



**HAL**  
open science

## Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues.

Alex Boulton

► **To cite this version:**

Alex Boulton. Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues..  
Texte et corpus, 2007, 3, pp.37-46. hal-00326988

**HAL Id: hal-00326988**

**<https://hal.science/hal-00326988>**

Submitted on 6 Oct 2008

**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.

# ESPRIT DE CORPUS : PROMOUVOIR L'EXPLOITATION DE CORPUS EN APPRENTISSAGE DES LANGUES\*

Alex Boulton  
CRAPEL – ATILF/CNRS, Nancy-Université

## INTRODUCTION

L'exploitation directe des corpus par les enseignants et surtout par les apprenants n'est pas aussi avancée que l'aurait laissé supposer de nombreux articles parus dans les années 90. Le retard est encore plus important en France, en partie dû à des raisons culturelles que nous exposerons brièvement. L'un des facteurs-clés semble être le manque d'études évaluatives de l'apport des corpus dans les processus d'apprentissage, un manque constaté à maintes reprises. Afin de voir plus clair, cet article examine l'état du terrain actuel en analysant 39 études pour en tirer les grandes lignes, les objectifs et les méthodologies employées.

## 1 LES CORPUS ET L'APPRENTISSAGE EN FRANCE

L'Agence Nationale de la Recherche a récemment lancé deux appels à projet pour des "corpus et outils de la recherche en sciences humaines et sociales". En attendant, le plus grand corpus actuellement disponible au public est Frantext avec ses quelques 200 millions de mots<sup>1</sup>. La composition même de ce corpus (principalement des textes littéraires datant du 16<sup>e</sup> au 20<sup>e</sup> siècle) est révélatrice de l'importance accordée au « bon usage » de la langue française. En effet, le manque relatif d'intérêt pour la linguistique de corpus en France peut être attribué à une méfiance envers une approche purement descriptive de l'emploi courant de la langue. Ce genre de barrière culturelle n'est donc pas à sous-estimer.

Les corpus ont également eu peu d'impact dans le domaine de l'apprentissage / enseignement des langues en France, que ce soit au niveau des syllabus ou des matériels didactiques. Il en est de même dans la salle de classe où les implications de la recherche en linguistique de corpus arrivent au compte-gouttes, loin du « torrent » prédit dans d'autres pays (Leech, 1997). En effet les corpus doivent encore faire leurs preuves dans les trois usages de base décrits par Fligelstone (1993), à savoir « *teaching about* » (enseigner la linguistique de corpus), « *exploiting to teach* » (exploiter les corpus pour enseigner les langues) et « *teaching to exploit* » (former les apprenants à exploiter les corpus pour apprendre une langue). Le peu d'intérêt dont ces trois domaines font l'objet, relève ici aussi plus d'une réaction instinctive que d'une décision réfléchie.

En ce qui concerne plus particulièrement l'approche que Johns (1991a) appelle le « *data-driven learning* » ou DDL (apprentissage à partir des données, c'est-à-dire l'exploitation directe des corpus par l'apprenant), les écarts par rapport aux pratiques courantes peuvent

---

\* Pour faire référence à cet article : Boulton Alex, « Esprit de corpus : Promouvoir l'exploitation de corpus en apprentissage des langues », revue électronique *Texte et corpus*, n°3 / août 2008, Actes des Journées de la linguistique de Corpus 2007, p. 37-46

(disponible sur [http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES\\_JLC07\\_boulton.pdf](http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_boulton.pdf))

<sup>1</sup> <http://www.frantext.fr/noncateg.htm> page consultée en décembre 2007.

sembler trop importants et manquer de « sérieux ». Ceci est d'autant plus vrai dans une culture d'enseignement traditionnelle telle que la décrit Willis (2003 : 167) :

Ton professeur est ton guide et mentor, il te montrera ce qu'il faut apprendre et comment le faire. Ecoute ton professeur et fais ce qu'il te dit.

