk*, *walked*, *walking* sont comptabilisés comme un seul « type ». Dans le RTO calculé sur les mots, ces trois occurrences morphologiques d'une même base lexicale sont comptabilisés comme trois types distincts ; les mesures de RTO basées sur un corpus lemmatisé donnent donc une mesure de la richesse lexicale plus fiable que les mesures basées exclusivement sur les mots (Nation 2001, p. 7). Ceux qui s'intéressent aux mesures de la richesse lexicale connaissent sans doute déjà la statistique « D » – mesure plus sophistiquée de la diversité lexicale, élaborée par une équipe de chercheurs à l'Université de Reading (Malvern & Richards 1997), qui donne une estimation probabilistique de la diversité lexicale, selon un algorithme d'échantillonnage appliqué au texte produit (pour une explication en français de l'utilité et la conception de D, voir Lété 2004). Le programme *vocd* de CLAN permet de générer cette mesure de la diversité lexicale automatiquement (prenant en compte la lemmatisation effectuée par *mor* et *post*) ; plus la valeur de D est élevée, plus la diversité lexicale d'une production est riche. Nous verrons, dans les comparaisons qui suivent, que la mesure D semble effectivement la plus fine des trois estimations de richesse lexicale effectuées sur notre corpus.

Directement ou indirectement, CLAN fournit des mesures plus générales de la productivité orale. En plus du calcul automatique des rapports type-occurrence, le programme *freq* permet le décompte du nombre et du type d'erreurs produites par chaque sujet (selon le codage préalable de ces erreurs par les transcripteurs), le nombre total de reprises, d'hésitations, de pauses vocalisées, etc. Le programme *mlu* indique le nombre d'énoncés produits par chaque sujet, et calcule la longueur moyenne des énoncés (en mots ou en morphèmes) – mesure classique du développement communicationnel dans le domaine de l'acquisition de la L1. Le programme *timedur* génère des fichiers qui peuvent être importés dans un tableur (comme *Open Office Calc* ou *Excel*), à partir desquels on peut facilement connaître

<sup>5</sup> Du point de vue psycholinguistique, les lemmes sont ces représentations lexicales – comprenant le sens et la fonction syntaxique du mot – qui sont actives avant la spécification morpho-phonologique dans la production orale (« lexical items unspecified for phonological form », Levelt 1992, p. 5). Du point de vue lexicologique, le lemme regroupe toutes les formes flexionnelles d'un même nom, verb, adjectif, ou adverbe (Nation 2001, p. 7).

<sup>6</sup> Sans doute impossible, vue la nécessaire récurrence des mots grammaticaux de la « classe fermée » dans toute production langagière – ces petits opérateurs (de nombre limité) que sont les déterminants, les pronoms, les prépositions, les conjonctions et les auxiliaires.

le temps de parole de chaque participant dans une transcription balisée. Cette mesure de temporalité sert ensuite de base aux calculs d'aisance productive tels le nombre de mots par minute (taux de parole), les taux d'erreur ou d'hésitation. Moins automatique, mais tout aussi important pour les considérations qui vont suivre, nous avons calculé le temps total d'hésitation pour chaque sujet, la longueur moyenne d'hésitation, la localisation de chaque hésitation, et le pourcentage du temps de production passé en hésitation.

### Les résultats

Quand nous confrontons les mesures des différents paramètres verbaux pour nos deux groupes, ainsi que leur performance aux tests complémentaires, nous trouvons des différences hautement significatives dans presque tous les domaines. Ces résultats, avec leur significativité statistique (comparaisons non-paramétriques, test de Mann-Whitney<sup>7</sup>), sont regroupés dans le tableau 1.

