



HAL
open science

Corpus d'apprenants et acquisition des langues

Sylvie de Cock, Henry Tyne

► **To cite this version:**

Sylvie de Cock, Henry Tyne. Corpus d'apprenants et acquisition des langues. Recherches en Didactique des Langues et Cultures - Les Cahiers de l'Acedle, 2014, Recherches en Didactique des Langues et des Cultures, 11 (1), pp.137-168. hal-01211359

HAL Id: hal-01211359

<https://hal.science/hal-01211359>

Submitted on 5 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Corpus d'apprenants et acquisition des langues

Sylvie De Cock¹ ; Henry Tyne²

¹ Université Catholique de Louvain et Université Saint-Louis, Belgique ; ² Université de Perpignan Via Domitia et VECT, France

Résumé

Ce que l'on appelle les "corpus d'apprenants" sont des recueils électroniques de productions orales ou écrites d'apprenants de langue seconde ou étrangère (L2) (Granger, 2004). Ces ensembles de données informatisées sont notamment utilisés dans le cadre de recherches qui explorent certains aspects de l'acquisition d'une L2. La première partie de l'article illustre la contribution d'études basées sur des corpus d'apprenants dans le domaine de l'acquisition des L2. La deuxième traite d'applications pédagogiques concrètes de ce genre d'étude. La troisième porte un regard plus critique sur certains aspects et soulève des points de discussion ou de débat qui ouvrent souvent sur le domaine plus général de l'acquisition des L2 ou sur l'utilisation de corpus pour l'étude de la langue.

Mots-Clés

corpus d'apprenants, acquisition des L2, applications pédagogiques

Abstract

Learner corpora are systematic "electronic collections of spoken or written texts produced by foreign or second language learners" (Granger, 2004: 124). These corpora have been used to explore specific aspects of second language acquisition. The first section of this paper illustrates the insights into L2 development that can be gained from learner corpus research. The second section presents and discusses possible pedagogical applications of this type of research. The article then offers a more critical perspective on learner corpus research and more generally on studies on second language acquisition and the use of corpora in language research.

Keywords

learner corpora, second language acquisition, pedagogical applications

1. Introduction

Ce que l'on appelle les "corpus d'apprenants" (parfois appelés "corpus d'interlangue"), sont des recueils électroniques de productions orales ou écrites d'apprenants de langue seconde ou étrangère (L2) (Granger, 2004). Ces ensembles de données informatisées, qui ont vu le jour vers la fin des années 1980, sont notamment utilisés dans le cadre de recherches qui explorent certains aspects de l'acquisition d'une L2. Le domaine des études sur l'acquisition des L2 travaille à partir de données attestées depuis longtemps ; c'est un champ qui s'appuie historiquement sur une démarche empirique et les premiers "corpus" utilisés par les chercheurs travaillant sur l'acquisition étaient des collections d'observations, des relevés d'erreurs, des productions sollicitées, etc. D'ailleurs, on peut constater une évolution au niveau des données utilisées pour l'étude de l'acquisition de L2 : les premières études se basent sur des inventaires ou listes ; on travaille ensuite sur des données sollicitées, et plus récemment sur des recueils de productions (orales comme écrites) en L2. Cette évolution s'accompagne d'autres changements, notamment au niveau de la manière dont on traite les données (développement de l'outillage informatique) et on peut citer l'apport récent de la linguistique de corpus en ce qui concerne les techniques d'observation et de description des données.

Il existe une grande variété de corpus (Granger, 2012). Même si la majorité des corpus existant dans le monde aujourd'hui contiennent des productions d'apprenants d'anglais, de plus en plus de corpus d'autres L2 sont rassemblés de nos jours (<http://www.uclouvain.be/en-ccel-lcworld.html>). Les études mentionnées dans cet article portent essentiellement sur des corpus d'apprenants d'anglais et de français L2 étant donné les domaines de recherche des deux auteurs.

Cet article est divisé en trois grandes parties. La première illustre la contribution d'études basées sur des corpus d'apprenants dans le domaine de l'acquisition des L2. La deuxième traite d'applications pédagogiques concrètes de ce genre d'étude. La troisième porte un regard plus critique sur certains aspects et soulève des points de discussion ou de débat qui ouvrent souvent sur le domaine plus général de l'acquisition des L2 ou sur l'utilisation de corpus pour l'étude de la langue.

2. Corpus d'apprenants et développement langagier

L'un des avantages de l'utilisation de corpus d'apprenants dans des études sur l'acquisition des langues, et ce comparé à des données empiriques plus traditionnellement utilisées, réside dans

le fait que, grâce à leur taille et au fait que ces corpus rassemblent des textes (oraux comme écrits) produits par un certain nombre d'apprenants, ils permettent d'observer des régularités d'utilisation et de quantifier certains aspects du développement langagier et de l'interlangue en général (Myles, 2005 ; Granger, 2012). Comme l'a fait remarquer Granger (2012 : 9), l'un des atouts majeurs des corpus d'apprenants, par rapport à des observations moins systématiques, réside dans leur taille et leur représentativité (nous reviendrons sur la question de l'échantillonnage plus loin). En outre, comme tout corpus informatisé, les corpus d'apprenants peuvent être étiquetés et traités par des logiciels d'analyse linguistique.

L'une des critiques souvent formulées à l'égard des études sur corpus d'apprenants (et sur corpus en général) est que ces dernières sont le plus souvent essentiellement descriptives. En effet, très peu parmi elles se focalisent plus particulièrement sur le développement langagier ou s'attachent à tenter d'expliquer les phénomènes observés. Cette section a pour but de discuter des études sur corpus d'apprenants qui se concentrent notamment sur le développement langagier ou sur l'explication de certains phénomènes observés dans le domaine de l'acquisition des L2. Etant donné qu'il est impossible de prétendre à une exhaustivité dans cette présentation, nous nous contenterons de citer quelques études à titre d'exemple.

Un corpus d'apprenants longitudinal semble être le plus indiqué pour explorer le développement langagier des apprenants car il contient les productions d'une cohorte d'apprenants pendant une période de plusieurs mois ou de plusieurs années, ce qui permet de suivre en quelque sorte le processus d'acquisition en cours. Mais rassembler un tel corpus représente en réalité de nombreux défis (moyens disponibles, taux d'attrition de la cohorte, etc. ; cf. Meunier & Littré, 2012). Ce genre de corpus n'est dès lors pas toujours disponible/accessible et une approche dite "quasi-longitudinale" (en temps apparent, avec des groupes d'apprenants de niveaux différents) est souvent adoptée. L'analyse d'erreurs assistée par ordinateur de Thewissen (2012a) illustre ce type d'approche. Cette étude vise à explorer le développement de la correction formelle d'apprenants d'anglais à l'écrit. Les données utilisées proviennent des sous-corpus d'apprenants francophones, germanophones et hispanophones d'ICLE (*International Corpus of Learner English*), un corpus qui rassemble des écrits argumentatifs d'apprenants d'anglais de niveau avancé. En tout, 223 textes ont été sélectionnés au hasard dans les trois sous-corpus, ce qui représente 50.000 mots (*tokens*) pour chacun des trois groupes et au total 150.000 mots. Le niveau dit "avancé" des apprenants qui ont contribué à ICLE a été déterminé sur la base d'un critère externe simple : il s'agit d'étudiants

d'anglais de troisième et quatrième année à l'université. Mais, quand on y regarde de plus près, on s'aperçoit que les textes produits par ces étudiants ne peuvent pas tous être considérés comme étant de niveau égal, ce qui permet à Thewissen d'étudier le développement de la correction formelle dans un corpus quasi-longitudinal stratifié selon les niveaux du *Cadre européen commun de référence pour les langues* (CECR ; Thewissen, 2012b : 79). Pour ce faire, les 223 textes argumentatifs sélectionnés pour l'étude ont été analysés de manière indépendante par deux évaluateurs professionnels. Il leur a été demandé d'attribuer à chaque texte un des niveaux du CECR⁴⁶. Parmi eux, 66 textes ont été jugés de niveau B1, 62 de niveau B2, 67 de niveau C1 et 28 de niveau C2.

Les textes ont ensuite été annotés suivant le système d'étiquetage *Louvain Error-Tagging System* (Dagneaux *et al.*, 1998, 2008) qui comporte deux principaux niveaux d'annotation. Le domaine d'erreurs, qui est le niveau le plus général, précise si l'erreur est formelle (ou orthographique) (F), grammaticale (G), lexicale (L), lexico-grammaticale (X), ponctuationnelle (Q), etc. Les domaines d'erreurs peuvent être subdivisés en catégories qui reflètent différents niveaux de granularité (par exemple, l'étiquette GV signale des erreurs sur des verbes alors que GVT signale plus précisément des erreurs de temps). Une fois le corpus annoté, les erreurs de chaque type ont été comptabilisées pour chacun des niveaux du CECR afin de pouvoir identifier des points de progrès, de stabilisation ou de régression éventuels.