Une caricature, bien évidemment, et en pleine contradiction avec tant de concepts actuellement promus dans la didactique des langues : les styles, stratégies et profils d'apprentissage différents, l'individualisation et l'apprentissage différencié, la motivation intrinsèque, l'apprentissage inductif, l'autonomisation, la compétence communicative, et ainsi de suite. En particulier dans ce modèle traditionnel, l'enseignant y est perçu comme étant le point central, une représentation carrément opposée à celle du DDL où l'on essaie « d'éliminer l'intermédiaire [c'est-à-dire l'enseignant] dans la mesure du possible pour donner à l'apprenant un accès direct aux données » (Johns, 1991b : 30). Pourtant, il s'agit d'une représentation qui perdure, avec des conséquences prévisibles : à la sortie du lycée, plus de 90% des élèves en France ont étudié l'anglais pendant sept ans pour atteindre un niveau de compétence médiocre selon les rapports européens<sup>2</sup>.

Certes, le DDL nécessite une remise en question des rôles traditionnels : les enseignants peuvent craindre une perte de contrôle, de pouvoir, de statut d'expert ; les apprenants peuvent refuser la responsabilité qu'ils auront pour leur propre apprentissage. Tous peuvent rejeter la nature « floue » du langage inhérent au travail sur corpus et préférer la certitude et la sécurité (même illusoire) qui accompagnent une approche traditionnelle à base de règles. Certains peuvent prétendre qu'ils manquent de ressources, surtout en termes d'ordinateurs et de logiciels, ou les moyens financiers pour les obtenir. Même lorsqu'elles existent, les ressources TIC font souvent peur et ne font que rebuter les utilisateurs ; un minimum de formation serait indispensable pour une utilisation productive du DDL, mais toute activité pour « apprendre à apprendre » peut être perçue comme une perte de temps qui ne sert qu'à retarder l'apprentissage de la langue elle-même.

Si de telles objections ne se limitent pas uniquement à la France, les cultures d'apprentissage / enseignement sont telles que l'on peut se demander si le DDL ne serait pas complètement inapplicable. Trois éléments de réponse nous paraissent importants. Tout d'abord, il semble évident que toute approche nouvelle a peu de chances de réussir si elle est imposée telle quelle de façon idéologique et rigide ; une sensibilité aux cultures locales est indispensable (Seidlhofer, 2002 : 220). Il convient donc d'expérimenter le DDL dans le contexte spécifique français. Ensuite, la philosophie en général en didactique veut que les apprenants (et les enseignants) soient informés et formés afin de prendre des décisions concernant leur propre parcours en toute connaissance de cause : styles, stratégies, techniques, méthodes... Une initiation progressive peut inciter les apprenants et les enseignants à tâter le terrain, à ne pas rejeter le DDL sans l'avoir essayé.

Enfin, il ne suffit pas de proclamer les avantages théoriques ou supposés du DDL : il nous incombe en tant que chercheurs de démontrer qu'il rapporte effectivement des résultats bénéfiques dans une variété de situations ; que pour les apprenants comme pour les enseignants, un investissement minime pour maîtriser les outils et techniques peut rapporter un rendement important au processus d'apprentissage.

---

<sup>2</sup> European Network of Policy Makers for the Evaluation of Education Systems. (2004). *The Assessment of Pupils' Skills in English in Eight European Countries 2002. A European Project.* ([www.cisad.adc.education.fr/revu/pdf/assessmentofenglish.pdf](http://www.cisad.adc.education.fr/revu/pdf/assessmentofenglish.pdf))

## 2 LES ÉTUDES ÉVALUATIVES

Pour ces raisons, le présent article passe en revue les recherches évaluatives à ce jour. Cresswell (2007) ne reconnaît que trois études de ce type (Stevens, 1991 ; Cobb, 1997b) ; Chambers (2007) en compare une douzaine ; dans cet article nous en analysons 39 qui ont été publiées depuis la parution de *Classroom concordancing* de Johns et King en 1991 (Tableau 1). La différence d'échelle s'explique du fait que nous comptabilisons comme étude « empirique » toute étude basée sur des observations ou des expérimentations, et non pas uniquement sur le raisonnement théorique ou bien sur l'appréciation subjective du chercheur. Cette approche devrait nous permettre de voir plus loin et de construire une meilleure appréciation de l'ensemble du domaine. Les conclusions générales sont discutées à la fin, mais dans un premier temps notre préoccupation principale sera une analyse des objectifs et des méthodologies de recherche. Cette analyse est nécessairement extrêmement succincte ; le lecteur est renvoyé aux études originales listées dans la bibliographie.