mesure	moyenne, groupe <i>disfluent</i>	moyenne, groupe <i>fluent</i>	Mann-Whitney	significativité
Test de compréhension de l'oral DIALANG	14,2	28,8	Ws(10,10)=55	p<,001***
Test estimant l'étendue du vocabulaire en anglais <sup>8</sup>	2835	7200	Ws(10,10)=57	p<,001***
Test de grammaire DIALANG (score/ 30)	23.3	28.7	Ws(10,10)=55	p<,001***
Test de répétition de pseudomots, L1 (score/ 35)	32,4	30	Ws(4, 10)=16,5	p<,20 (n.s.)
Test de répétition de pseudomots, L2	24,2	26,6	Ws(10,10)=82,5	p<,05*
Test d'aptitude à l'analyse grammaticale	65,5	67	Ws(10,10)=107,5	p<,40 (n.s.)
Longueur moyenne des énoncés	7,9	11,2	Ws(10,10)=70	p<,01**
Mots par minute	65,3	102,1	Ws(10,10)=56	p<,001***
Rapport type-occurrence (RTO), mots	0,391	0,463	Ws(10,10)=70	p<,01**
RTO, lemmes	0,437	0,493	Ws(10,10)=76	p<,05*
Diversité lexicale D	36,66	58,79	Ws(10,10)=56	p<,001***
taux d'erreurs (erreurs par énoncé)	1,84	0,77	Ws(10,10)=55	p<,001***
% d'hésitation	50,3	27,8	Ws(10,10)=58	p<,001***
longueur moyenne d'hésitation (en secondes)	1,890	0,690	Ws(10,10)=57	p<,001***

**Tableau 1 : comparaisons entre les groupes, tests complémentaires et mesures productives**

Lors de la table ronde du colloque de septembre 2006, la question fut posée de l'utilité d'une telle série de chiffres – apparemment loin des réalités de l'interaction

<sup>7</sup> Nous remercions Christel Leuwers, du laboratoire de psychologie expérimentale à l'Université de Savoie, pour son aide précieuse dans l'analyse statistique de ces données.

<sup>8</sup> Ce test donne une estimation du nombre de mots connus en anglais (jusqu'à 20.000 mots).

humaine – dans l'étude de l'acquisition ou de la didactique des langues. Malgré nos préoccupations pour les aspects sociaux de la communication, nous ne devons pas, dans nos recherches scientifiques, négliger la puissance de l'analyse statistique. N'importe quel enseignant de langues doit régulièrement se poser la question des mécanismes exacts de l'amélioration des compétences : en quoi faut-il qu'un apprenant s'améliore, s'il veut passer d'un stade de compétence donné au stade supérieur ? Les descripteurs du *CECRL* nous indiquent les performances qui caractérisent chaque niveau de référence, mais laisse aux « utilisateurs » du *Cadre* le soin de définir ce qui peut amener un apprenant du palier A2 (« peut se faire comprendre, dans une brève intervention même si la reformulation, les pauses et les faux démarrages sont évidents ») au niveau B2 (« peut parler relativement longtemps avec un débit assez régulier » ; descripteurs de l'aisance, *CECRL*, p. 28). Dans notre tableau de résultats nous voyons – clairement et objectivement – que différents aspects de la compétence lexicale distinguent bel et bien les apprenants de niveau avancé de ceux qui semblent figés au niveau élémentaire (après plus de sept années d'étude de la langue, rappelons-le). Ceux qui ont du mal à prendre la parole, dans une simple tâche de description, ont des connaissances linguistiques limitées dans la L2 : un vocabulaire réduit (moins de 3000 mots<sup>9</sup>), un niveau B1 en connaissances grammaticales, une mémoire phonologique limitée pour les sons de la L2 (alors qu'ils sont légèrement supérieurs au groupe fluent en ce qui concerne la mémoire phonologique en L1). Ce manque de connaissances linguistiques est sans doute à l'origine d'une bonne partie de la variance constatée entre les productions de ces deux groupes : diversité lexicale moindre chez les disfluents, forts taux d'hésitation et d'erreur ; un taux de production nettement inférieur et un taux d'hésitation nettement supérieur à ceux du groupe fluent.