Il ressort de l'étude que, pour les niveaux intermédiaires et avancés de l'échantillon d'ICLE, le développement de la correction formelle est caractérisé par le progrès et la stabilisation ou par la stabilisation uniquement. Il en ressort également que seuls deux types d'erreurs sont caractérisés par un progrès constant et significatif du niveau B1 au niveau C2 : il s'agit du nombre total d'erreurs et du choix de mot lexical simple. L'étude montre en outre que la plus grande marge de progrès se situe entre le niveau B1 et le niveau B2 et qu'il y a stabilisation du niveau B2 au niveau C2. De manière générale, les résultats semblent indiquer que le développement de la correction formelle est en grande partie non-linéaire. Thewissen souligne également que les erreurs ne doivent pas nécessairement être considérées comme des éléments négatifs dans le développement d'une L2 – certains types d'erreurs sont même

⁴⁶ On reproche parfois au CECR (voir dans Bartning *et al.*, 2010) le manque de précision à certains égards dans la délimitation des différents niveaux. Ce problème, inhérent à toute évaluation critériée, renvoie surtout à l'interprétation des descripteurs et habitudes des évaluateurs. Il est à noter, cependant, que dans l'étude de Thewissen le niveau de fiabilité inter-évaluateurs est élevé. Nous reviendrons sur l'apport des études récentes au service des référentiels en langues.

qualifiés de positifs car elles témoignent d'un progrès plus général de développement ou de mise en place de l'interlangue.

L'étude de Meunier et Littré (2012), s'inscrivant dans un domaine classique dans les recherches sur l'acquisition de L2, à savoir l'utilisation des temps et de l'aspect, examine le développement dans des productions écrites en anglais par des apprenants francophones sur une période de trois ans. Le corpus d'apprenants utilisé dans cette étude est le *Longitudinal Database of Learner English (Longdale)*, un véritable corpus longitudinal pour lequel des profils de niveau de langue sont disponibles pour chacun des apprenants (test de langue standardisé passé par chaque participant) pour chacun des moments de collecte. Dans le cadre de cette étude, les erreurs de temps et d'aspect ont été annotées dans des écrits argumentatifs produits par une cohorte d'étudiants de langue et littérature anglaise (Université Catholique de Louvain). A côté des données de corpus, l'étude repose également sur des données expérimentales recueillies grâce à un test dans le cadre duquel les étudiants devaient juger de l'acceptabilité de phrases⁴⁷ dans un contexte donné. Les étudiants avaient ensuite refait un test semblable trois semaines plus tard et, cette fois, ils devaient également justifier leur choix. De plus en plus d'études sur le développement langagier combinent des données de corpus et des données expérimentales, ce qui permet de dépasser certaines limites associées aux corpus d'apprenants informatisés et de mieux pouvoir cerner et expliquer certains phénomènes.

L'étude de corpus de Meunier et Littré révèle, entre autres, que la plupart des erreurs se situent au niveau du choix du présent continu ou du présent simple. Ce choix constitue encore une importante source d'erreurs en troisième année. Les données expérimentales permettent de mieux cerner comment les apprenants comprennent les contraintes d'utilisation liées au choix du temps et de l'aspect. Par exemple, les résultats des tests montrent que, alors que les étudiants avancés sont capables de juger et de justifier correctement l'utilisation du présent continu dans des cas où l'aspect en cours est étendu/prolongé, ils éprouvent des difficultés avec l'utilisation du présent continu pour parler d'événements planifiés (Meunier & Littré, 2012 : 72).

⁴⁷ Soulignons au passage que l'approche qui consiste à croiser des observations sur corpus et des tests d'acceptabilité n'est pas sans problème, et a suscité de vifs échanges sur la liste d'informations *Corpora List* (voir Worlock Pope, 2010). En effet, si pour les uns le lien entre l'approche psycholinguistique et l'approche sur données attestées est acceptable (voire même souhaitable – Gries, 2010), pour les autres le fait de compléter des données attestées par de l'introspection n'a pas de sens dans une démarche proprement empirique (Louw, 2010 ; Teubert, 2010a, 2010b).

L'étude d'Osborne (2008) vise à expliquer la présence d'erreurs grammaticales persistantes/résiduelles dans des productions écrites d'apprenants d'anglais de niveau post-intermédiaire (par exemple l'omission du "-s" à la troisième personne au présent simple). Contrairement aux études mentionnées ci-dessus, elle ne se penche pas sur le développement sur une période donnée mais sur des phénomènes à un moment précis du développement. Les données de corpus proviennent des onze sous-corpus d'ICLE et du *Chambéry Corpus*, qui rassemble des productions rédigées par des étudiants d'anglais francophones en deuxième et troisième année à l'université. Les résultats de l'étude semblent indiquer que la persistance de certaines erreurs à un niveau relativement avancé est généralement induite par le contexte et est déclenchée par le fait que les apprenants s'essaient à des constructions plus complexes. L'étude distingue un certain nombre de contextes dans lesquels les apprenants d'anglais de niveau post-intermédiaire omettent encore le "-s" à la troisième personne singulier du présent simple. Ces contextes sont marqués par une certaine complexité syntaxique :

*Not to mention the long hours, the highly stressful environment, and **a workload** that *get heavier the better and more committed you are to your job. (Osborne, 2008 : 75, mots mis en évidence dans la version originale).*

*[Sans parler des longues heures, de l'environnement extrêmement stressant, et d'une charge de travail qui *devenir plus lourd au plus on est dévoué à son travail.]*

Osborne explique que, dans ce genre de contexte, l'utilisation de subordonnées peut entraîner un affaiblissement du lien entre le sujet et le verbe conjugué.

Paquot (2008) illustre comment un corpus qui contient des productions d'apprenants de plusieurs langues maternelles différentes (un corpus d'apprenants multi-L1) associé au modèle contrastif intégré ou *integrated contrastive model* (Granger, 1996 ; Gilquin, 2000, 2001) permet potentiellement d'identifier des cas de transfert de la langue maternelle et de comprendre ce qui favorise le transfert. Le modèle contrastif intégré combine l'analyse contrastive ou *contrastive analysis* et l'analyse contrastive d'interlangue ou *contrastive interlanguage analysis*. L'analyse contrastive d'interlangue associe deux types de comparaison :

- la comparaison de productions de locuteurs natifs et de locuteurs non-natifs, qui peut mettre au jour certains aspects de l'interlangue (par exemple des cas de suremploi et de sous-emploi) ;
- la comparaison de productions de locuteurs non-natifs de langues maternelles différentes (par exemple d'apprenants francophones et

d'apprenants germanophones), qui peut faire ressortir des différences entre plusieurs groupes de L1 pour la même L2.

L'étude de Paquot se concentre sur la phraséologie de l'exemplification à l'écrit et plus particulièrement sur les séquences *for example* et *for instance*, et les séquences qui contiennent les mots *example*, *illustrate* et *exemplify* dans des écrits argumentatifs d'apprenants d'anglais L2 de cinq langues maternelles différentes d'ICLE : néerlandais, français, allemand, polonais et espagnol. Le *Louvain Corpus of Native Speaker Essays (Locness)* est utilisé pour la comparaison natif vs non-natif. L'étude révèle que la séquence *let's/let us take the example of* est utilisée de manière récurrente par les apprenants francophones alors qu'elle n'apparaît pas dans le corpus *Locness* et qu'elle est plutôt rare dans les autres variétés d'apprenants étudiées. Paquot (2008 : 112) suggère que ce suremploi est très probablement un cas de transfert de la L1 étant donné qu'il existe un équivalent congruent en français (*prenons l'exemple de...*). Le modèle contrastif intégré comporte également un volet contrastif. Afin de tenter d'expliquer le suremploi de *let's/let us take the example of* par les étudiants francophones, Paquot étudie, d'une part, l'utilisation de cette séquence dans des corpus écrits de locuteurs natifs (*Locness*, le sous-corpus d'anglais académique du corpus *BNC-Baby* ainsi que le *British Academic Written Corpus, Bawe*) et, d'autre part, l'utilisation de la séquence *prenons l'exemple de...* dans le *Corpus de dissertations françaises (Codif)*, un corpus de productions écrites en français par des étudiants francophones. Il ressort de l'analyse contrastive que la séquence anglaise n'apparaît qu'une seule fois dans les corpus natifs utilisés, tandis que la séquence équivalente en français est plutôt fréquente dans le *Codif*.

Paquot examine ensuite l'utilisation de *let's/let us* dans les corpus natifs et non-natifs. Les résultats montrent que, alors qu'il n'y a que peu d'occurrences de *let's/let us* dans *Locness*, la séquence est sur-employée par les apprenants francophones dans ICLE. De plus, une analyse de l'expression équivalente en français, à savoir l'impératif à la première personne du pluriel (par exemple *citons, considérons*), dans *Codif* révèle que l'impératif est fréquemment utilisé pour organiser le discours dans des productions écrites formelles en français et que la fréquence de ces formes impératives dans *Codif* est très semblable à la fréquence des séquences introduites par *let's/let us* dans les productions écrites en anglais des apprenants francophones.

Paquot conclut donc que le transfert de forme peut s'accompagner d'un transfert de fréquence et de registre (cf. la notion de "savoir élocutionnel" et la question de la compétence

sociolinguistique en L2 – Tyne, à paraître ; cf. aussi la production à l'oral de "syntagmes erronés mais fluides correspondent à des préfabriqués dans la L1" – Hilton, 2008 : 84). Par ailleurs, comme le montre l'étude de Dubois *et al.* (à paraître), les erreurs à l'écrit en français L2 ne sont pas toujours comparables selon la L1, ce qui permet de voir que pour certains types d'apprenants (et pour certains types d'erreurs) le développement se fera plus ou moins vite et/ou facilement du fait de la connaissance des règles de la L1.

