

## Développement d'un corpus oral d'apprenants : Apport à la didactique du français langue étrangère dans les Caraïbes anglophones.\*

*Hugues Peters*

The University of New South Wales, Sydney, Australie  
avec la collaboration de Françoise CÉVAËR

### Résumé

Cet article présente quelques réflexions sur la constitution, l'élaboration et l'analyse d'un corpus de productions orales en français faites par des apprenants jamaïcains du français langue étrangère en milieu universitaire. Je m'intéresse plus précisément à la pertinence pédagogique d'un tel corpus dans le contexte de la Caraïbe anglophone. J'illustre mon propos par l'exemple des structures de détachement à gauche avec reprise pronominale et de négation partielle et par l'analyse de la diversité lexicale chez les apprenants.

### Abstract

*This article reflects on the constitution, the development and the analysis of a French learner corpus of oral productions of Jamaican learners of French in a university setting. I am particularly interested in the pedagogical relevance of such a corpus in the context of the English-speaking Caribbean. I illustrate my point with an analysis of the structures of left detachment and partial negation, as well as by an analysis of lexical diversity among learners.*

### Introduction

Cet article propose une réflexion sur la constitution, l'élaboration et l'analyse d'un corpus de productions orales en français faites par des apprenants jamaïcains du français langue étrangère en milieu universitaire. Nous nous intéressons plus précisément à la pertinence pédagogique d'un tel corpus dans le contexte de la Caraïbe anglophone quant au développement des compétences de production orale. Nous illustrons notre propos par des exemples tirés du corpus d'apprenants de l'University of the West Indies (UWI), Mona, Jamaïque, en cours de réalisation par l'auteur.

Dans la première partie, nous décrivons brièvement notre corpus constitué de transcriptions d'entretiens en français d'apprenants de niveaux intermédiaire à avancé de UWI, Mona. Dans la deuxième partie, nous abordons la question de l'apport potentiel des corpus d'apprenants pour la recherche en acquisition et la pédagogie des langues étrangères en illustrant notre propos d'exemples tirés de notre corpus quant aux structures de détachement à gauche avec reprise pronominale et de négations partielles et quant à l'examen de la richesse lexicale des apprenants.

### Brève présentation du corpus d'apprenants de UWI

Rappelons qu'un corpus est simplement une collection de textes, le plus souvent stockés dans des banques de données informatiques. Notre corpus suit un groupe de neuf étudiants de licence inscrits dans le programme de langue française à UWI, sur une période d'environ dix-neuf mois, c'est-à-dire de la fin de la première année (avril 2003) à la fin du premier semestre de la troisième année de leur licence

\* Je remercie Françoise CÉVAËR pour sa collaboration éditoriale à la conception de cet article.

(novembre 2004).<sup>1</sup> Chaque étudiant a été enregistré six ou sept fois au cours de cette période dans des tâches de natures différentes : conversations durant des examens oraux de fin de semestre en contexte formel ou durant des entretiens libres dans une atmosphère détendue et informelle, sur des sujets tels que la vie universitaire, les loisirs, les plans de carrières, des commentaires sur des événements récents de l'actualité ou une discussion sur une personnalité publique, la narration d'une petite histoire présentée sous forme de bande dessinée, la description d'une photographie ou encore des jeux de rôles sur scénarios imposés (Peters, 2005). Ces interviews forment la base des transcriptions et des annotations diverses sur lesquelles des analyses qualitatives et quantitatives peuvent être réalisées.<sup>2</sup>

Les transcriptions respectent les conventions du système CHILDES (McWhinney, 2000a, 2000b). Les transcriptions graphémiques, comme en (1), sont effectuées en alphabet orthographique. Elles tentent de rendre compte le plus exactement possible de ce que les locuteurs disent et incluent donc répétitions (marquées par le code [/]), interjections (« oh ! »), pauses (#), chevauchements de paroles ([>] [<]), divers marqueurs d'hésitation (euh, hm, hum) et indiquent autant que possible des événements extra-linguistiques tels que bruit, toux, rire ([=! laughing]).

(1) Transcription : Locutrice 17 : Année I Semestre 2 : Examen oral final.<sup>3</sup>

\*INV: ton père, il voyage beaucoup?

\*STU: euh pas euh maintenant [=! laughing].<sup>4</sup>

\*INV: pas maintenant?

\*STU: oui.

\*INV: mais il voyage beaucoup?

\*STU: oui.

\*INV: <pour son...> [/] <pour son> [//] pour sa profession +/-?<sup>5</sup>

\*STU: oui.

\*INV: +, ou pour le plaisir?<sup>6</sup>

\*STU: hum # euh quelquefois pour hum sa profession.

\*INV: ++ et quelquefois pour son travail, oui.<sup>7</sup>

\*STU: et ma sœur +/-.

\*EXA: oui.

\*STU: +, elle travaille beaucoup [=! laughing].<sup>8</sup>

\*STU: # hum elle hum quitte la Jamaïque aujourd'hui.