	PAYS	L1	L2	CONTEXTE	SPÉCIALITÉ	NIVEAU	SUJETS
<b>Ilse, 1991</b>	Allemagne	allemand	anglais	post Abitur	finance	inter ?	26
<b>Mparutsa et al., 1991</b>	Zimbabwe	shona ?	anglais	ES	divers	inter +	variable
<b>Stevens, 1991</b>	Oman	arabe	anglais	ES 1	sciences du langage	inter ?	54
<b>Aston, 1996</b>	Italie	italien	anglais	ES	traduction	avancé	?
<b>Aston, 1997</b>	Italie	italien	anglais	ES	traduction	avancé	?
<b>Cobb, 1997a ; 1997b ; 1999</b>	Oman	arabe	anglais	ES 1	commerce	faible	11 (centaines)
<b>Johns, 1997</b>	Grande-Bretagne	malais	anglais	ES, enseignants	L2	avancé	?
<b>Maia, 1997</b>	Portugal	portugais	anglais	ES	traduction	avancé	12-16
<b>Bowker, 1998</b>	Irlande	anglais	français	ES 3	sciences du langage	avancé	14
<b>Turnbull &amp; Burston, 1998</b>	Australie	mixte	anglais	ES 4+	japonais, gestion	avancé	2
<b>Bowker, 1999</b>	Irlande	anglais	français	ES 4	traduction	avancé	5
<b>Bernardini, 2000</b>	Italie	italien	anglais	ES 3	traduction	avancé	6
<b>Seidlhofer, 2000 ; 2002</b>	Autriche	allemand	anglais	ES 1/2	L2	avancé	10
<b>Ciezielska-Ciupek, 2001</b>	Pologne	polonais	anglais	lycée	scolaire	faible	33
<b>Cobb et al., 2001</b>	Canada	anglais	français	(adulte)	?	inter	1
<b>Kennedy &amp; Miceli, 2001 ; 2002</b>	Australie	anglais	italien	ES 2	5 L2, 5 autres	inter	17
<b>St John, 2001</b>	Grande-Bretagne	anglais	allemand	ES 2	sciences du langage	débutant	1
<b>Todd, 2001</b>	Thaïlande	thaïlandais	anglais	ES 4+	sciences, ingénieur	inter -	23
<b>Bernardini, 2002</b>	Italie	italien	anglais	ES 4	traduction	avancé	10
<b>Curado Fuentes, 2002</b>	Espagne	espagnol	anglais	ES	commerce	inter ?	50
<b>Fan &amp; Xun-Feng, 2002</b>	Hong Kong	chinois	anglais	ES 3	traduction	avancé	21
<b>Hadley, 2002</b>	Japon	japonais	anglais	ES 1	économie	faux débutant	25
<b>Cheng et al., 2003</b>	Hong Kong	chinois	anglais	ES 2	L2	avancé	29
<b>Curado Fuentes, 2003</b>	Espagne	espagnol	anglais	ES 3	commerce	inter	20 (c=10)
<b>Sun &amp; Wang, 2003</b>	Taiwan	chinois	anglais	lycée	scolaire	inter ?	81 (c=40)
<b>Gaskell &amp; Cobb, 2004</b>	Canada	chinois	anglais	ES	divers	inter -	20
<b>Chambers, 2005</b>	Irlande	anglais	mixte	ES 2	L2	avancé	14