A l'oral, certaines études se penchent sur la « fluence » dans les corpus d'apprenants, en particulier dans une perspective psycholinguistique, cherchant à mettre en avant le lien entre le traitement (c'est-à-dire tout le processus qui a pour résultat la mise en place du message verbal) et la production plus ou moins fluide (par ex. Towell *et al.*, 1996 ; Hilton, 2008). Dans une étude récente, Brand et Götz (2013) explorent la corrélation potentielle entre la fluence (mesurée en termes de vitesse d'élocution et de fréquence et position des pauses remplies et vides) et la correction formelle⁴⁸ (mesurée en termes de nombre d'erreurs par 100 mots) dans le sous-corpus d'apprenants d'anglais germanophones de niveau intermédiaire-avancé du *Louvain International Database of Spoken English Interlanguage* : Lindsei ; Gilquin *et al.*, 2010). Il apparaît qu'il n'y a pas de corrélation claire entre les paramètres de la fluence temporelle étudiés et la correction formelle dans les interviews informelles du corpus (Brand & Götz, 2013 : 131). En marge de cette étude, il a été demandé à des locuteurs natifs d'évaluer le niveau de compétence langagière de certains apprenants. Les résultats semblent suggérer que les variables temporelles ont un peu plus d'impact sur la perception de la compétence par les locuteurs natifs que le niveau de correction formelle (voir plus loin la question de la fluence apparente, notamment d'après Guillot, 1999). Ceci étant, ni les variables temporelles, ni la correction formelle ne paraissent avoir un impact significatif sur cette perception, ce qui indiquerait que des bonnes performances au niveau de plusieurs variables sont plus pertinentes quand on se concentre sur la perception (Brand & Götz, 2013 : 133).

Par ailleurs, Osborne (2013) examine les liens entre la fluence temporelle (vitesse d'élocution en mots par minute, temps d'hésitation et longueur moyenne des contributions en mots), la complexité syntaxique et la présentation de l'information (*informational content*) dans des sous-corpus de productions orales en anglais et en français du corpus *Parole*. Les productions orales étudiées sont des récits basés sur une même séquence vidéo. Les récits ont été produits

⁴⁸ Il apparaît que dans la plupart des études sur l'acquisition de L2, qui adoptent une approche plutôt cognitiviste, le travail sur la « fluence » s'apparente à celui sur le degré de correction (*accuracy*) et à celui de la complexité (voir Housen & Kuiken 2009). Voir plus loin sur ce point.

par des locuteurs non natifs et par des locuteurs natifs de l'anglais et du français. Il ressort de l'étude que les locuteurs les plus « fluents » sont ceux qui parviennent à communiquer le contenu du message de manière efficace en alliant peu d'hésitations et qui font des choix lexicaux et syntaxiques qui leur permettent de présenter et transmettre l'information de manière plutôt économique (Osborne, 2013 : 139). Les locuteurs les moins « fluents », au contraire, ont tendance à utiliser des structures syntaxiques complexes dans lesquelles l'information est plus diluée (par exemple, lorsque des circonlocutions sont utilisées).

Nous reviendrons sur la question de la « fluence » et notamment sur le problème de la prise en compte de la fluence interactionnelle en L2.

3. Applications pédagogiques

Dans un article publié en 2009 (*The contribution of learner corpora to second language acquisition and foreign language teaching*), Granger (2009 : 14) déplore le fait que les recherches basées sur les corpus d'apprenants n'ont encore que très rarement donné lieu à des applications pédagogiques concrètes. Ce constat est en grande partie encore valable aujourd'hui – et ne concerne pas uniquement les recherches sur des corpus L2 mais bien l'ensemble des recherches sur corpus pour l'enseignement des langues (McCarthy, 2008 ; Römer, 2006, 2009). Si l'on parcourt les actes du premier colloque entièrement consacré à la recherche sur les corpus d'apprenants (*first Learner Corpus Research Conference, LCR 2011* : Granger *et al.*, 2013) ainsi que les résumés des communications de la deuxième édition du colloque en septembre 2013 (<http://lcr2013.b.uib.no/files/2013/09/abstracts-book.pdf>), il apparaît clairement que, bien qu'une série de contributions mentionne brièvement à la fin des applications possibles, seule une poignée d'articles consacre le corps du texte à ce genre d'applications. Certes, les caractéristiques de l'interlangue mises au jour par des études de corpus d'apprenants ne doivent pas nécessairement donner lieu à des applications pédagogiques (Granger, 2009). Il convient de poser la question de la pertinence pédagogique : les besoins spécifiques d'un groupe d'apprenants, les objectifs d'un programme d'enseignement donné sont autant d'éléments qui doivent guider les choix opérés par les chercheurs et les professionnels de l'enseignement et de l'acquisition des langues. Mais au-delà de ce constat, il existe des freins, des contraintes, qui font qu'il n'est pas toujours évident de relier le monde de la recherche sur l'acquisition et le monde de l'enseignement (voir Boulton & Tyne, 2015).

Dans cette section les applications pédagogiques abordées sont divisées en trois sous-catégories qui concernent (1) le CECR, (2) les outils de références, et (3) les activités en classe.

3.1. Corpus d'apprenants et CECR

Le CECR a été développé dans le but de pouvoir offrir "*une base commune pour l'élaboration de programmes de langues vivantes, de référentiels, d'examens, de manuels, etc. en Europe*" (Conseil de l'Europe, 2000 : 9). Six niveaux communs de référence, à savoir les niveaux A1, A2, B1, B2, C1 et C2, et une série de descripteurs de compétence ont été mis au point. Le cadre et les descripteurs ont été formulés de manière à ce qu'ils puissent être transposables à n'importe quelle langue cible pour des apprenants de langues maternelles différentes. Les descripteurs ne sont dès lors délibérément pas spécifiques à une langue. Ce manque de spécification est cependant problématique quand on se penche sur des contextes bien précis. Par exemple, Thewissen (2012a : 236) fait remarquer que, pour que les descripteurs et les niveaux de compétence puissent être plus aisément utilisés dans des situations concrètes, il conviendrait de pouvoir formuler, d'une part, des descripteurs spécifiques à une L2 déterminée (voir par ex. Beacco *et al.*, 2005, pour le français), et, d'autre part, des descripteurs spécifiques à des couples L1-L2 bien précis. Par exemple, les recherches menées dans le cadre du projet *English Profile* montrent comment des études basées sur un corpus d'apprenants de très grande taille (le *Cambridge Learner Corpus* – corpus qui a été enrichi d'étiquettes d'erreurs) permettent de spécifier les descripteurs pour l'anglais L2 et d'identifier des caractéristiques linguistiques discriminantes pour chacun des six niveaux de compétence (Hawkins & Buttery, 2009, 2010). Thewissen (2012a) est un exemple de tentative de formulation de descripteurs spécifiques pour un couple L1-L2 donné : l'étude vise à proposer des descripteurs pour les francophones apprenant l'anglais en tant que L2. Thewissen prétend que ce genre d'approche est particulièrement pertinent pour l'enseignement et l'évaluation à un niveau local car elle permet aux enseignants de s'adapter aux besoins spécifiques d'un groupe d'apprenants d'une L2 déterminée qui partagent la même langue maternelle. Les descripteurs pour les francophones apprenant l'anglais ont été formulés suite à des recherches menées dans le sous-corpus francophone enrichi d'étiquettes d'erreurs d'ICLE (cf. plus haut). Les descripteurs proposés concernent plus particulièrement la correction formelle pour les niveaux de compétence B2 et C1 – ces deux niveaux sont traités ensemble, les recherches de

Thewissen ayant montré qu'il n'y avait pas de différences significatives entre ces deux niveaux quand les erreurs sont prises en compte.

L'étude qualitative des erreurs dans le corpus permet à Thewissen de proposer des grilles de description des erreurs attestées aux niveaux B2/C1 pour les francophones apprenant l'anglais pour la correction grammaticale, la maîtrise du vocabulaire, la maîtrise de l'orthographe, la maîtrise de la ponctuation et la cohésion. Thewissen (2012a : 251) précise qu'à ce stade ces grilles donnent un aperçu des erreurs qui sont en théorie typiques/discriminantes pour les francophones qui ont atteint les niveaux B2/C1 en anglais. Ces grilles sont accompagnées de banques d'exemples tirés du corpus analysé afin d'illustrer concrètement ce à quoi les descripteurs font référence. Pour mieux comprendre la contribution de l'étude de Thewissen, le Tableau 1 contient à la fois les descripteurs non spécifiques du CECR ainsi qu'un extrait des descripteurs (et des exemples) pour les francophones apprenant l'anglais proposés par Thewissen pour la correction grammaticale aux niveaux B2/C1.

Tableau 1 - Descripteurs pour la correction grammaticale aux niveaux B2/C1.

CECR	<p>[B2] Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.</p> <p>[B2+] Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.</p> <p>[C1] Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.</p>
Thewissen, 2012a	<p>Good grammatical control overall but some errors remain, both in more subtle but also sometimes more basic grammatical areas. There is still noticeable mother tongue influence from French. Errors are likely to be found mainly, though not exclusively, in the following areas.</p> <hr/> <p>Articles, especially $\emptyset \rightarrow the$ and $the \rightarrow \emptyset$ pairs. <i>In this essay I will deal with the intricate and tricky problem of (GA) the \$0\$ feminism. (the$\rightarrow \emptyset$)</i></p> <p>0 \$The\$ law is a very strict institution and we should all respect it. ($\emptyset \rightarrow the$)</p> <hr/> <p>Tenses, sometimes despite the presence of overt temporal markers: errors occur on a variety of tenses, including cases of confusion between present continuous vs. simple present. <i>Actually, some people are sitting or lying \$sit or lie\$ in front of television for hours and receive the sound and the picture in a passive way.</i></p> <p><i>For instance, from 92 we will no longer use francs, pounds, pesetas but ECUs. When we go from one country to another we don't need \$will not need\$ to show our passports or identity cards.</i></p>

Noun agreement, in contexts such as generic use, distributive plurals, after the use of the phrase 'one of', or when nouns are used in adjectival function.