### L'aisance orale

Ces constats d'ordre quantitatif autour du lexique dans le corpus nous ont incités à regarder de plus près un éventuel lien entre connaissances lexicales et hésitations dans le corpus. Lors de la transcription des productions, nous avons systématiquement codé toutes les hésitations supérieures à 200ms – pauses silencieuses ou vocalisées, ainsi que tout phénomène paralinguistique les accompagnant (soupirs, claquements de langue, etc.) – selon les conventions de CHAT. Les pauses vocalisées sont transcrites par « uh », « um » (ou « euh », dans le cas d'une vocalisation francophone plus fermée) ; les répétitions, reprises et faux démarrages sont indiqués par des combinaisons convenues du symbole « / » ; les fragments et bégaiements par le symbole « & », les pauses par un dièse. D'après les mesures effectuées sur la courbe acoustique de CLAN (figure 1), nous avons pu inclure sur la ligne principale de transcription la valeur temporelle de chaque hésitation : une pause de 1350 millisecondes de durée étant transcrite « #1\_350 » (avec un trait de soulignement séparant les secondes des millisecondes). Constatant l'existence de groupes complexes d'hésitation dans ces productions en L2, où pauses silencieuses alternent avec pauses vocalisées et autres bruitages, nous avons décidé de coder ces regroupements selon les normes de CHAT, les enfermant entre crochets, suivis du temps total d'hésitation. Dans cet exemple, tiré d'une production disfluente, il y a quatre groupes d'hésitation, figurant en caractères gras :

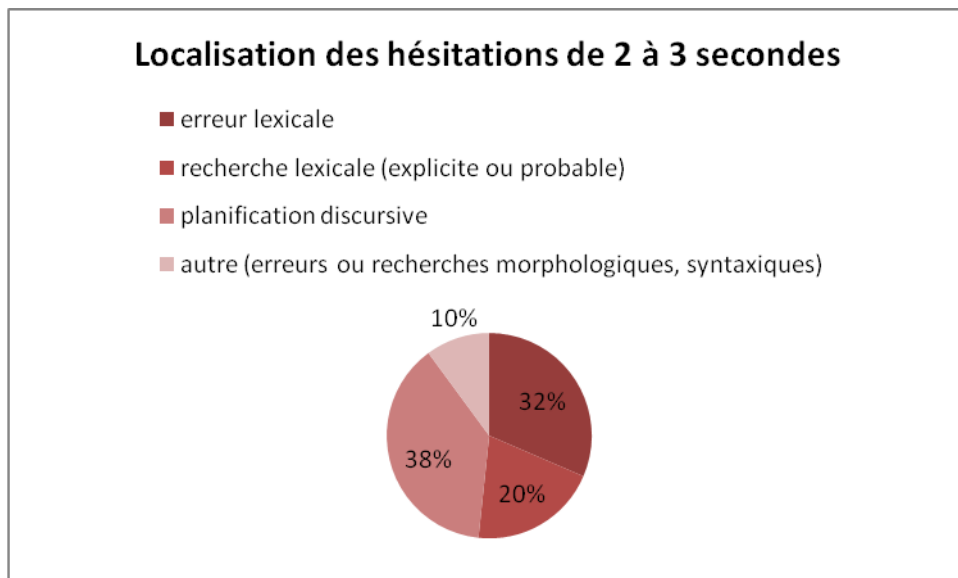
and <u:m #> [**#1\_904**] (h)e: <uh #> [**#1\_939**] he give <uh #> [**#1\_445**] to the man <# uh # uh &=bouche #> [**#6\_109**] +... (énoncé abandonné).

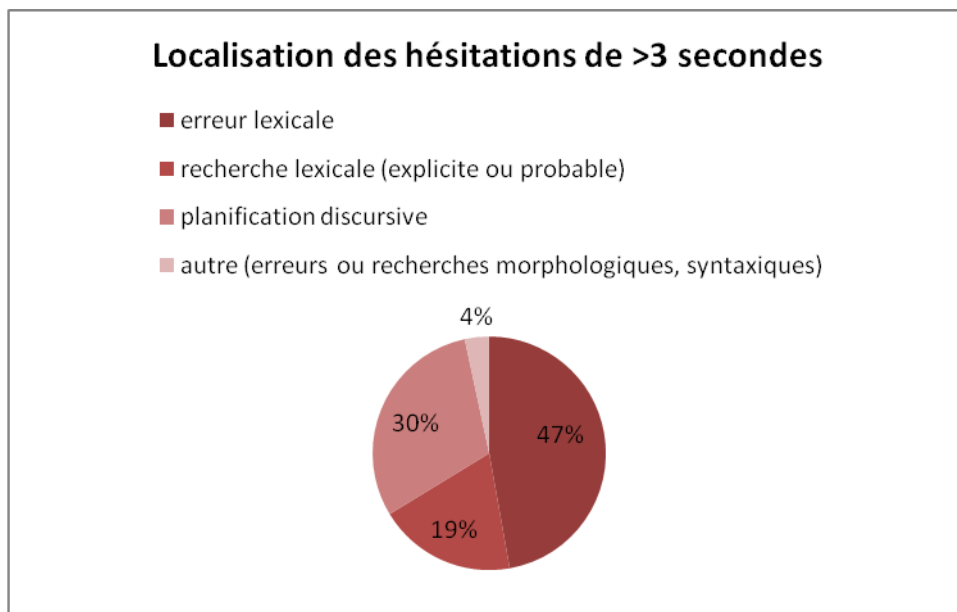
<sup>9</sup> Le seuil estimé de l'aisance communicative de base étant fixé vers 5000 mots (Laufer 1997, 23-24).