<b>Frankenberg-Garcia, 2005c</b>	Portugal	portugais	anglais	ES 4	traduction	avancé	16
<b>Horst et al., 2005</b>	Canada	mixte	anglais	ES	divers	inter	75
<b>Allan, 2006</b>	Irlande	mixte	anglais	centre de langues	Divers	inter +	18 (c=5)
<b>Boulton, 2006a ; Boulton &amp; Wilhelm, 2006</b>	France	français	anglais	ES 4	L2	avancé	30
<b>Boulton, 2006b</b>	France	français	anglais	ES 1	Ingénieur	inter -	132 (c=64)
<b>Koosha &amp; Jafarpoo, 2006</b>	Iran	farsi	anglais	ES	L2	avancé	200 (c=100)
<b>Lee &amp; Swales, 2006</b>	États-Unis	chinois	anglais	ES 5+	Divers	avancé	4
<b>Boulton, 2007</b>	France	français	anglais	ES 1	Ingénieur	inter -	104 (c=51)
<b>Boulton, à paraître</b>	France	français	anglais	ES 1	Ingénieur	inter -	113
<b>Cresswell, 2007</b>	Italie	italien	anglais	ES 3	Traduction	avancé	126 (c=65)
<b>Curado Fuentes, 2007</b>	Espagne	espagnol	anglais	ES	Tourisme	inter +	20 (c=10)
<b>Lavid, 2007</b>	Espagne	espagnol	anglais	ES 4+	sciences du langage	avancé	?

*Tableau 1 : 39 études empiriques en DDL, selon la date de publication. Lorsqu'une seule étude est citée dans plusieurs articles, toutes les références sont fournies*

Légende : Pays : lieu du déroulement de l'étude. L1 : langue maternelle des participants. L2 : langue cible. Contexte : pas toujours précisé, mais par ex. ES 1 = première année d'études supérieures. Spécialité : domaine d'études principales des participants. Niveau : selon les termes utilisés dans les études mêmes, dans la mesure du possible ; par ex. inter+ = intermédiaire fort. Sujets : nombre de participants, dont c = groupe de contrôle.

## 2.1 Contextes

Les études ont été menées dans 19 pays différents, démontrant une certaine variété de contextes même si 24 des études ont été effectuées dans des pays européens : cinq en Italie ; quatre en Espagne, en Irlande et en France (les différentes études de Boulton sont les seules dans le contexte français) ; deux en Grande-Bretagne et au Portugal ; une seule en Allemagne, en Autriche, et en Pologne. Les autres pays comprennent le Canada (trois études), l'Australie, Hong-Kong et Oman (deux études), les États-Unis, l'Iran, le Japon, Taiwan, la Thaïlande et le Zimbabwe (une étude).

Dans la plupart des cas, les participants avaient comme langue maternelle la langue du pays où s'est déroulée l'étude ; cinq des études (toutes dans des pays anglophones) avaient des populations de langue maternelle mixte. En tout, 18 études proviennent de pays romanophones (dont le Canada), 15 de pays anglophones (y compris le Canada, le Zimbabwe et Hong Kong). Dans six études, les participants étaient des anglophones natifs, dont les deux menées par Bowker (1998 ; 1999) où des apprenants irlandais du français L2 ont employé des corpus comme aide en traduction vers la L1.

La langue cible (L2) était l'anglais dans 33 des 39 études, une langue européenne dans toutes les autres. Ceci s'explique en grande partie par la demande pour l'anglais L2 à l'échelle mondiale, ainsi que par la gamme d'outils, de logiciels et de corpus disponibles pour ces langues ; le fait que les publications sont souvent en anglais contribue certainement à une plus grande prise de conscience pour cette langue. Quelles que soient les raisons, il semble que le DDL soit une approche qui s'inspire largement du monde anglo-saxon.

Sur l'ensemble des 39 études, deux seules s'intéressent aux jeunes apprenants en lycée (Ciezielska-Ciupek, 2001 ; Sun & Wang, 2003) ; dans tous les autres cas il s'agit d'adultes d'un niveau d'études post-baccalauréat, le plus souvent dans un contexte d'enseignement supérieur, dont huit en master ou thèse de doctorat. Dans la moitié de ces cas (18) les apprenants avaient besoin de l'anglais sur objectifs spécifiques ou académiques, les autres

étant spécialistes de la langue cible ou en traduction de cette langue, ou bien dans trois cas des étudiants en sciences du langage.