*The next century should be based on better human **relationship** **\$relationships\$** and understanding.*
[=generic reference]

*The rise of barren land in Africa is due to **cash-crops** **\$cash-crop\$** policies.*
[=noun with adjectival function]

*One of the **function** **\$functions\$** of religion was to prevent the masses from having access to political consciousness*
[= 'one of' + plural]

Adverb placement, with recurrent use of the erroneous verb+adverb+noun phrase pattern.

*One should try to **define precisely** **\$precisely define\$** the term "nation".*

*Solidarity **plays also** **\$also plays\$** an important part*

Osborne (2011 : 183), quant à lui, s'est attaché à examiner la contribution possible des corpus de productions orales d'apprenants et de locuteurs natifs pour l'évaluation de la compétence orale, et plus particulièrement de la fluence, dans le cadre du CECR. Osborne part du constat que les descripteurs pour la fluence dans le CECR décrivent certains phénomènes de dysfluence (tels que les pauses, les faux démarrages et les reformulations) comme étant plus évidents jusqu'au niveau B1. Il n'est cependant pas précisé si ces phénomènes sont absents ou moins évidents à partir du niveau B2. Osborne fait également remarquer que les faux démarrages et les reformulations ne se retrouvent pas uniquement dans les productions orales des apprenants ; ils figurent aussi dans les corpus oraux de locuteurs natifs (cf. André & Tyne, 2012). Il propose de mesurer la fluence de manière quantitative et qualitative, ce qui lui permet de définir des profils de fluence pour chaque niveau. Par exemple, un locuteur qui affiche une vitesse d'élocution de moins de 90 mots par minute ou un taux de pauses de plus de 40 % n'aura très probablement pas atteint un niveau supérieur au niveau B1. Ces profils sont accompagnés d'exemples de productions orales d'apprenants pour illustrer les différents niveaux du CECR du point de vue de la fluence. De tels exemples peuvent en outre s'avérer être très utiles lors de la formation des personnes qui sont amenées à évaluer la compétence orale.

Quant à l'élaboration du CEFRL pour le Japon (par ex. Tono & Negishi, 2012) les différents paramètres pris en compte dans l'élaboration des descripteurs combinent les données de sources variées (corpus de descripteurs, corpus de manuels, corpus d'apprenants, etc.) et ce afin d'avoir un maximum d'informations servant à la formulation des descripteurs sous forme de "can-do statements" ("je peux faire...") pour la mise en application du référentiel.

Le volume de Bartning *et al.* (2010) présente une série d'études qui traitent de la question du développement de la L2 (cf. Bartning & Schlyter, 2004) et les correspondances possibles entre ce développement et les niveaux du CECR. Si les différents niveaux du CECR trouvent des échos dans l'étude des corpus d'apprenants, il apparaît que les descripteurs ne sont pas toujours précis quant au développement de la L2 ou que les différences entre les niveaux sont plus perceptibles dans certains cas que dans d'autres. Ce type de travail est d'autant plus important qu'il permet de confronter l'approche du CECR qui met l'accent sur la compétence communicative (le fait de pouvoir réaliser telle ou telle tâche convenablement d'un point de vue communicatif) et l'approche linguistique qui s'appuie sur des données produites par les apprenants en train de réaliser certaines activités.

3.2. Corpus d'apprenants et outils de référence

Le *Macmillan English Dictionary for Advanced Learners* (MED2 ; Rundell, 2007) constitue un exemple de dictionnaire d'apprenants de référence qui a intégré des composantes élaborées à partir de recherches basées sur un corpus d'apprenants afin d'aider les utilisateurs à faire face à des difficultés récurrentes à l'écrit. Les recherches reposent sur l'analyse minutieuse d'ICLE et d'un corpus de 15 millions de mots d'anglais académique. Trois composantes ont ainsi été intégrées : 100 notes d'erreur *Get it right (GIR)*, six sections grammaticales (*Grammar sections*) et douze sections portant sur l'anglais académique (*EAP Writing sections*). Le rôle des notes d'erreurs est d'attirer l'attention des apprenants sur certaines utilisations erronées d'un mot ou d'une séquence de mots. Les mots et séquences qui font l'objet de ces notes ont été identifiés et sélectionnés sur base de l'étude minutieuse d'une partie d'ICLE enrichi d'étiquettes d'erreurs (680.000 mots). Les éléments sélectionnés ont ensuite été analysés dans l'entièreté d'ICLE disponible lors de cette étude (3,5 millions de mots). Etant donné que le dictionnaire est destiné à des apprenants d'anglais de niveau avancé quelles que soient leurs langues maternelles, la fréquence et la présence d'une erreur dans les productions d'apprenants d'au moins cinq langues maternelles différentes ont été prises en compte lors de la sélection. Comme le montre la Figure 1, les notes d'erreurs, qui sont intégrées au sein des entrées pour les mots et séquences sélectionnés, comprennent plusieurs éléments : des exemples d'usages erronés tirés du corpus d'apprenants (ces exemples sont précédés d'une croix et les éléments erronés sont barrés), une explication de la source du problème ainsi que des conseils pour aider les apprenants à corriger ces erreurs et à les éviter.

Figure 1 - Exemple de note d'erreur *Get it right (MED2)*

Get it right: damage

When **damage** means 'harm or injury' it is an uncountable noun, and so:

- it is never used in the plural
- it never comes after **a** or a number

✗ These toxins can cause **damages** to the lungs and brains.

✓ These toxins can cause **damage** to the lungs and brains.

✗ They should consider the serious **damages** that their decisions may cause.

✓ They should consider the serious **damage** that their decisions may cause.

✗ **A great damage** has been done to agriculture, forests, and people's health.

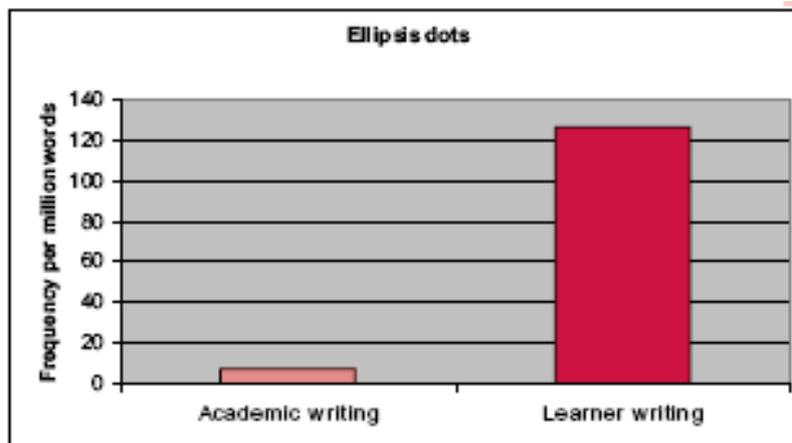
✓ **Great damage** has been done to agriculture, forests, and people's health.

The plural form **damages** is a specialized legal term meaning 'money that a court orders you to pay someone because you have harmed them or their property'.

Mr Galloway was awarded substantial **damages**.

Les six sections grammaticales et les douze sections portant sur l'anglais académique apparaissent au milieu du dictionnaire au sein d'une partie intitulée *Improve your writing skills (Améliorez vos compétences écrites)*. Les sections grammaticales sont basées sur l'analyse d'erreurs mentionnée ci-dessus. Elles ciblent les difficultés rencontrées fréquemment dans le corpus d'apprenants par rapport à l'utilisation des articles, de la complémentation (noms, verbes et adjectifs), des noms (in)dénombrables ou de la ponctuation, par exemple (De Cock *et al.*, 2007). Les sections sur l'anglais académique se focalisent sur l'organisation du discours dans des écrits académiques et traitent notamment de l'exemplification, l'expression d'opinions personnelles ou de la causalité. La rédaction de ces sections a été guidée par une analyse détaillée de 350 marqueurs d'anglais académique identifiés par Paquot (2010) dans l'ensemble d'ICLE et dans des corpus natifs (voir Gilquin *et al.*, 2007 ; Gilquin & Paquot, 2008). Outre des exemples issus d'ICLE et des notes d'erreurs spéciales, les sections grammaticales et académiques incluent également des données de fréquence qui visent à permettre aux apprenants de prendre conscience de certains cas de suremploi à l'écrit (cf. la Figure 2).

Figure 2 - Fréquence de points de suspension dans des corpus d'anglais natifs et non-natif (section sur la Ponctuation, MED2)



Le *Louvain English for Academic Purposes Dictionary (Lead)* est une ressource électronique qui est à la fois un dictionnaire d'anglais académique et un outil d'aide à la rédaction ("*dictionary-cum-writing aid*" – Paquot, 2012). L'élaboration de cette ressource électronique repose en partie sur l'analyse de deux corpus multi-L1 : ICLE et *Vespa (Varieties of English for Specific Purposes dAtabase*, Paquot *et al.*, 2013), un corpus d'apprenants qui contient des textes académiques d'anglais sur objectifs spécifiques. L'analyse des corpus d'apprenants a guidé la sélection des entrées et des collocations, l'inclusion d'exercices ciblés, ainsi que l'intégration de notes d'erreurs. Une des particularités de *Lead* est que cette ressource est très souple ; elle peut être adaptée au domaine d'étude (les exemples inclus dans les entrées pour les étudiants de médecine proviennent de revues scientifiques de médecine) et à la langue maternelle de l'utilisateur (les étudiants francophones peuvent entrer un mot en français et voir les traductions qui sont proposées ; ils peuvent ensuite avoir accès à l'entrée du mot anglais avec des notes d'erreurs qui concernent uniquement les apprenants francophones).