Cette transcription minutieuse du temps d'hésitation nous a permis (après l'importation dans un logiciel-tableur des fichiers générés par *freq* dans CLAN) d'effectuer les indices d'hésitation figurant dans les deux dernières lignes du tableau 1 (pourcentage, et longueur moyenne d'hésitation).

Des études de l'aisance (*fluency*) en production orale en L1 fixent quelques repères de base : le débit normal est de 130 à 170 mots par minute (2-3 mots par seconde, Levelt 1989 : 22) ; l'hésitation occupe environ un tiers du temps de parole, variant un peu selon la complexité de la tâche (Good & Butterworth 1980). Pour la production normale en L1, nos hésitations intra-propositionnelles dépassent rarement 500ms (les phrases complexes pouvant comporter des pauses plus longues) ; les pauses inter-propositionnelles dépassent rarement 2 secondes. Après 3 secondes, l'hésitation devient silence, et l'interlocuteur intervient pour remettre la communication sur les rails. (Ces chiffres concernent l'anglais, principalement, et sont tirés de Kowal & O'Connell 1980 ; Beattie 1980 ; Pawley & Syder 1983 ; Tauroza & Allison 1990 ; Griffiths 1991 ; Riegenbach 1991 ; Rieger 2003).

Prenant donc le chiffre de deux secondes comme un seuil ordinaire d'hésitation dans des productions de niveau expert, nous avons manuellement codé l'emplacement et la durée de toutes les pauses supérieures à deux secondes dans les productions de nos vingt sujets (178 pauses et groupes d'hésitations de deux secondes et plus au total, avec 89 hésitations entre deux et trois secondes de durée, et 89 hésitations de plus de trois secondes). Toute hésitation en début d'énoncé ou entre deux propositions fut considérée comme une pause de planification – il y en a 60 au total. Les 128 autres hésitations se trouvent à l'intérieur d'une proposition ; leur emplacement exact fut codé, et le mot (ou les mots) se trouvant immédiatement après la pause furent analysés. Les résultats de ces analyses détaillées sont résumés dans les graphiques de la figure 2.





**Figure 2 : localisation des hésitations disfluentes, corpus PAROLE**

On voit ici que plus de la moitié des hésitations longues de notre corpus sont directement liées à la compétence lexicale. 32% des pauses de 2 à 3 secondes et 47% des pauses de plus de 3 secondes de durée sont suivies directement par une erreur lexicale. Les « recherches lexicales » qui expliquent environ 20% des pauses disfluentes sont soit totalement explicites (le sujet déclarant qu'il ne trouve pas ou ne connaît pas un mot, ou demandant directement à l'investigateur de le lui fournir), soit extrapolées mais hautement probables – la pause étant suivie d'un mot correct de fréquence basse (*parade, crane*), par un hyperonyme à la place d'un mot plus précis mais moins fréquent (*animal* pour *elephant, nose* pour *trunk*), ou par une paraphrase. A titre d'illustration, voici un exemple de chaque type de localisation<sup>10</sup> :

**Pause discursive** (début/ fin d'énoncé ou de proposition) :

\*033: <&=bouche # um # er> [#2\_379] he's watching a parade .

**Erreur lexicale** (immédiatement après l'hésitation) :

\*012: uh they: see <uh # u:h> [#3\_477] frigo@n [\*] .

**Recherche lexicale** (explicitée) :

\*002: a fridge #0\_511 which <u:h # &=bouche #> [#7\_664] [/ -] (be)cause I [/] I don't know uh how [/] #0\_383 how we say uh monter@s .

**Recherche lexicale** (extrapolée ; paraphrase pour *parade*) :

\*019: uh he's watching a kind of <um # um> [#2\_241] party uh with elephant [\*] and something like a circus .