Il est donc sans surprise que peu de recherches s'intéressent aux niveaux « faibles » ou « débutants » (quatre études en tout), mais une lecture attentive soulève des doutes quant aux définitions. Les « faux débutants » de Hadley (2002) ont certainement étudié l'anglais pendant plusieurs années dans le système éducatif au Japon avant d'arriver à l'université. De la même façon, le débutant dans l'étude de cas de St John avait étudié la langue cible pendant un an lors de sa licence en sciences du langage et était « l'un des meilleurs étudiants de sa promotion et adorait la grammaire » (St John, 2001 : 188). La quasi-totalité des recherches à ce jour vise des apprenants de niveau relativement avancé, en partie sans doute parce que les chercheurs dans des milieux universitaires y ont plus facilement accès. S'il est d'une évidence pour certains que le DDL se prête mieux aux niveaux avancés, les quelques tentatives avec des niveaux plus faibles (par ex. Boulton 2006b ; 2007 ; à paraître) sont prometteuses et indiquent que la possibilité n'est pas à exclure.

## 2.2 Outils

Les recherches sont basées sur une grande variété de corpus, de 2.000 occurrences (Turnbull & Burston, 1998) à plus de cent millions. La plupart était monolingue, mais certains ont utilisé des corpus parallèles ou comparables à des fins de traduction. Les corpus existants comprennent le British National Corpus et la Bank of English, ainsi que les corpus Brown, ICE et MICASE ; le web a été utilisé comme corpus en tant que tel dans deux études seulement (Todd, 2001 ; Boulton, 2006b). Plus souvent les chercheurs ont créé leurs propres corpus en fonction des besoins des apprenants, récoltant les données à partir d'internet ou de CD-ROMs (par ex. des journaux), ou tout simplement en scannant des ouvrages papier (par ex. des manuels utilisés dans les autres cours). Seidlhofer (2000 ; 2002) et Lee et Swales (2006) ont utilisé les productions de leurs apprenants dans leurs corpus. Parfois les étudiants pouvaient choisir les textes à inclure (par ex. Horst *et al.*, 2005), ou bien créer des corpus eux-mêmes (par ex. Boulton & Wilhelm, 2006 ; Maia, 1997), généralement lorsqu'ils participaient à un projet en linguistique de corpus.

Il était parfois question de concordances imprimées, mais dans la plupart des cas, les apprenants pouvaient interroger le corpus directement, soit pendant les cours soit en accès libre dans une salle informatique. Les interfaces dédiées (Sara, VIEW,...) étaient utilisées avec certains grands corpus, mais WordSmith Tools reste le logiciel de préférence, utilisé dans 12 études, suivi par MicroConcord dans cinq autres ; un petit nombre se servait d'outils spécialement conçus (par ex. Cobb, 1997a ; 1997b ; 1999a).

## 2.3 Méthodologies

Ces études sont tout aussi variées en ce qui concerne leurs objectifs. Ces derniers sont souvent multiples, ce qui complexifie toute tentative de synthèse. Toutefois, on peut les regrouper dans trois catégories principales. Un premier groupe analyse les *attitudes* des participants envers l'exploitation des corpus, un axe essentiel car le DDL implique des changements importants de rôles et de pratiques et la motivation semble être un facteur-clé de la réussite. Un deuxième recense les *pratiques* utilisées : que font les apprenants et est-ce qu'ils le font bien, c'est-à-dire, est-ce que les méthodologies de la linguistique de corpus sont suffisamment accessibles à des non-spécialistes ? Cet axe est également essentiel, mais on peut remarquer qu'aucun de ces deux premiers types d'études n'évalue directement l'*efficacité* du DDL.

Tel est l'objectif du dernier groupe : est-ce que les apprenants peuvent tirer profit d'une analyse de corpus ? Ici on peut détecter deux grands courants. D'abord, plusieurs études examinent l'emploi des corpus et les méthodologies afférentes comme outil de *référence* pour la production écrite ou pour la traduction, ou bien pour la correction de ses propres erreurs. Il

se peut en effet que l'intérêt principal des corpus en L2 soit de compléter les outils de référence traditionnels : dictionnaires, grammaires, etc. Seule une poignée d'études se fixe l'objectif d'évaluer directement l'efficacité de l'exploitation des corpus à des fins *d'apprentissage* – le DDL en tant que tel. Celles-ci visent le plus souvent différents aspects de l'apprentissage du lexique qui, à en juger par l'ensemble de ces recherches, semble bien se prêter à cette approche.