Chuang et Nesi (2006) expliquent comment elles ont pu mettre au point un programme de remédiation en ligne axé essentiellement sur la grammaire (*GrammarTalk*) pour répondre aux besoins d'étudiants chinois qui étudient au sein d'universités anglophones. Ces étudiants, qui éprouvent le besoin de progresser en grammaire, suivent des cours d'anglais académique dans le cadre desquels il n'est pas possible d'aborder la grammaire. Les modules du programme de remédiation en ligne accessible aux étudiants en dehors de ces cours ont été élaborés suite à une analyse d'erreurs dans un corpus d'écrits académiques en anglais par des étudiants dont le mandarin est la langue maternelle. Le corpus de 88.000 mots a été enrichi d'étiquettes d'erreurs et l'analyse a permis d'identifier une série de problèmes récurrents. La majorité des

erreurs identifiées dans le corpus concerne l'utilisation des articles. Ainsi, le premier module proposé porte sur cet aspect de la grammaire anglaise. Ce module comprend des explications et des exercices et inclut des extraits du corpus d'apprenants ainsi que du corpus *Bawe* (cf. ci-dessus).

3.3. Utilisation de corpus d'apprenants en cours de langue

Ce que l'on appelle l'apprentissage sur corpus (*data-driven learning* en anglais – Johns, 1991), c'est-à-dire le fait de mettre l'apprenant dans une situation où il peut observer des données réunies sous forme de corpus, peut s'appliquer à des données provenant de locuteurs natifs ou non, même si la plupart des études existantes se basent sur des données produites par des locuteurs natifs, ou, en tout cas, par des données produites par les apprenants eux-mêmes.

Or, de façon générale, comme prétendent Boulton et Tyne (2014), la méthodologie de travail sur corpus (quel que soit le travail spécifique) est favorable à des pratiques d'observation de ses propres productions, ou des productions que l'on "aurait pu produire" si l'on travaille à partir de données d'autres apprenants. Comme le montre Seidlhofer (2002), le fait de faire travailler des apprenants sur leurs propres productions sous forme de corpus appelés parfois corpus locaux d'apprenants ou *local learner corpora* (Mukherjee & Rohrbach, 2006) permet aux apprenants de se sentir concernés, de chercher à améliorer leurs productions. De plus, des activités sur corpus locaux d'apprenants apportent la possibilité aux enseignants de donner aux apprenants du *feedback* sur mesure (Mukherjee, 2009) et d'examiner de manière détaillée les performances de leurs propres élèves/étudiants tant au niveau individuel qu'au niveau du groupe (Pérez-Paredes, 2003). Ce genre d'approche présuppose que les enseignants ne soient pas complètement étrangers à certains aspects de la linguistique de corpus grâce à des programmes de formation continue ou de publications (comme Millar & Lehtinen, 2008, par exemple). Cependant, il n'est pas évident des enseignants, qui n'ont du reste pas tous un goût prononcé pour la linguistique, qu'ils deviennent des linguistes sur corpus. Voilà pourquoi pour certains une approche plus "*soft*" (cf. Gabrielatos, 2005), ou cherchant simplement à mettre en avant un certain "*esprit de corpus*" (Boulton & Tyne, 2014 : 81) serait plus favorable (du moins dans un premier temps).

Les activités d'apprentissage sur corpus locaux d'apprenants apparaissent comme étant tout particulièrement indiquées quand il s'agit d'aider les apprenants à prendre conscience des erreurs persistantes qu'ils commettent (fonction correctrice) ou du suremploi de certains mots ou séquences de mots (Granger & Tribble, 1998 ; Granger, 2002 ; Nesselhauf, 2004 ;

Mukherjee, 2009). Comme l'a fait remarquer Granger (2002 : 26), l'utilisation de données issues de corpus d'apprenants en classe est cependant controversée : exposer les apprenants à des erreurs pourrait renforcer ces dernières, même si dans le domaine de l'acquisition des L2 on constate les effets positifs du *noticing* (Schmidt, 1990) dans des cas où les données comportent des erreurs (qui sont corrigées ou renégociées par l'apprenant avec la possibilité de modifier l'output ; McDonough, 2005). Selon Nesselhauf (2004) ainsi que Granger et Tribble (1998), ce risque est considérablement réduit si l'activité d'apprentissage comporte également de nombreux exemples d'utilisation correcte (preuves positives) et si elle est suivie par des exercices de consolidation. Lors de l'élaboration d'activités d'apprentissage basées sur des corpus d'apprenants, Mukherjee (2009) souligne, à juste titre, qu'il est important de ne pas assimiler les corpus d'apprenants uniquement à des sources d'utilisations erronées (preuves négatives) ; ces corpus contiennent également des exemples d'utilisations correctes qui devraient être intégrées dans les activités. Mukherjee (2009) illustre cette approche, qui permet par ailleurs de renforcer la motivation des apprenants, avec une activité qui vise à améliorer les compétences collocationnelles des apprenants. Cette activité a été élaborée suite au constat que les apprenants d'anglais de la composante germanophone de *Lindsei* (cf. ci-dessus) ont tendance à utiliser le verbe *make* avec le substantif *experience*. Dans un premier temps, les apprenants sont invités à observer et à commenter des lignes de concordance du corpus d'apprenants où la forme *experience* est accompagnée, d'une part, de formes du verbe *make* (input négatif), et, d'autre part, de formes du verbe *have* (input positif). Cette première phase est suivie d'exercices de consolidation.

Par ailleurs, dans un travail sur les productions orales, Lynch (2001, 2007) se sert de la transcription afin d'amener les apprenants à travailler sur leurs propres productions par le biais du *noticing*. Si dans ce cas de figure le travail ne porte pas sur le corpus comme produit fini, il concerne néanmoins un des éléments de constitution d'un corpus. Comme le soulignent Boulton et Tyne (2013, 2014), de telles activités contribuent à leur façon à une rencontre fructueuse entre apprenants et données, nous invitant à adopter une définition plus ouverte du *data-driven learning*. Justement, l'apprentissage sur corpus ne doit pas s'arrêter à l'étude des données réunies sous forme de corpus. Ainsi, par exemple, dans le travail autour de la transcription de données orales, le fait d'être au contact de la langue permet aux apprenants de saisir des détails qui ne sont généralement pas présentés dans les manuels (cf. la question de la compétence sociolinguistique dans Tyne, 2009a), sans parler des atouts de l'auto-transcription en tant qu'activité de *noticing* et d'auto-correction (voir aussi Stillwell *et al.*, 2010 ; Mennim,

2012). C'est un domaine encore peu développé mais qui laisse entrevoir des résultats tout à fait intéressants pour l'enseignement des langues (voir Paternostro, à paraître).

4. Eléments de discussion et perspectives pour l'acquisition des L2

Les corpus d'apprenants sont pour la plupart des corpus de données contrôlées, pour ne pas dire "sollicitées" pour reprendre le terme historique (Corder, 1973). Si le contrôle des données n'est certes plus ce qu'il était à ses débuts, dans la mesure où les corpus comprennent aujourd'hui notamment des textes entiers produits par des apprenants, il n'en demeure pas moins que derrière les données ainsi réunies se trouvent en général les "tâches" des apprenants, avec parfois des hypothèses spécifiques préexistantes en amont de la collecte de données. Par ailleurs, les corpus plus généraux, échantillonnés ou orientés selon des critères plutôt fonctionnels, génériques ou sociolinguistiques font défaut.

Si les corpus de données en L1 s'organisent généralement en fonction des types ou genres de production ou par période, etc., les corpus d'apprenants s'organisent en revanche en fonction des types de tâches : rédaction sans préparation ; description ou narration à partir d'un stimulus, d'une image, etc. Ainsi, il apparaît que les corpus d'apprenants sont résolument non écologiques (à quelques exceptions près – voir par exemple Dubois *et al.*, 2014, et le corpus *Vespa*, Paquot *et al.*, 2013), et résolument non littéraires. Leur richesse générique n'est toutefois pas comparable à celle des corpus que l'on possède pour les locuteurs natifs (articles de journaux, œuvres littéraires, débats radiophoniques, etc.). Ce problème, que Granger (2002) aborde en termes d'"authenticité" des données (voir également la notion de corpus d'apprenants "prototypique" de Nesselhauf, 2004), a pour conséquence de rendre assez difficiles certaines comparaisons entre données issues de L1 et données issues de L2. Prenons l'exemple de la richesse lexicale (cf. Milton, 2010). On sait que celle-ci relève de deux critères en particulier (outre le niveau en langue, par exemple) : le type de production et la longueur de la production (plus un texte est long, plus il devient difficile d'augmenter le nombre de nouveaux items lexicaux par rapport au nombre total de mots produits). Ainsi, toute généralisation à partir de mesures de ce type reste extrêmement problématique étant donné la difficulté qu'il y a à comparer des types de productions et des longueurs de corpus différents, même s'il existe des formules pour comparer la densité lexicale dans des corpus ou textes de longueurs différents, notamment à travers un échantillonnage répété. Certes, il est possible de combiner la méthodologie de l'étude de corpus avec des tests de réception de formes cibles (généralement hors contexte), mais ceci ne résout pas le problème de la comparabilité des

données collectées sous forme de corpus. Par ailleurs, s'il est possible de se baser sur des critères communicationnels généraux (cf. Bilger & Tyne, 2009), comme formel/informel, monologue/dialogue afin d'observer les mêmes types de variation, par exemple, dans des corpus de L1 et L2, la comparaison de corpus strictement équivalents demeure problématique. Il est à noter que certains projets de corpus d'apprenants comportent un volet de collecte de corpus natifs afin de tenter de minimiser ces problèmes de comparabilité. Ainsi, dans le cadre du projet *Lindsei*, le corpus natif *Locnec* (*Corpus of Native English Conversation – Locnec* : De Cock, 2004) a été rassemblé suivant la même procédure que les sous-corpus d'apprenants d'anglais de langues maternelles différentes ; *Lindsei* et *Locnec* comprennent des interviews informelles (format et tâches identiques) avec des étudiants à l'université (voir aussi De Cock, 2002).