**Autre**: (erreurs syntaxiques, morphologiques, etc.)

\*006: in the first parts [\*] um there is a #3\_199 children [\*] .

<sup>10</sup> Afin que ces extraits de transcriptions soient plus lisibles, nous précisons que : le suffixe « @n » est rajouté aux créations lexicales (néologismes) ; « @s » indique un mot articulé dans une langue autre que la L2 ciblée ; l'ensemble « [\*] » est placé directement après toute erreur ; l'ensemble « [/] » marque une reprise ou reformulation ; l'ensemble « &=bouche » désigne un claquement de langue ou tout autre bruitage para-linguistique.



Nous penchant plus particulièrement sur les hésitations intra-propositionnelles de plus de trois secondes, nous constatons un phénomène intéressant et sans doute important : 84% de ces hésitations longues se trouvent directement devant un nom ou un verbe ; la recherche lexicale portant sur le verbe semble compromettre un peu plus sérieusement la fluidité de la prise de parole, car les hésitations longues ont une durée moyenne de 3,360 secondes devant un nom manquant, et de 4,690 secondes devant un verbe. 74% des hésitations de plus de six secondes se trouvent en position immédiatement pré-verbale – illustrant la difficulté de l’encodage quand l’énonciateur n’a pas un accès facile aux lexèmes qui vont déterminer les structures syntaxiques de l’énoncé.

### **Implications didactiques**

Il est important d’insister sur le fait que cette comparaison de quelques paramètres dans deux niveaux acquis en anglais L2 ne porte pas de jugement sur les apprenants participant au projet PAROLE. La référence à des indices d’aisance pour une population native ne veut en rien suggérer que nous prenons comme but ultime de l’acquisition ce même niveau d’aisance experte. Notre comparaison de deux groupes d’apprenants très différents est faite dans un but scientifique, pour voir si c’est possible d’identifier quelques facteurs contribuant à une plus grande aisance en production orale. La nature artificielle des données collectées – manque d’interactivité, tâches très contraintes – ne nous permet pas de nous prononcer sur les capacités interactionnelles des sujets. Mais cette même artificialité est utile dans l’exploration de certains aspects psycholinguistiques de l’aisance orale, qui seraient difficiles à mesurer dans la complexité et l’imbrication de l’interaction orale.

Dans un article décrivant l’infrastructure théorique « chancelante » du *CECRL*, Hulstijn (2007) insiste sur la nécessité d’une investigation scientifique des liens entre connaissances et processus linguistiques (*lower- and higher-order cognitive knowledge and skills*) d’un côté, et les descripteurs performatifs du *Cadre* de l’autre. Notre analyse des structures de disflueance dans un corpus oral révèle précisément l’intersection entre certains types de connaissance, et certains aspects temporels de la production orale. Dans la littérature scientifique consacrée à l’aisance orale, les hésitations sont considérées comme des indices précieux d’un manque d’automatisme dans le traitement verbal :

Hesitations are especially useful in showing us where it is easy to move on [in speech production] and where it is difficult. (*Les hésitations sont particulièrement utiles pour révéler où on peut facilement avancer en production, et où cela devient difficile* ; Chafe 1980, p. 171.)

Such characteristics of L2 fluency as speech rate and the length of fluent runs between pauses may reflect automaticity fairly directly. (*Il se pourrait que les caractéristiques d’aisance en L2, tels le taux de parole ou la longueur des énoncés entre deux pauses, reflètent l’automatisme [du traitement langagier] assez directement* ; Schmidt 1992: 362.)

Une partie importante de la disflueance temporelle dans PAROLE est due à la non-automatisme de l’accès lexical (Meara 1993, p. 293 ; Hilton 2002 et 2003) : mots connus mais difficiles à trouver rapidement à l’oral, mots ou propriétés collocationnelles des mots inconnus. Les auteurs du *CECRL* consacrent une page et demie seulement à ce qu’ils appellent la « compétence lexicale » (87-89) – où il est question de connaissances collocationnelles et d’étendu du lexique en L2, mais jamais de la question du fonctionnement de ce réseau sémantique lors du traitement réceptif ou productif. Nous pouvons constater ici que le lien entre les

indices temporels de l'aisance productive et le réseau lexical en L2 est forte, voire fondamentale.