\*INV: oui [>]?

\*EXA: oui [<].

\*STU: et hum elle [/] hum # elle travaille pour l'Espagne.

<sup>1</sup> Les enregistrements et la transcription ont été réalisés par l'auteur.

<sup>2</sup> Le corpus contient 46.941 mots, dont 29.678 mots d'apprenants. La transcription des interactions orales d'un dixième participant est en cours de réalisation.

<sup>3</sup> \*STU est l'étudiante, \*INV le chercheur et \*EXA l'examineur. Dans cet exemple, la situation est celle d'un examen final de première année avec deux examinateurs dont un est aussi le chercheur principal.

<sup>4</sup> La notation [=! texte] indique que l'événement décrit par le texte entre crochets, par exemple "rire", se passe en même temps que le mot précédent: ici la locutrice prononce "maintenant" en riant.

<sup>5</sup> Le symbole & introduit une transcription phonétique sur la ligne principale. [/] indique une répétition et [//] une auto-correction.

<sup>6</sup> Le symbole +, indique que la ligne qui suit forme un seul énoncé avec la ligne précédente terminée par le symbole +/-.

<sup>7</sup> Le symbole ++ indique qu'un locuteur termine l'énoncé d'un autre locuteur.

<sup>8</sup> « Travailler » est probablement utilisé avec le sens de « voyager » à la suite d'une interférence avec le verbe anglais « to travel » (voyager), car le thème de la conversation est celui des voyages.

\*INV: comment?

\*STU: <à Espagne> [//] en Espagne.

Comme on peut le voir en (2), la transcription orthographique peut en outre être enrichie par l'ajout d'une ligne supplémentaire, précédée de l'entête %mor, sur laquelle chaque mot de la ligne principale est décomposé en ses morphèmes constitutifs. Un parseur morphologique du français, intégré à la plateforme CHILDES, permet de générer une telle ligne de manière automatique, même si une intervention humaine est nécessaire pour vérifier cet encodage.

(2) Transcription avec %mor : Locutrice 17 : Année I Semestre 2 : Examen oral.<sup>9</sup>

\*STU: et ma soeur +/.

%mor: conj|et det:poss|ma&fem&sing n|soeur&FEM +/.

\*EXA: oui .

\*STU: +, elle travaille beaucoup [=! laughing] !

%mor: +, pro:subj|elle&FEM&3S v|travailler-PRES&13S adv|beaucoup !

Cette ligne morphologique permet, comme nous le verrons plus loin, d'effectuer des analyses grammaticales plus précises et des recherches plus complexes.

L'étude de l'acquisition du français en contexte guidé est particulièrement adaptée au contexte jamaïcain car la plupart des étudiants n'ont pas la possibilité d'effectuer des séjours de langue ou des voyages touristiques dans des pays ou territoires de langues étrangères, et c'est donc en salle de classe que les langues étrangères sont le plus souvent apprises.<sup>10</sup> En fait, pour ce corpus, n'ont été sélectionnés que des étudiants résidant en Jamaïque depuis la naissance et n'ayant fait aucun séjour dans un pays francophone, ce qui assure une relative homogénéité dans la mesure où nos participants sont tous locuteurs diglossiques d'anglais et de créole jamaïcains. Cela n'empêche pas qu'une hétérogénéité subsiste quant à leur parcours d'études en langues étrangères, notamment dans le fait que, pour certains étudiants, le français est en fait la seconde langue étrangère qu'ils étudient, la première étant l'espagnol.<sup>11</sup>

De plus, les contraintes institutionnelles, qui imposent de placer au même niveau du programme de langue universitaire des étudiants ayant des niveaux de compétences très différents, ont aussi une influence sur la constitution de ce corpus. Nous avons ainsi trois types d'apprenants classés selon leur apprentissage préalable du français :

- ceux ayant le niveau de certification du Caribbean Secondary Certificate (C-SEC) (150-200 heures d'instruction de langue française sur quatre ans) ;
- ceux ayant le niveau de certification du Caribbean Advanced Proficiency Examination (CAPE) ou du A Level de Cambridge (300 heures d'instruction additionnelles au C-SEC, sur deux ans) (Cévaër, 2002; Cévaër & Peters, 2005) ;

<sup>9</sup> Seuls les énoncés des apprenants ont été traités par l'analyseur morphologique car ce sont les seuls qui nous intéressent ici. L'analyseur ignore les répétitions et corrections transcrites sur la ligne principale et ignore les mots commençant par &.

<sup>10</sup> Il est vrai néanmoins que de nos jours l'Internet permet un accès beaucoup plus aisé qu'il y a même cinq ans à du matériel en français. Reste à voir quel usage les étudiants en font.

<sup>11</sup> Grâce au support du gouvernement jamaïcain, l'espagnol est en effet enseigné dans la plupart des lycées du pays, alors que le français n'est le plus souvent enseigné que dans les lycées privés de grandes agglomérations comme Kingston ou Manchester (Cévaër, 2002).