Si le domaine des études sur l'acquisition de L2 paraît parfois relativement homogène (ou tout simplement moins éclaté que celui de l'étude des données en L1) avec des protocoles et des conventions largement partagés d'une étude à l'autre, d'un groupe de chercheurs à l'autre (par ex. *Chat / Clan* : <http://chilides.psy.cmu.edu>), d'une époque à l'autre, on pourrait reprocher au champ un relatif manque de vue d'ensemble malgré la complicité sur un bon nombre de points, notamment méthodologiques. Cette critique est en fait plus générale, et s'applique aussi (et avant tout) à la linguistique de corpus, souvent taxée de description stérile (cf. la caricature décrite par Fillmore, 1992). Mais en ce qui concerne le domaine de l'acquisition des L2, à partir des années 1990 on assiste à un certain morcellement du champ, notamment avec la prise en compte de facteurs personnels et sociaux affectant l'acquisition (Mitchell *et al.*, 2013) : à chaque étude son corpus, à chaque étude ses résultats, etc. La recherche sur l'acquisition des L2 entre alors dans une phase où la multiplication des monographies montre qu'il est pratiquement possible de tout prouver... à condition d'avoir les « bonnes » données, le « bon » corpus, etc.

Par ailleurs, même dans le travail à partir de corpus d'apprenants (pris au sens large), on trouve certaines restrictions, comme :

- des cadres théoriques dominants (on peut citer le poids de la Grammaire universelle, par exemple – Mitchell *et al.*, 2013) ;
- des modèles descriptifs dominants (notamment celui de la grammaire des constituants pour ce qui est de la description syntaxique) ;
- des hypothèses dominantes (par ex. dans l'analyse de la fluence – voir ci-après) ;

- des cadres analytiques dominants (par ex. le variationnisme labovien pour l'étude de la compétence sociolinguistique – voir Dewaele & Mougeon, 2002 ; Dewaele, 2007) ;
- etc.

Dans la littérature traitant de l'acquisition, la fluence en L2 renvoie avant tout aux facteurs temporels, ou à la régularité dans les productions, avec l'étude du débit mais aussi de phénomènes d'hésitation. En termes de modélisation de la fluence, on s'intéresse beaucoup à la façon dont l'apprenant mobilise les différents éléments du système L2 pour que la production devienne automatisée (par ex. Towell *et al.*, 1996). Il s'agit donc principalement de voir comment les apprenants développent une certaine spontanéité ou rapidité d'exécution dans leurs productions. Mais comme le précisent André et Tyne (2012 : 24), une telle approche "*ne permet pas de saisir d'autres aspects davantage liés au fait que le langage dans l'échange n'est pas simplement le résultat d'un processus de planification-production par l'individu*". Et comme le montre l'étude d'André et Tyne (2012), à l'instar de Guillot (1999), en travaillant sur des éléments dus au mode oral de production établis par le Gars pour le français L1 (par ex. Blanche-Benveniste *et al.*, 1990), il est tout à fait possible de voir en quoi la soi-disant "dysfluence" en L2 peut être vue comme le témoin d'une fluence effective : *hésitation* devient alors *gestion de la parole en interaction* (cf. McCarthy, 2010) ; *correction d'erreurs* devient *stratégie de travail sur le texte oral*, etc. (cf. Bilger & Tyne, 2007). Les travaux sur la langue parlée (L1), mettant en avant la question du travail sur le texte oral, le phénomène de locuteur collectif, etc., déjà présents dans l'ouvrage de Guillot (1999), restent largement à l'écart des travaux sur la L2. Le paradigme traitant de la fluence (aux côtés de l'étude du taux d'erreurs et de la complexité) y semble davantage tourné vers une prise en compte de l'élaboration linguistique et du traitement cognitif. Mais comme nous l'évoquions plus haut, dans une prise en compte de la fluence en lien avec le CECR, il paraît important d'élargir la prise en compte du phénomène afin de valoriser une certaine gestion "dysfluente" de la parole en situation de communication réelle lorsque cette dysfluence contribue à la construction du discours, à l'implication dans l'interaction, etc. (cf. McCarthy, 2010).

En termes d'analyse des données, la dichotomie classique opposant natif et non-natif ou bien L1 et L2 est potentiellement problématique. Cette distinction s'appuie sur une certaine idéologie des groupes ou des objets d'étude qui, en réalité, sont sans doute moins stables, ou plus perméables qu'on ne le pense (Davies, 2003). La question de la double ou de la multicompétence a déjà été posée dans le contexte du travail sur l'acquisition (par ex. Cook, 1992), ou pour ce qui est de l'enseignement des langues (Davies, 1995). Le CECR s'appuie

notamment sur l'existence d'un répertoire plurilingue et non plus monolingue des locuteurs dans le contexte européen, ce qui laisse entrevoir des situations en cours ou à venir où il deviendrait difficile d'ériger des frontières nettes entre catégories historiquement étanches (du moins dans le contexte européen classique) dans la mesure où la présence conjointe et précoce de plusieurs langues maîtrisées à des degrés divers est une réalité que connaissent de plus en plus d'individus.

Par ailleurs, d'un point de vue didactique et sociolinguistique, on pourrait plaider en faveur d'approches plus ouvertes, traitant de manière plus générale des données comme, par exemple, celles provenant de locuteurs dits en situation d' "insécurité langagière" (cf. André & Adami, 2012) quelle que soit leur L1, L2, L3, etc. On pourrait alors peut-être trouver des recoupements entre publics que l'on oppose *a priori* mais qui sont semblables à certains égards ou pour certains types de productions, etc. On envisagerait alors une approche basée non pas sur tout ce qui oppose les différents locuteurs des langues (qu'ils soient natifs, pseudo-natifs, quasi-natifs, néo-natifs, natifs tardifs, etc.) mais sur ce qu'ils ont en commun. C'est précisément une démarche sur corpus qui permettra de bâtir une telle approche.

5. Conclusion

Comme nous l'avons vu dans cet article, les recherches en acquisition de L2 s'appuyant sur des corpus d'apprenants ont permis d'avancer considérablement dans la prise en compte du développement des diverses L2. De plus, les applications du domaine forment une base de réflexions permettant de faire évoluer notre prise en compte des données de L2, comme en ce qui concerne la place de la lexico-grammaire ou de la phraséologie, par exemple. Elles ont aussi permis de compléter un certain nombre d'éléments qui relèvent du champ de la didactique des langues, notamment en lien avec le CECR. Mais, comme nous l'avons souligné également, il y a encore des développements possibles ou des éléments qui pourraient être renforcés, en particulier pour les types de corpus utilisés, tels les corpus écologiques, travail sur les genres, comportant davantage de données orales et en interaction, pour le suivi longitudinal des apprenants (plus de vrais corpus longitudinaux) ou encore en prenant en compte des données pour davantage de L2 et de couples L1-L2 différents. Si certaines applications pédagogiques font encore preuve d'un manque d'utilisation des corpus et des techniques de travail sur corpus (sans parler des corpus d'apprenants), nous pouvons souligner que les applications indirectes, la production d'outils de référence ou d'apprentissage, sont d'une grande utilité et devraient continuer de progresser. Les avancées technologiques dans le

monde de la didactique font qu'il est de plus en plus facile d'envisager pour les applications plus directes (utilisation de corpus) un certain type d'approche des langues basé sur requêtes et moteurs de recherche (Boulton & Tyne, 2014). Enfin, il serait utile que les enseignants qui travaillent déjà à partir de corpus d'apprenants locaux se fassent davantage l'écho de leurs initiatives pour que l'on sache ce qui se passe et ce qui est efficace pour eux et leurs apprenants.

Références

Bibliographie

André, V. & Adami, H. (2012). "Vers le français langue d'intégration et d'insertion (FL2I)". In Adami, H. & Leclercq, V. (dir.). *Les migrants face aux langues des pays d'accueil : acquisition en milieu naturel et formation*. Lille : Septentrion. pp. 277-289.

André, V. & Tyne, H. (2012). "Compétence sociolinguistique et dysfluence en L2". In Kamber, A. & Skupien, C. (dir.). *Recherches récentes en FLE*. Berne : Peter Lang. pp. 21-46.

Bartning, I. & Schlyter, S. (2004). "Itinéraires acquisitionnels et stades de développement en français L2". *Journal of French Language Studies*, vol. 14, n° 3. pp. 281-99.