Les conclusions didactiques d'un tel constat sont importantes : l'étendue, la composition et l'accessibilité du lexique en L2 jouent un rôle important dans la compétence productive ; il est donc important d'intégrer une méthodologie de l'acquisition lexicale au cursus d'apprentissage de la langue. L'acquisition lexicale concerne la mise en mémoire d'une vaste quantité de représentations lexicales (phonologiques, orthographiques, collocationnelles), structurées dans un réseau performant pour les besoins du traitement langagier en temps réel. Elle concerne aussi l'acquisition de compétences permettant de contourner de façon efficace les lacunes lexicales inévitables en situation de production ou de réception en L2. Nous constatons, effectivement, que les sujets disfluents de cette étude semblent manquer de stratégies de compensation lexicale, ayant tendance à fournir le mot manquant en L1, ou à le chercher presque indéfiniment (l'hésitation montant à plus de 21 secondes dans un cas extrême). Les sujets fluents, par contre, ont développé des stratégies de contournement lexical, passent rapidement à la paraphrase simple ou à l'hyperonyme, bloquant rarement sur un lexème activé dans la L1.

Des analyses ultérieures de notre corpus parallèle nous permettront de voir si les mêmes structures d'hésitation existent dans les productions en langues morphologiquement plus complexes que l'anglais (le français et l'italien). D'autres études pourront porter sur l'effet de différents niveaux de compétence lexicale dans la production plus interactive.

#### **Références :**

- Arnaud, P. J. L., Béjoint, H., Thoiron, P., 1985, A quoi sert le programme lexical ? *Les Langues modernes*, 3-4, p. 72-85.
- Beattie, G. W., 1980, Encoding units in spontaneous speech : some implications for the dynamics of conversation, in Dechert & Raupach, p. 131-143.
- Bell, C. D. S., 2003, L2 speech rate in monologic and dialogic activities, *Linguagem & Ensino*, 6(2), p. 55-79.
- Casalis, S., 2003, *Test de répétition de logotomes*. Outil non-publié, utilisé par le Département de psychologie, Université de Savoie.
- Chafe, W. L., 1980, Some reasons for hesitating, in Dechert & Raupach, p. 168-180.
- Chambers, F., 1997, What do we mean by fluency? *System*, 25(4), p. 535-544.
- Dechert, H. W., Raupach, M., éd. 1980, *Temporal variables in speech*, den Haag, Mouton.
- de Robien, G., 2006, Mise en œuvre du socle commun de connaissances et de compétences : l'acquisition du vocabulaire à l'école primaire, *Bulletin officiel*, 12, 23 octobre.
- Conseil de l'Europe, Division des Politiques Linguistiques, 2000, *Un cadre européen commun de référence pour les langues*, Paris, Didier.
- Gardner, R., 2004, *Attitude & motivation test battery*, University of Western Ontario.
- Gathercole, S. G., Baddeley, A., 1996, *The children's test of nonword repetition*, Londres, Psychological Corporation.
- Gathercole, S. E., Thorn, A. S. C., 1998, Phonological short-term memory and foreign language learning, in A. F. Healy & L. E. Bourne, éd. *Foreign language and learning : psycholinguistic experiments on training and retention*, Hillsdale, Lawrence Erlbaum, p. 142-158.
- Good, D. A., Butterworth, B. L., 1980, Hesitancy as a conversational resource : some methodological implications. In Dechert & Raupach, 145-152.
- Granger, S., & Monfort, G., 1994, La description de la compétence lexicale en langue étrangère : perspectives méthodologiques, *AILE*, 3, p. 55-76.
- Griffiths, R., 1991, Pausological research in an L2 context : a rationale, and review of selected studies, *Applied Linguistics*, 12(4), p. 345-362.