- ceux issus d'un programme de formation intensif en langue française pour débutants administré par UWI (ou d'autres établissements d'enseignement) (jusqu'en 2004, celui de UWI comprenait 260 heures d'instruction sur un an) (Nzengou-Tayo, 2005).

Notre corpus inclut des participants de ces trois catégories majeures dans une proportion équivalente (3 CSEC, 3 débutants, 3 A Level). Cette information sur la formation préalable des étudiants constitue une indication précieuse de leur niveau de compétence, mais n'implique nullement que des étudiants de même formation auront nécessairement une compétence identique.

### Pourquoi les corpus électroniques d'apprenants ?

Les objectifs de la constitution du corpus de UWI sont de :

- rassembler, transcrire et encoder des données de productions orales d'apprenants jamaïcains adultes en contexte universitaire ;
- analyser le développement de traits linguistiques spécifiques et la progression globale de la compétence orale de ces apprenants ;
- apporter une contribution aux débats théoriques actuels sur l'acquisition des langues secondes, spécifiquement en ce qui concerne l'acquisition en contexte diglossique ;
- contribuer à l'amélioration de l'enseignement du français langue étrangère et au développement de matériel pédagogique dans le contexte de la Jamaïque et, plus généralement, des Antilles anglophones.

Bien que le développement de corpus électroniques d'apprenants de langue seconde ou étrangère soit un phénomène relativement récent (Granger, 1998 ; Granger *et al.*, 2002 ; McCarthy, 2004 ; Myles, 2005, 2007), la linguistique de corpus joue un rôle de plus en plus important non seulement dans la recherche en acquisition, mais aussi dans la méthodologie de l'enseignement. Les corpus sont en effet utiles aussi bien pour les chercheurs qui peuvent avec un simple clic de souris tester leurs hypothèses sur des banques de données de taille variable (corpus de quelques dizaines de milliers à plusieurs millions de mots) ou de plus en plus spécialisées (le langage des affaires, le langage de la presse, le langage des apprenants, oral ou écrit, etc.) que, comme nous allons le montrer, pour les didacticiens, qui peuvent raffiner leurs méthodes d'instruction ou développer du matériel pédagogique plus efficaces et mieux adaptés aux caractéristiques de leurs élèves.

### Intérêt pour la recherche en acquisition.

Myles (2005) passe en revue les recherches effectuées dans le domaine de l'acquisition des langues secondes dans le cadre de la linguistique de corpus. Le but de la recherche en acquisition, nous rappelle-t-elle, est de construire des « modèles des représentations mentales et des processus de développement sous-jacents » (2005, p.374). Or, la production orale des apprenants constitue une source sinon essentielle du moins centrale d'accès à ces représentations et processus cognitifs, dans la mesure où elle permet de tester des hypothèses d'acquisition ou de mettre à jour des régularités statistiques dans le langage produit.

Les corpus de productions orales d'apprenants sont malheureusement relativement plus rares que ceux de productions écrites. Cela est dû en grande partie à l'investissement beaucoup plus lourd en

temps que la transcription et l'annotation de ces données orales exigent.<sup>12</sup> Cette relative pénurie de corpus oraux est regrettable car l'oral spontané, produit en temps réel, paraît offrir à la recherche un accès plus direct que l'écrit aux représentations, aux processus cognitifs et aux compétences implicites sous-jacentes. La production écrite implique en effet une planification du dire plus réfléchi et plus élaborée, met en jeu un contrôle métalinguistique plus poussé, et implique donc des connaissances linguistiques de nature plus explicite et moins intuitive, auxquelles s'ajoutent, en outre, des considérations orthographiques et stylistiques. Les conversations impromptues ne sont bien sûr pas dénuées d'une forme de contrôle grammatical, comme en témoignent les nombreux exemples d'auto-correction (notés par [/]) que l'on rencontre chez les apprenants aussi bien que chez les locuteurs natifs. Cette faculté d'auto-correction révèle le système grammatical sous-jacent en cours d'élaboration. Pour s'auto-corriger, il faut en effet avoir perçu l'erreur, ce qui présuppose chez l'apprenant une prise de conscience des régularités de son propre système grammatical et de la distance entre l'expression linguistique produite (« ce qu'il a dit ») et l'intention communicative (« ce qu'il voulait dire »). L'oral permet donc d'avoir un accès en temps réel à ce processus mental car on ne peut pas « gommer » les formulations successives.<sup>13</sup>

Une dernière caractéristique de notre corpus est qu'il s'agit d'un corpus longitudinal, c'est-à-dire d'un corpus qui suit des apprenants spécifiques dans la durée. Cette approche permet d'observer les différentes étapes de l'évolution de la compétence linguistique d'apprenants particuliers. Pour toutes ces raisons, notre corpus constitue un apport original à la recherche en acquisition.