On peut regrouper ces études autrement : d'un côté celles qui prennent comme point de départ un cours existant qu'elles cherchent ensuite à évaluer ; de l'autre celles qui commencent par une question de recherche qu'elles cherchent à tester. La distinction se reflète dans les instruments de recherches utilisés : à part une évaluation informelle, les outils les plus fréquemment utilisés comprennent les discussions en classe (onze études), des questionnaires (dix) et des entretiens (huit), souvent en combinaison. Quelques-unes se sont servies d'un tracking automatique, de(s) cahiers d'apprenants ou d'observations dans la salle de classe, tandis que d'autres ont imposé des tâches très contrôlées afin de permettre une analyse des activités proposées. Dans onze cas, généralement avec des étudiants spécialistes de la langue, l'analyse se porte sur des rapports (oraux ou écrits) fournis par les participants sur un point de langue qu'ils ont étudié en utilisant des techniques de corpus. Cette diversité garantit des découvertes très riches, mais rendent les comparaisons complexes et les conclusions générales sont difficiles à établir.

## 2.4 Quelques études quantitatives

Parmi les recherches examinées ici, huit études comptent dix apprenants au plus, dont deux études de cas avec un seul apprenant. A l'autre bout de l'échelle, cinq comptent plus de 100 participants, la moyenne étant de 38.97, une taille suffisante pour des conclusions statistiques. Par ailleurs, on peut noter dans le Tableau 1 que les populations sont généralement plus importantes dans les études plus récentes, facilitant les évaluations quantitatives. Celles-ci sont généralement assez rares : seules huit études comptent un groupe de contrôle (bien que quatre autres aient été conçues de façon que les mêmes apprenants servent à la fois de groupe de contrôle et de groupe expérimental pour différentes tâches) ; neuf autres comparent des résultats avant et après la condition expérimentale. En tout, seize ne présentent aucun résultat quantitatif, et neuf fournissent quelques chiffres bruts ou des pourcentages sans analyse statistique proprement dite.

Parmi les études quantitatives, il n'y en a que cinq qui tentent d'évaluer l'apprentissage même de la langue (*cf.* la section précédente) : Cobb (1997a ; 1997b ; 1999a ; 1999b), Horst & Cobb (2001), Allan (2006), Koosha & Jafarpour (2006), Sun & Wang (2003). On remarque en particulier que quatre de ces études évaluent des programmes de formation expérimentale en cours et fournissent ainsi des données longitudinales très riches.

## 3 DISCUSSION

Si l'on ose une synthèse très générale de programmes de recherches aussi variés, on ne peut que conclure que dans l'ensemble, les résultats sont très prometteurs : les attitudes sont majoritairement positives, les apprenants maîtrisent vite les outils et techniques proposés, et les corpus semblent être bénéfiques à des fins de référence autant que pour l'apprentissage même. Bien entendu, la plupart des chercheurs qui embarquent sur ce type d'étude évaluative sont des partisans du DDL ; toutefois, les conclusions des articles examinés ici sont formulées avec prudence, comme en témoigne la conclusion générale de Cresswell (2007 : 280) :

Overall, given that the students were advanced and the items already partially known it is possible to conclude, albeit tentatively, that, given language items at the right level, DDL has an observable (though slight) positive effect on actual use.

Les chercheurs soulignent souvent les limitations de l'approche, et on peut en conclure que le DDL ne conviendrait pas à tous les profils d'apprenants à tous les niveaux dans tous les contextes d'apprentissage et pour tout aspect d'une L2. Mais nous manquons cruellement de recherches pour déterminer ces limites – non seulement pour démontrer que le DDL peut apporter quelque chose à l'apprentissage, mais dans quelles conditions.

Rares sont ceux qui réclameraient une « révolution » à base de corpus dans la salle de classe, mais on peut facilement imaginer l'inclusion d'activités inspirées par le DDL dans des manuels (Tribble & Jones, 1997). De même, les maisons d'édition pourraient concevoir des corpus « pédagogiques » (Braun, 2005) créés spécialement pour des apprenants avec des interfaces didactiquement appropriées, ou encore des sites qui complètent des manuels pour un travail autonome ou pour permettre à l'enseignant de créer des activités pertinentes. Plusieurs centaines d'articles prônent les mérites de ce type d'approche et présentent des arguments sérieux, mais vu le peu d'impact à ce jour, il semble que ces arguments soient insuffisants pour convaincre un public plus large. Peut-être que des preuves plus concrètes pourraient susciter l'intérêt auprès des décideurs et des maisons d'édition et permettre au DDL de sortir de ses confins universitaires.