- Grossman, F., Paveau, M-A., Petit, G., 2005, *Didactique du lexique : langue, cognition, discours*, Grenoble, ELLUG.
- Hart, B., Risley, T. R., 1995, *Meaningful differences in the everyday experience of young American children*, Baltimore, Brookes Publishing.
- Hever, B., sans date, *English vocabulary test*, Göteborg, ForumEducation.
- Hilton, H. E., 2002, Modèles d'acquisition lexicale en L2 : où en sommes-nous? *ASp*, 35, p. 201-217.
- Hilton, H. E., 2003, L'accès au lexique mental dans une langue étrangère : le cas de francophones apprenant l'anglais, *CORELA*, 2, p. 25-43.
- Hulstijn, J. H. 2007, The shaky ground beneath the CEFR: quantitative and qualitative dimensions of language proficiency, *Modern Language Journal*, 91(4), p. 663-667.
- Kowal, S., O'Connell, D. C., 1980, Pausological research at Saint Louis University. In Dechert & Raupach, p. 61-66.
- Laufer, B., 1997, The lexical plight in second language reading : words you don't know, words you think you know, and words you can't guess, in J. Coady & T. Huckin, éd. *Second Language Vocabulary Acquisition*. Cambridge, Cambridge University Press, p. 20-34.
- Lennon, P., 1990, Investigating fluency in EFL : a quantitative approach, *Language Learning*, 40(3), p. 387-417.
- Lété, B., 2004, MANULEX: Le lexique des manuels scolaires de lecture, implications pour l'estimation du vocabulaire des enfants de 6 à 11 ans, in E. Calaque & J. David, éd. *Didactique du langage: contextes, démarches, supports*, Bruxelles, De Boeck, p. 241-257.
- 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, p. 1-22.
- Levelt, W. J. M., 1999, Producing spoken language : a blueprint of the speaker, in C. M. Brown & P. Hagoort éd. *The Neurocognition of Language*, Oxford, Oxford University Press, p. 83-122.
- MacWhinney, B., Spektor, L. (2003-2007) Child Language Data Exchange System (CHILDES), Carnegie-Mellon University, <http://childes.psy.cmu.edu/>.
- Malvern, D., Richards, B., 1994, A new measure of lexical diversity, in A. Ryan & A. Wray éd. *Evolving models of language*, Clevedon, Multilingual Matters, p. 58-71.
- Meara, P. M., 1993, The bilingual lexicon and the teaching of vocabulary, in R. Schreuder & B. Weltens, éd. *The Bilingual Lexicon*, Amsterdam, John Benjamins, p. 279-297.
- Meara, P. M., Milton, J., Lorenzo-Dus, N., 2001, *Language aptitude tests*, Swansea, Express Publishing.
- Nagy, W. E., Herman, P. A., 1987, Breadth and depth of vocabulary knowledge : implications for acquisition and instruction, in M. G. McKeown & M. E. Curtis, éd. *The Nature of Vocabulary Acquisition*, Hillsdale, Laurence Erlbaum, p. 19-35.
- Nation, I. S. P., 2001, *Learning vocabulary in another language*, Cambridge, Cambridge University Press.
- O'Connell, D. C., Kowal, S., 1980, Prospectus for a science of pausology, in Dechert & Raupach, p. 3-10.
- Pawley, A., Syder, F. H., 1983, Two puzzles for linguistic theory : nativelike selection and nativelike fluency, in J. C. Richards & R. W. Schmidt, éd. *Language and Communication*, Londres, Longman, p. 191-226.
- Papagno, C., Valentine, T., Baddeley, A., 1991, Phonological short-term memory and foreign-language vocabulary learning, *Journal of Memory and Language*, 30, p. 331-47.
- Raupach, M., 1980, Temporal variables in first and second language speech production, in Dechert & Raupach. p. 263-270.
- Rieger, C., 2003, Disfluencies and hesitation strategies in oral L2 tests, in R. Ecklund, éd. *Gothenburg Papers in Theoretical Linguistics 90*, Göteborg University, Sweden. p. 41-44.
- Riggenbach, H., 1991, Towards an understanding of fluency : a microanalysis of nonnative speaker conversations, *Discourse Processes*, 14, p. 423-441.

- Schmidt, R., 1992, Psychological Mechanisms underlying second language fluency, *Studies in Second Language Acquisition*, 14, p. 357-85.
- Semel, E., Wiig, E. H., Secord, W. A., 1994, *Clinical Evaluation of Language Fundamentals*, Londres : The Psychological Corporation.
- Smith, M. K., 1941, Measurement of the size of general English vocabulary through the elementary grades and high school, *Genetic Psychological Monographs*, 24, p. 311-345
- Tauroza, S., Allison, D., 1990, Speech rates in British English, *Applied Linguistics*, 11(1), p. 90-105.