### Intérêt pour la pédagogie des langues étrangères.

Quoique le lien entre recherche en acquisition et didactique des langues soit objet de controverses, essentiellement pour des raisons de positionnement dans le champ universitaire et de traditions intellectuelles différentes ayant des finalités théoriques (perçues comme) incompatibles, - la recherche en acquisition s'intéressant au développement de la structure linguistique en elle-même, dans un cadre théorique à dominante linguistique alors que la didactique aurait surtout des préoccupations pratiques de développement de compétences communicatives et interculturelles dans un cadre conceptuel plus ouvert à l'éclectisme (voir notamment Ellis, 1997 ; Véronique 2005) - un domaine où la recherche de corpus pourrait apporter un éclairage pertinent à la pratique des enseignants semble être celui de l'évaluation des difficultés et de la correction / remédiation des erreurs de production orale. Les études de corpus d'apprenants pourraient être d'autant plus pertinentes pour informer la pratique didactique qu'il s'agit d'apprenants en contexte guidé.

Selon notre expérience, les enseignants sont en effet souvent démunis lorsqu'il s'agit d'évaluer la production orale de leurs élèves, de diagnostiquer les zones de difficultés ou de mettre en place des stratégies pédagogiques de remédiation personnalisées. Étant donné que la parole ne laisse pas de traces, les enseignants, que les conditions de travail ou les contraintes de temps empêchent de faire un relevé exhaustif de données, n'ont le plus souvent à leur disposition que leur propre jugement intuitif. Celui-ci repose tantôt sur leur expérience professionnelle et l'utilisation de grilles de correction standardisées qui ne prennent pas en compte les caractéristiques spécifiques des apprenants, notamment en ce qui

<sup>12</sup> Il faut en effet environ 20 minutes de travail de transcription et d'encodage pour chaque minute d'enregistrement. Un exemple de tel corpus est notamment le « French Learner Oral Corpora » (FLLOC) (Rule, 2004 ; Myles, 2005, 2007) accessible en ligne: <http://www.flloc.soton.ac.uk/>

<sup>13</sup> Il est vrai, cependant, que cette distinction « spontané » / « élaboré » pour distinguer l'oral et l'écrit a tendance à s'estomper suite au développement de formes d'écrits immédiats comme le « chat » sur ordinateur et que, d'autre part, les conversations impromptues ne sont bien sûr pas dénuées d'une forme de contrôle grammatical, ni d'un emploi de formules figées ou stéréotypées.

concerne leur langue maternelle, tantôt sur des débuts d'analyse *ad hoc* d'enregistrements de tests ou d'activités de classe.

Cette approche parcellaire oblige l'enseignant à se concentrer sur une sélection des erreurs qu'il juge les plus marquantes, identifiées à un moment donné de l'apprentissage. Or, idéalement, il faudrait disposer d'un panorama complet du processus d'acquisition qui prendrait en compte aussi bien les structures correctes que les structures erronées (selon le standard de la langue cible) à différents moments de l'acquisition, et qui permettrait en outre d'identifier les structures 'absentes' par suite de stratégies d'évitement employées par les étudiants face à des constructions que, à tort ou à raison, ils ont l'impression de ne pas maîtriser.

Précisons d'emblée, en ce qui concerne le concept « d'erreur, » que la didactique des langues a adopté l'idée issue de la recherche en acquisition, que « l'erreur » dans les productions d'apprenants n'est pas « faute » fortuite ou fruit du hasard, mais constitue une étape nécessaire dans le processus d'acquisition ainsi qu'un révélateur du système linguistique transitoire en construction. Il est néanmoins vrai que dans l'urgence de la salle de classe et sous l'influence de pressions institutionnelles qui obligent à classer les étudiants de manière hiérarchique par l'attribution d'une évaluation sommative sanctionnée par une note, aussi bien les enseignants que les apprenants se focalisent le plus souvent sur l'erreur de manière pénalisante plutôt que constructive.

De plus, l'importance excessive accordée à l'écrit dans le contexte éducatif, plus encore au niveau universitaire, intimement liée à l'idéologie de la norme standard contribue à la négation de la variabilité inter-individuelle (régionale, sociale, etc.) et intra-individuelle (registres formels et informels, styles écrits ou oraux), qui existe pourtant chez les locuteurs natifs, et à l'adoption en classe de langue d'un hyper standard artificiel et inatteignable (Train, 2007). Une approche systématique de données orales d'apprenants devrait donc permettre à l'enseignant de mettre en place une pédagogie de l'oral moins influencée par la norme de l'écrit, et mieux informée des particularités de la communication orale en général et des spécificités linguistiques de ses apprenants en particulier.<sup>14</sup>

Il est ainsi bien connu (voir notamment Blanche-Benveniste, 1997) que le français parlé a recours à des constructions syntaxiques que l'on retrouve plus rarement ou pas du tout à l'écrit, comme par exemple les « constructions détachées à gauche » qui introduisent le « thème » ou le « topique » de l'énoncé, c'est-à-dire « ce dont on va parler. » Lorsque ce détachement concerne un argument du verbe (sujet ou complément), le groupe nominal détaché est repris par le pronom personnel correspondant dans le noyau verbal. Dans notre corpus, nous en trouvons de nombreux exemples dans les interventions du chercheur ou des examinateurs natifs:

- (3) Exemples de constructions détachées à gauche avec reprise pronominale chez un natif en conversation avec le locuteur 17.
- a. **ton père, il** voyage beaucoup? (A1 s2)<sup>15</sup>
  - b. et euh **cette maîtrise ou ce doctorat**, tu veux l'obtenir euh ici en Jamaïque? (A2 s1)

Or, bien que cette structure de détachement d'un groupe nominal avec reprise pronominale ne fasse pas l'objet d'un apprentissage particulier en cours de français, sauf en ce qui concerne l'enseignement des pronoms toniques en utilisation de topiques contrastifs pour mettre l'accent sur un

<sup>14</sup> Parmi les spécificités que, faute de place, nous n'aborderons pas dans cet article, il faut évidemment placer le contexte linguistique diglossique typique des Caraïbes (Christie, 2003; Kouwenberg, 2007; Mather, 2007). Voir Peters (2009) pour un examen de cette question du transfert de traits du créole.

<sup>15</sup> Dans les exemples, « A » est utilisé pour l'année de licence et « s » pour le semestre. Pour faciliter la lecture des exemples, nous supprimons les indications de pauses, d'événements extralinguistiques et les encodages superflus.

pronom personnel sujet (« moi, je... toi, tu... »), nos apprenants semblent bien l'avoir acquise au contact des apprenants natifs, mais seulement en ce qui concerne le détachement du sujet, dont on voit quelques exemples en (4 a-c) chez la locutrice 17 :

- (4) Constructions détachées avec reprise pronominale sujet.
- a. \*STU: et **ma sœur** +/.
  - \*STU: +, **elle** travaille beaucoup. (AI s2)
  - b. hum mais **les professeurs de sciences, ils** n'aiment pas euh l'idée de hum faire des études dans les deux facultés. (AII s1)
  - c. par exemple, **ma sœur, elle** habite à Miami. (AII s1)
  - d. parce que hum, par exemple, **moi, je** n'ai pas étudié hum le français pendant deux années. (AII s2)
  - e. quelquefois **les politiciens, ils** ont &pl beaucoup de pouvoir. (AIII s1)

Cette structure de dislocation est en revanche évitée aux autres fonctions que celle de sujet, ou bien donne lieu à une utilisation erronée :

- (5) Constructions détachées avec reprise pronominale objet chez la locutrice 17.  
hum **les examens**, nous doit fait les [/] les [/] &z **les**. (AIII s1)

Dans l'exemple (5.a), l'apprenant détache correctement à gauche le groupe nominal complément direct du verbe, mais ne réalise que partiellement la structure de redoublement pronominal, car le pronom personnel objet « les » est placé dans une position post-verbale incorrecte (à moins qu'il ne s'agisse de l'article « les » dont le nom serait laissé en suspens).

Selon Bartning et Schlyter (2004), la position préverbale des pronoms compléments semble pourtant devoir être acquise ou en cours d'acquisition aux niveaux intermédiaires et avancés. D'ailleurs, nous trouvons des exemples de leur emploi en position correcte chez ce même apprenant, en (6.a-b), parfois avec beaucoup d'hésitations comme en (6.c) ou encore avec des difficultés diverses en rapport avec le choix du temps et du mode du verbe, comme en (6.d-f) ou en rapport avec le choix du pronom, comme en (6.g) où « le » est utilisé à la place de « y ».

- (6) Les pronoms personnels objets chez la locutrice 17 :
- a. peut-être nous avons besoin de beaucoup de temps pour **le** faire. (AII s2)
  - b. hum je ne sais pas si je peux **le** gagner parce+que hum ce n'est pas mon spécialité, le français. (AII s2)
  - c. mais <elle veulent> [//] elle <va euh me visiter> [//], euh **me** visitera. (AII s1)
  - d. hum je pense que je **me** sens que hum je **me** cultiver mais +//. (AIII s1)
  - e. et je gagne de l'argent pour **me** soutenir. (AIII s1)
  - f. qui ont hum visité la plage pour recevoir une vue finale de hum la plage avant que le cyclone **le** détruire. (AIII s1)
  - g. je vais **le** penser. (AIII s1)

L'absence de structure de dislocation avec des pronoms objets semble donc liée à des difficultés plus générales que la simple connaissance grammaticale des pronoms compléments chez ce locuteur. Si l'on analyse les productions d'étudiants plus avancés, ayant commencé leurs études universitaires de français avec un diplôme de niveau A Level / CAPE, tel que les locutrices 14 et 20, le même phénomène

est observé : des exemples généralement corrects d'emplois des pronoms compléments peuvent être identifiés, mais à nouveau on ne trouve qu'un seul exemple de dislocation redoublée par un pronom objet. Ce qui semble donc confirmer une absence de sensibilité à cette structure typique de l'oral et plus généralement, une difficulté qui se situe au niveau de l'articulation entre la grammaire et l'emploi des structures grammaticales en discours, difficulté qui s'explique en partie par l'absence de cette structure en Anglais.