Bartning, I., Martin, M. & Vedder, I. (2010). *Communicative proficiency and linguistic development : intersections between SLA and language testing research*. Eurosla Monographs Series, 1. Disponible en ligne. <http://eurosla.org/monographs/EM01/EM01tot.pdf>

Beacco, J.-C., Ferrari, M. de, Lhote, G. & Tagliante, C. (2005). *Niveau A1.1 pour le français. Public adultes peu francophones, scolarisés, peu ou non scolarisés. Référentiel et certification (DILF) pour les premiers acquis en français*. Paris : Didier.

Bilger, M. & Tyne, H. (2007). "Speaking style and oral text construction in second language French". In Abecassis, M., Ayooso, L. & Vialleton, E. (dir.). *Le français parlé au 21^e siècle : normes et variations dans les discours et en interaction* (tome 2). Paris : L'Harmattan. pp. 173-189.

Bilger, M. & Tyne, H. (2009). "Variation in first and second language French : the case of 'parce que'". In Beeching, K., Armstrong, N. & Gadet, F. (dir.). *Sociolinguistic variation in French*. Amsterdam : John Benjamins. pp. 233-259.

Blanche-Benveniste, C., Bilger, M., Rouget, C. & van den Eynde, K. (1990). *Le français parlé : études grammaticales*. Paris : Éditions du CNRS.

Boulton, A. & Tyne, H. (2013). "Corpus linguistics and data-driven learning : a critical overview". *Bulletin Suisse de Linguistique Appliquée*, vol. 97. pp. 97-118.

Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris : Didier.

Boulton, A. & Tyne, H. (2015). "Corpus-based study of language and teacher education". In Bigelow, M. & Ennsner-Kananen, J. (dir.). *Handbook of educational linguistics*. New York : Routledge.

Brand, C. & Götz, S. (2013). "Fluency versus accuracy in advanced spoken learner language : a multi-method approach". In Gilquin, G. & De Cock, S. (dir.). *Errors and disfluencies in spoken corpora*. Amsterdam/Philadelphia : John Benjamins. pp. 117-137.

Chuang, F.-Y. & Nesi, H. (2006). "An analysis of formal errors in a corpus of L2 English produced by Chinese students". *Corpora*, vol. 1, n° 2. pp. 251-271.

Cook, V. (1992). "Evidence for multicompetence". *Language Learning*, vol. 42. pp. 557-591.

Conseil de l'Europe (2000). *Un cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Disponible en ligne. http://www.coe.int/t/dg4/linguistic/Source/Framework_FR.pdf

Corder, S. P. (1973). "The elicitation of interlanguage". *Errata : papers in error analysis*. Lund : CWK Gleerup.

Dagneaux, E., Denness, S. & Granger, S. (1998). "Computer-aided error analysis". *System*, vol. 26. pp. 163-174.

Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J. & Thewissen, J. (2008). *Error tagging manual*, version 1.3. Centre for English Corpus Linguistics. Louvain-la-Neuve : Université Catholique de Louvain.

Davies, A. (1995). "Proficiency or the native speaker : what are we trying to achieve in ELT ?". In Cook, G. & Seidlhofer, B. (dir.). *Principle and practice in applied linguistics : studies in honour of H.G. Widdowson*. Oxford : Oxford University Press. pp. 145-157.

Davies, A. (2003). *The native speaker : myth and reality*. Clevedon : Multilingual Matters.

De Cock, S. (2002). "Pragmatic prefabs in learners' dictionaries". In Braasch, A. & Povlsen, C. (dir.). *Proceedings of the "Tenth EURALEX international congress, EURALEX 2002"*. Copenhagen : Center for Sprogteknologi. pp. 471-781 (vol II).

De Cock, S. (2004). "Preferred sequences of words in NS and NNS speech". *Belgian Journal of English Language and Literatures (BELL), New Series*, vol. 2. pp. 225-246.

De Cock, S., Gilquin, G., Granger, S., Lefer, M.-A., Paquot, M. & Ricketts, S. (2007). "Improve your writing skills". In Rundell, M. (dir.). *Macmillan English dictionary for advanced learners*, second edition. Oxford : Macmillan Education. pp. IW1-IW50.

Dewaele, J.-M. (2007). "Diachronic and/or synchronic variation ? The acquisition of sociolinguistic competence in L2 French". In Ayoun, D. (dir.). *French applied linguistics*. Amsterdam : John Benjamins. pp. 208-36.

Dewaele, J.-M. & Mougeon, R. (dir.) (2002). "Acquisition de la variation par les apprenants du français langue seconde". *AILE*, vol. 17.

Dubois, M., Kamber, A. & Skupien Dekens, C. (2014). "A quantitative and qualitative analysis of French L2 students' spelling problems : the case of adjective agreement". In Tyne, H., André, V., Benzitoun, C., Boulton, A. & Greub, Y. (dir.). *French through corpora : ecological and data-driven perspectives in French language studies*. Newcastle : Cambridge Scholars.

Fillmore, C. (1992). "'Corpus linguistics' or 'computer-aided armchair linguistics'". In Svartvik, J. (dir.). *Directions in corpus linguistics*. Berlin : Mouton de Gruyter. pp. 35-60.

Gabrielatos, C. (2005). "Corpora and language teaching : just a fling or wedding bells ?". *TESL-EJ*, vol. 8, n° 4. n.p. Disponible en ligne. <http://tesl-ej.org/ej32/a1.html>

Gilquin, G. (2000/2001). "The integrated contrastive model : spicing up your data". *Languages in Contrast*, vol. 3. pp. 95-123.

Gilquin, G., De Cock, S. & Granger, S. (dir.) (2010). *The Louvain international database of spoken English interlanguage : handbook and CD-ROM*. Louvain-la-Neuve : Presses Universitaires de Louvain.

Gilquin, G., Granger, S. & Paquot, M. (2007). "Learner corpora : the missing link in EAP pedagogy". In Thompson, P. (dir.). *Corpus-based EAP pedagogy. Journal of English for Academic Purposes*, vol. 6, n° 4. pp. 319-335.

Gilquin, G. & Paquot, M. (2008). "Too chatty : learner academic writing and register variation". *English Text Construction*, vol. 1, n° 1. pp. 41-61.

Granger, S. (1996). "From CA to CIA and back : an integrated approach to computerized bilingual and learner corpora". In Aijmer, K., Altenberg, B. & Johansson, M. (dir.). *Languages in contrast : text-based cross-linguistic studies*. Lund : Lund University Press. pp. 37-51.

Granger, S. (2002). "A bird's eye view of learner corpus research". In Granger, S., Hung, J. & Petch-Tyson, S. (dir.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam : John Benjamins. pp. 3-33.

Granger, S. (2004). "Computer learner corpus research : current status and future prospects". In Connor, U. & Upton, T. (dir.). *Applied corpus linguistics : a multidimensional perspective*. Amsterdam/Atlanta : Rodopi. pp. 123-145.

Granger, S. (2009). "The contribution of learner corpora to second language acquisition and foreign language teaching : a critical evaluation". In Aijmer, K. (dir.). *Corpora and language teaching*. Amsterdam/Philadelphia : John Benjamins. pp. 13-32.

Granger, S. (2012). "How to use foreign and second language learner corpora". In Mackey, A. & Gass, S.M. (dir.). *Research methods in second language acquisition : a practical guide*. Oxford : Wiley-Blackwell. pp. 7-29.

Granger, S., Gilquin, G. & Meunier, F. (dir.) (2013). *Twenty years of learner corpus research. Looking back, moving ahead. Proceedings of the "First learner corpus research conference (LCR 2011)"*. Louvain-la-Neuve : Presses Universitaires de Louvain.

Granger, S. & Tribble, C. (1998). "Learner corpus data in the foreign language classroom : form-focused instruction and data-driven learning". In Granger, S. (dir.). *Learner English on computer*. London/New York : Addison Wesley Longman. pp. 199-209.

Gries, S. (2010). "Corpus linguistics and theoretical linguistics : a love-hate relationship ? Not necessarily...". *International Journal of Corpus Linguistics*, vol. 15, n° 3. pp 327-343.

Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon : Multilingual Matters.

Hawkins, J. & Buttery, P. (2010). "Criterial features in learner corpora : theory and illustrations". *English Profile Journal*, vol. 1. pp. 1-23. Disponible en ligne. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7908278&fulltextType=RA&fileId=S2041536210000103>

Hawkins, J. & Buttery, P. (2009). "Using learner language from corpora to profile levels of proficiency : insights from the English Profile Programme". In Taylor, L. & Weir, C. J. (dir.), *Language testing matters: investigating the wider social and educational impact of assessment*. Cambridge : Cambridge University Press. pp. 158-175.

Hilton, H. (2008). "Connaissances, procédures et production orale en L2". *AILE*, vol. 27. pp. 63-89.

Housen, A. & Kuiken, F. (dir.) (2009). "Complexity, accuracy, and fluency (CAF) in second language acquisition research". *Applied Linguistics*, vol. 30, n° 4. pp. 461-473.

Johns, T. (1991). "Should you be persuaded : two examples of data-driven learning". In Johns, T. & King, P. (dir.). *Classroom concordancing*. *English language Research Journal*, vol. 4. pp. 1-16.

Louw, B. (2010). "'The painting is where the paint is not' : reflections on the 'bootcamp debate'". *International Journal of Corpus Linguistics*, vol. 15, n° 3. pp. 344-353.

Lynch, T. (2001). "Seeing what they meant : transcribing as a route to noticing". *ELT Journal*, vol. 55, n° 2. pp. 124-132.

Lynch, T. (2007). "Learning from the transcripts of an oral communication task". *ELT Journal*, vol. 61, n° 4. pp. 311-320.