- (7) mais je suis chrétienne et je crois que **n'importe qui le dieu veut pour moi**, il le permet@s:v [= permet] à passer pour moi. (Loc.20, A1 s2)

Une séquence didactique de révision des pronoms personnels pourrait se faire avantageusement dans un cadre communicatif oral qui fournirait le contexte discursif adéquat à leur utilisation et inciterait par la même les apprenants à utiliser cette structure de dislocation à gauche d'une manière qui viendrait en complément d'exercices structurels plus classiques axés sur l'écrit. Le bénéfice serait double puisqu'on renforcerait la capacité à employer les pronoms de manière grammaticalement correcte et qu'on développerait une façon de parler plus proche de celle des locuteurs natifs par la prise de conscience du rôle de la topicalisation en discours.

Véronique (2005) suggère que les résultats de recherches en acquisition pourraient être avantageusement utilisés pour aider à la réalisation de grammaires pédagogiques et de curriculum d'apprentissage. Les corpus d'apprenants fournissent une aide essentielle dans ce processus car ils permettent l'identification de problèmes et d'erreurs typiques (Granger, 1999 ; Véronique, 2005). Selon McCarthy (2004), la linguistique de corpus peut aussi être utilisée par les apprenants eux-mêmes pour observer et évaluer à tête reposée leurs propres erreurs et les comparer de manière anonyme et émotionnellement détachée, à celles d'autres apprenants. Cette comparaison aussi bien pour le professeur que pour l'apprenant dédramatise l'erreur par la prise de conscience que faire des erreurs est un phénomène largement partagé et normal pour tous les apprenants.

En outre, McCarthy (2004) suggère que la comparaison des corpus d'apprenants avec des corpus de locuteurs natifs permettra de mettre en évidence les stratégies que les apprenants en contexte guidé n'utilisent pas à l'oral. Ainsi, dans Peters (2005), il est montré que les adverbes de négation partielle (tels que « jamais », « plus » ou « pas encore ») sont souvent évités dans les productions orales d'apprenants, même aux niveaux les plus avancés, et remplacés par des combinaisons telles que « pas maintenant », « pas déjà », « pas avant » d'adverbes positifs avec la négation totale alors même que ces formes à adverbes de négation partielle sont utilisées dans les productions écrites de ces mêmes apprenants. Cette observation suggère, qu'ayant acquis une connaissance explicite de ces mots grammaticaux à l'écrit, les apprenants sont néanmoins incapables de mobiliser cette connaissance dans l'urgence de productions orales spontanées.

- (8) Une stratégie d'évitement : les négations partielles (voir Peters, 2005)
- je **ne** trouve **pas** un professeur spécifique **maintenant** (dans le sens 'je n'en ai pas encore trouvé') (Loc. 18, AII s1)
  - je **ne** sais **pas maintenant** (dans le sens 'je ne sais plus') (Loc. 38, AI s2)
  - et bien je **n'** avais voyagé à la France **avant** (dans le sens 'je n'ai jamais voyagé en France) (Loc. 20, AIII s1)
  - je **n'** ai **pas** trouvé **déjà** (dans le sens 'je n'ai pas encore trouvé') (Loc. 17, AII s1)

Enfin, dans un contexte institutionnel qui place dans le même programme de langue des étudiants ayant reçu 150-200 heures d'instruction et d'autres en ayant reçu le double, disposer de données chiffrées précises sur les différences de compétence linguistique sera particulièrement utile pour identifier les zones de difficultés et établir des programmes plus ciblés de remédiations pédagogiques différenciées. Grâce au programme VocD, la plateforme électronique CHILDES offre notamment la possibilité d'obtenir une mesure de diversité lexicale, c'est-à-dire une mesure du nombre de types de mots par rapport au nombre total de mots produits par un locuteur.<sup>16</sup>

Ainsi, une analyse de mesure de diversité lexicale chez les apprenants montre d'une manière relativement claire qu'une différence entre les étudiants les moins avancés (ceux de niveau C-SEC et ceux issus du programme intensif pour débutants) et les étudiants les plus avancés (de niveau A-Level / CAPE) se situe précisément au niveau de la richesse lexicale en production orale.<sup>17</sup>

Tableau 1 : Mesure de la diversité lexicale (VocD) des apprenants.

Formation	Locuteurs	Année et semestre de l'entretien			
		AI s2	AII s1	AII s2	AIII s1
A-Level ou CAPE	Loc.14	60.53	<b>83.09</b>	47.64	<b>81.40</b>
	Loc.20	56.26	59.79	55.75	<b>85.97</b>
	Loc.31	65.59	54.00	63.07	54.44
C-SEC	Loc.17	76.49	69.46	57.66	74.59
	Loc.18	35.63	43.70	44.18	50.59
	Loc.38	53.56	51.88	54.34	51.17
Programme pour débutants	Loc.08	X <sup>18</sup>	62.79	65.52	68.61
	Loc.12	X	56.04	68.78	60.75
	Loc.16	61.60	51.04	36.76	54.12

Nous observons en effet que les trois scores de diversité lexicale les plus élevés du tableau (indiqués en caractères gras) appartiennent tous à des étudiants de formation A Level : locutrice 20 : **85.97** (Année III sem.1) et locutrice 14 : **83.09** (Année II sem.1) et **81.40** (Année III sem.1), deux de ces scores ayant été obtenus lors d'entretiens réalisés au cours de la troisième année de licence et un au

<sup>16</sup> Le programme VocD est le programme le plus élaboré permettant d'obtenir une mesure de la richesse lexicale (MacWhinney, 2000a : p.118). Il s'agit d'un programme qui calcule statistiquement la diversité de types de mots utilisés par rapport au nombre total d'occurrences de mots (token) produits sur des échantillons pris au hasard au cours d'une interaction. Ce programme a l'avantage sur les autres méthodes d'éliminer l'effet de longueur de texte, - la mesure de variabilité calculée par simple proportion types / occurrences ayant en effet tendance à décroître à mesure que la longueur du texte augmente -, effet qui rend les autres méthodes peu fiables. La version la plus récente de VocD sur CHILDES permet en outre d'effectuer ce calcul de diversité sur la ligne morphologique, %mor, plutôt que sur la ligne principale des énoncés (Voir la version de décembre 2009 du manuel électronique des programmes de CLAN, p.116, sur <http://chilides.psy.cmu.edu>). Effectuer ce calcul sur la ligne morphologique permet d'éviter de traiter comme mots différents les formes différentes d'un lexème unique : « (nous) marchons » et « (tu) marches » ne sont que deux formes d'un seul mot-type « marcher » et seront comptés comme tel. Inversement le calcul sur la ligne %mor permet de traiter comme mots-types différents des mots ayant superficiellement la même forme : « (tu) marches » et « (les) marches » seront ainsi comptés comme deux occurrences de mots-types différents, respectivement le verbe « marcher » et le nom « marche ».

<sup>17</sup> Nous ne fournissons les scores de diversité lexicale que pour les entretiens réalisés au cours des examens finals (et pas au cours des entretiens informels). Pour cette opérations, j'ai pris soin d'exclure du décompte les répétitions, les mots anglais, les interjections et onomatopées diverses.

<sup>18</sup> Les scores des Locutrices 08 et 12 pour l'Année I sem.2 ne sont pas fournis parce que nous ne disposons pour ces dernières que de données relevés lors d'une activité différente des autres locuteurs : un exposé formel au lieu d'un examen sous forme d'entretien.

cours de la deuxième année. Inversement, les trois scores les plus bas du tableau (indiqués en italiques) appartiennent tous à des étudiants de formation C-SEC ou sortis du programme pour débutants : locutrice 16 : 36.76 (Année II sem.2), locutrice 18 : 35.63 (Année I sem.2) et 43.70 (Année II sem.2), lors d'entretiens réalisés au cours de la première ou de la deuxième année de licence.

De plus, lorsque l'on observe les résultats à chaque semestre du programme colonne par colonne, nous remarquons que les deux scores les plus élevés de chaque colonne appartiennent toujours à au moins un apprenant de formation A Level, à l'exception de l'entretien réalisé lors de l'examen oral de niveau Année II sem.2, et qu'en revanche les deux scores les plus bas de chaque colonne n'appartiennent jamais à ces étudiants.<sup>19</sup>

Sur la base de ces résultats, grâce notamment à des listes de fréquence de mots typiques utilisés par les étudiants, il est possible de mettre en place des stratégies de remédiations pédagogiques axées sur le développement lexical et dirigées spécifiquement vers les étudiants les moins avancés, notamment par le choix de thèmes de conversation spécifiques. À ce propos, notons le cas de la locutrice 17 qui, bien que de formation CSEC, réalise de manière consistante des scores relativement élevés, équivalents à ceux d'étudiants de A level. Cela pourrait s'expliquer par le fait que cette locutrice prenne la peine, comme elle nous l'avait déclaré au cours d'un entretien, d'écouter des dvd de films français en version originale et qu'en outre elle suive des cours de littérature française l'obligeant à lire romans et pièces de théâtre en français, et qu'elle se soit donc, si l'on peut dire, préparé un programme pédagogique personnalisé qui porterait ses fruits au niveau du vocabulaire.

## Conclusion

L'analyse de corpus permet une approche systématique de données de production orale en français d'apprenants du français langue étrangère. L'analyse d'interactions orales devrait permettre la prise de conscience chez l'enseignant de la spécificité de l'oral par rapport à l'écrit. En outre, cette approche devrait permettre aux enseignants de distinguer les constructions typiques à un niveau de compétence donné pour tous les apprenants de celles qui sont propres à tel ou tel apprenant, sans se focaliser uniquement sur les erreurs, et donc d'aider les enseignants à choisir la meilleure stratégie d'intervention pédagogique aux niveaux grammatical, discursif ou lexical.