McCarthy, M. (2008). "Accessing and interpreting corpus information in the teacher education context". *Language Teaching*, vol. 41, n° 4. pp. 563-574.

McCarthy, M. (2010). "Spoken fluency revisited". *English Profile Journal*, vol. 1, n° 1. pp. 1-15, e4.

McDonough, K. (2005). "Identifying the impact of negative feedback and learner' responses on ESL question development". *Studies in Second Language Acquisition*, vol. 27. pp. 79-103.

Mennim, P. (2012). "Learner negotiation of L2 form in transcription exercises". *ELT Journal*, vol. 66, n° 1. pp. 52-61.

Meunier, F. & Littré, D. (2012). "Tracking learners' progress : adopting a dual 'corpus cum experimental data' approach". *Modern Language Journal*, vol. 97. pp. 61-76.

Millar, N. & Lehtinen, B. (2008). "DIY local learner corpora : bridging gaps between theory and practice". *JALT CALL Journal*, vol. 4, n° 2. pp. 61-72.

Milton, J. (2010). "The development of vocabulary breadth across the CEFR levels : a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe". In Bartning, I. Martin, M. & Vedder, I. (dir.). *Communicative proficiency and linguistic development : intersections between SLA and language testing research*. pp. 210-231. Disponible en ligne. <http://eurosla.org/monographs/EM01/EM01tot.pdf>

Mitchell, R., Myles, F. & Marsden, E. (2013). *Second language learning theories*, third edition. Londres : Routledge.

Mukherjee, J. (2009). "The grammar of conversation in advanced spoken learner English : learner corpus data and language-pedagogical implications". In Aijmer, K. (dir.). *Corpora and language teaching*. Amsterdam/Philadelphia : John Benjamins. pp. 203-230.

Mukherjee, J. & Rohrbach J.-M. (2006). "Rethinking applied corpus linguistics from a language-pedagogical perspective : new departures in learner corpus research". In Kettemann, B. & Marko, G. (dir.). *Planing, gluing and painting corpora : inside the applied corpus linguist's workshop*. Frankfurt : Peter Lang. pp. 205-232.

Myles, F. (2005). "Interlanguage corpora and second language acquisition research". *Second Language Research*, vol. 21. pp. 373-391.

Nesselhauf, N. (2004). "Learner corpora and their potential in language teaching". In Sinclair, J. (dir.). *How to use corpora in language teaching*. Amsterdam/Philadelphia : John Benjamins. pp. 125-152.

Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English : the case of more advanced learners. In Meunier, F. and Granger, S. (dir.). *Phraseology in foreign language learning and teaching*. Amsterdam/ Philadelphia : John Benjamins, 67-83.

Osborne, J. (2011). Oral learner corpora and the assessment of fluency in the common European framework. In Frankenberg-Garcia, A., Flowerdew, L. & Aston, G. (dir.). *New trends in corpora and language learning*. London/New York : Continuum. pp. 181-197.

Osborne, J. (2013). "Fluency, complexity and informativeness in native and non-native speech". In Gilquin, G. & De Cock, S. (dir.). *Errors and disfluencies in spoken corpora*. Amsterdam/Philadelphia : John Benjamins. pp. 139-161.

Paquot, M. (2008). "Exemplification in learner writing : a cross-linguistic perspective". In Meunier, F. & Granger, S. (dir.). *Phraseology in foreign language learning and teaching*. Amsterdam/ Philadelphia : John Benjamins. pp. 101-120.

Paquot, M. (2010). *Academic vocabulary in learner writing : from extraction to analysis*. New York : Continuum.

Paquot, M. (2012). "The LEAD dictionary-cum-writing aid : an integrated dictionary and corpus tool". In Granger, S. & Paquot, M. (dir.). *Electronic lexicography*. Oxford : Oxford University Press. pp. 163-187.

Paquot, M., Hasselgård, H. & Oksefjell Ebeling, S. (2013). "Writer/reader visibility in learner writing across genres : a comparison of the French and Norwegian components of the ICLE and VESPA learner corpora". In Granger, S., Gilquin, G. & Meunier, F. (dir.). *Twenty years of learner corpus research : looking back, moving ahead. Proceedings of the "First learner corpus research conference (LCR 2011)"*. Louvain-la-Neuve : Presses Universitaires de Louvain. pp. 377-387.

Paternostro, R. (à paraître). "L'éveil à la variation phonétique en didactique du français langue étrangère : enjeux et outils". *LIDIL*, vol. 50.

Pérez-Paredes, P. (2003). "Integrating networked learner oral corpora into foreign language instruction". In Granger, S. & Petch-Tyson, S. (dir.). *Extending the scope of corpus-based research : new applications, new challenges*. Amsterdam/Atlanta : Rodopi. pp. 249-261.

Römer, U. (2006). "Pedagogical applications of corpora : some reflections on the current scope and a wish list for future developments". *Zeitschrift für Anglistik und Amerikanistik*, vol. 54, n° 2. pp. 121-134.

Römer, U. (2009). "Corpus research and practice : what help do teachers need and what can we offer ?". In Aijmer, K. (dir.). *Corpora and language teaching*. Amsterdam : John Benjamins. pp. 83-98.

Rundell, M. (dir.). (2007). *Macmillan English dictionary for advanced learners*, second edition. Oxford : Macmillan Education.

Schmidt, R. (1990) "The role of consciousness in second language learning". *Applied Linguistics*, vol. 11, n° 2. pp. 129-158.

Seidlhofer, B. (2002). "Pedagogy and local learner corpora : working with learner-driven data". In Granger, S., Hung, J. & Petch-Tyson, S. (dir.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam : John Benjamins. pp. 213-234.

Stillwell, C., Curabba, B., Kamsin, A., Kidd, A., Kim, E., Stone, P. & Wyle, C. (2010). "Students transcribing tasks : noticing fluency, accuracy, and complexity". *ELT Journal*, vol. 64, n° 4. pp. 445-455.

Teubert, W. (2010a). "Our brave new world ?". *International Journal of Corpus Linguistics*, vol. 15, n° 3. pp. 354-358.

Teubert, W. (2010b). "My brave old world ?". *International Journal of Corpus Linguistics*, vol. 15, n° 3. pp. 395-399.

Thewissen, J. (2012a). *Accuracy across proficiency levels : insights from an error-tagged EFL learner corpus*. Unpublished PhD Thesis. Université Catholique de Louvain.

Thewissen, J. (2012b). "Capturing L2 accuracy developmental patterns : insights from an error-tagged EFL learner corpus". *Modern Language Journal*, vol. 97. pp. 77-101.

Tono, Y. & Negishi, M. (2012). "The CEFR-J : adapting the CEFR for English Language Teaching in Japan". *Framework & Language Portfolio (FLP) SIG Newsletter*, vol. 8, septembre 2012. Disponible en ligne. <https://sites.google.com/site/flpsig/publications>

Towell, R., Hawkins, R. & Bazergui, N. (1996). "The development of fluency in advanced learners of French". *Applied Linguistics*, vol. 17, n° 1. pp. 84-119.

Tyne, H. (2009a). "Corpus oraux par et pour l'apprenant". *Mélanges Crapel*, vol. 31. pp. 91-111. Disponible en ligne. <http://www.atilf.fr/spip.php?rubrique584>

Tyne, H. (2009b). "Style in L2 : the icing on the cake ?" In Labeau, E. & Myles, F. (dir.). *The advanced learner variety : the case of French*. Berne : Peter Lang. pp. 243-268.

Tyne, H. (à paraître). "Traitement des expressions idiomatiques en français L2 : questions d'influence". *Mélanges Crapel*.

Worlock Pope, C. (2010). "The bootcamp discourse and beyond". *International Journal of Corpus Linguistics*, vol. 15, n° 3. pp. 323-325.

À propos des auteurs

Sylvie De Cock est professeure de linguistique anglaise, langue anglaise et anglais des affaires à l'Université catholique de Louvain (Belgique). Elle est également professeure invitée à l'Université Saint-Louis (Belgique). Elle poursuit ses recherches sur les corpus d'apprenants et l'anglais des affaires au sein du Centre for English Linguistics (Université catholique de Louvain) et est notamment impliquée dans le projet *Louvain International Database of Spoken English Interlanguage (Lindsei)*. Elle s'intéresse plus particulièrement à la phraséologie et à la lexicographie pédagogique. Elle est membre de la *Learner Corpus Association*.

Courriel : sylvie.decock@uclouvain.be

Adresse : Centre for English Corpus Linguistics, Collège Erasme, Faculté de philosophie, arts et lettres, Université catholique de Louvain, 1 Place Blaise Pascal, B-1348 Louvain-la-Neuve, Belgique.

Henry Tyne est enseignant-chercheur à l'Université de Perpignan Via Domitia où il est aussi membre de l'équipe de recherche VECT. À l'Université de Perpignan, il assure des enseignements dans le Master professionnel « Didactique des langues, français langue étrangère et seconde ». Ses travaux de recherche portent notamment sur la notion de variation en langue, sur les implications en termes de description et en termes de d'acquisition ou d'apprentissage. Il est co-auteur avec Alex Boulton d'un ouvrage paru chez Didier en 2014 (*Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*). Il est membre du bureau de l'*Association for French Language Studies (AFLS)*.

Courriel : henry.tyne@univ-perp.fr

Adresse : Université de Perpignan Via Domitia, Département de Lettres classiques, modernes et de FLE, 52 avenue Paul Alduy, 66000 Perpignan, France