Dans le cadre de l'intervention pédagogique à l'oral, l'enseignant aura le choix entre plusieurs options: (i) soit l'éradication d'une construction erronée par une focalisation sur la forme, des exercices structuraux adéquats et des activités communicatives pertinentes; (ii) soit l'élargissement des compétences de production en mettant l'accent sur des constructions que les apprenants auraient tendance à éviter, par un renforcement de leur maîtrise des dites structures grâce à des séquences didactiques incluant aussi bien des activités structurales systématiques que l'utilisation de situations de communications facilitant l'emploi de ces structures; (iii) soit la relativisation d'une construction à un registre de langue, par exemple l'omission du « ne » utilisée le plus souvent par les étudiants dans des formules stéréotypées, telles que « c'est pas... » ou « je sais pas » et qui bien que manifestant une sensibilité à une caractéristique réelle des productions orales de locuteurs natifs, doit être évitée en contextes formels oraux ou écrits.

<sup>19</sup> Un élément d'explication pourrait être que l'entretien de fin du deuxième semestre de la deuxième année était plus court que celui des années précédentes, et donc en général, la quantité de sujets abordés était moindre, ce qui ne permettait pas aux apprenants plus avancés d'exprimer la gamme de leur vocabulaire.

## Références

- Bartning, I., & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French Language Studies*, 14, 281-299.
- Blanche-Benveniste, Cl. (1997). *Approches de la langue parlée en français*. Paris: Ophrys.
- Cévaër, F. (2002). Educational policies in the English-speaking countries of the Caribbean: The case of CAPE French in Jamaica. *Caribbean Journal of Education*, 24(2), 151-185.
- Cévaër, F., & Peters, H. (2005). Stratégie de politique éducative pour une meilleure intégration régionale: évaluation du nouvel examen de français des pays anglophones de la Caraïbe, le CAPE. In C. Lerat (Éd.), *Le monde Caraïbe: Défis et dynamiques* (pp.227-240). Presses de la Maison des Sciences de l'Homme d'Aquitaine.
- Christie, P. (2003). *Language in Jamaica*. Kingston: Arawac Publications.
- Ellis, R. (1997). SLA and language pedagogy: An educational perspective. *Studies in Second Language Acquisition*, 20, 69-92.
- Granger, S. (1998). *Learner English on computer*. London/New York: Addison Wesley Longman.
- \_\_\_\_\_. (1999). Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. In H. Hasselgård, et al. (Éds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp.191-202). Amsterdam/Atlanta: Rodopi.
- Granger, S., Hung, J., & Petch-Tyson, S. (Éds). (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: Benjamins.
- Kouwenberg, S. (2007). Bringing language awareness into the high school curriculum: The opportunities offered by CAPE communication studies. *Caribbean Journal of Education*, 29(2), 206-221.
- MacWhinney, B. (2000a). *The CHILDES project: Tools for analyzing talk. Volume 1: Transcription format and programs*. Mahwah, NJ: Lawrence Erlbaum.
- \_\_\_\_\_. (2000b). *The CHILDES project: Tools for analyzing talk. Volume 2: The database*. Mahwah, NJ: Lawrence Erlbaum.
- Mather, P.A. (2007). Creole studies. In D. Ayoun (Éd.), *French Applied Linguistics* (pp.401-424). Amsterdam: John Benjamins.
- McCarthy, M. (2004). Using a corpus in language teaching. Calper Professional Development Document CDDP0410. [En ligne: <http://calper.la.psu.edu>]
- Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373-391.
- \_\_\_\_\_. (2007). Using electronic corpora in SLA research. In D. Ayoun (Éd.). *French Applied Linguistics* (pp.377-400). Amsterdam: John Benjamins.
- Nzengou-Tayo, M.J. (2005). Computer-assisted language learning at UWI, Mona: using technology to face strategic challenges. *Caribbean Journal of Education*, 27(2), 142-162.
- Peters, H. (2005). Development of a longitudinal oral interlanguage corpus of Jamaican learners of French. *Caribbean Journal of Education*, 27(2), 75-96.
- \_\_\_\_\_. (2009). Lexical features in oral interlanguage: The acquisition of multifunctional words by adult Jamaican learners of French. *The International Journal of the Humanities*, 7(6), 97-108.
- Rule, S. (2004). French interlanguage oral corpora: Recent developments. *Journal of French Language Studies*, 14(3), 343-356.
- Train, R.W. (2007). Language ideology and foreign language pedagogy. In D. Ayoun (Éd.). *French Applied Linguistics* (pp.238-69). Amsterdam: John Benjamins.

Véronique, G.D. (2005). Les interrelations entre la recherche sur l'acquisition du français langue étrangère et la didactique du français langue étrangère : Quelques pistes de travail. *Acquisition et Interaction en Langue Etrangère*, 23 [En ligne: <http://aile.revues.org>].