La quarantaine d'études examinées ici peut sembler une quantité conséquente, mais elle représente le travail d'un petit groupe de chercheurs produisant en moyenne deux études par an depuis la publication de l'ouvrage fondateur de Johns & King en 1991. En particulier, il manque des études qui examinent les apports en termes d'apprentissage, qui testent les limites du DDL. Surtout, comme le dit Chambers (2007 : 5) dans sa propre revue du domaine, « on est en droit de se demander pourquoi il n'y a pas plus d'études quantitatives ». Certes, on manque de « méthodologie clairement définie » (Chambers & O'Sullivan, 2004 : 161) qui pourrait informer les études à venir, et il se peut que la perspective d'effectuer ce genre de recherche est plutôt décourageante au regard de la quantité de variables à contrôler. Mais il n'est pas évident que les obstacles soient plus intimidants que dans d'autres domaines de recherches en didactique des langues, et les quelques études expérimentales (par ex. Boulton, 2006b ; 2007 ; à paraître ; Sun & Wang, 2003) démontrent qu'il est tout à fait possible d'isoler des questions précises.

La plupart des recherches à ce jour, nous l'avons vu, se concentrent sur des apprenants relativement sophistiqués dans un contexte d'enseignement supérieur bien équipé et à un niveau avancé de la L2. Il ne s'agit pas uniquement de contraintes pratiques, car on lit souvent que le DDL se prête mieux à ce type d'apprenant dans ce type de contexte. Mais cette position reste intuitive : très peu de recherches dépassent ce contexte limité. Qu'en est-il d'apprenants plus jeunes dans un contexte scolaire ? Quelles sont les possibilités du DDL dans une salle de classe non équipée d'ordinateurs, par exemple avec des concordances préparées par l'enseignant ou intégrées directement dans des manuels ? Quels sont les profils d'apprenants, les styles d'apprentissage, où le DDL est le plus bénéfique ? Y a-t-il des langues ou des éléments linguistiques pour lesquels le DDL est particulièrement approprié ?

## 4 CONCLUSION

Il existe de très bons arguments pour promouvoir différents aspects du DDL, mais il est pour le moins ironique qu'une approche qui s'inspire de la linguistique de corpus – une science « empirique » par excellence – ne se mette pas à l'épreuve sur le terrain de façon scientifique. Le manque d'études concrètes se fait cruellement ressentir ; sans elles, tout ce que l'on peut espérer sont des affirmations du type « il me semble que » ou « à notre avis », des affirmations qui sont effectivement très fréquentes dans les publications en DDL.

Les recherches sur le terrain ne sont pas plus difficiles à effectuer que dans d'autres domaines de la didactique des langues, et les résultats peuvent être utiles même avec des moyens modestes. En effet, il semble peu probable qu'une seule étude puisse « prouver »

grand-chose ; ce qui manque essentiellement, c'est une multiplicité de projets de recherches sur le terrain dans des conditions variées, chacune avec ses propres questions. Seulement alors, le poids des résultats pourra montrer si le DDL peut tenir ses promesses pour différents apprenants sur différents points de langue dans différentes cultures et situations.

Bon nombre de publications réclament plus de recherches « empiriques ». Même si les arguments théoriques sont attrayants, de toute évidence ils ne suffisent pas pour convaincre les décideurs et pour pénétrer les courants dominants de l'enseignement / apprentissage des langues. Si un fondement purement théorique est à l'origine de différentes approches humanistes, il ne suffit plus de nos jours, comme le détaille Nunan (2007: 9) dans sa revue des pratiques humanistes des années 1980 :

The 1970s were punctuated with the so-called designer methods movement – Suggestopedia, the Silent Way, Community Language Learning, and so on. Most of these methods were 'data free', drawing sustenance from rhetoric rather than empirical support. While these designer methods continued to flourish into the 1980s, teachers, perhaps influenced by the growth of research into instructed second language acquisition, were less inclined to embrace pedagogical proposals without some kind of evidence.

Sans preuves à l'appui, le DDL risque de connaître le même sort que ces autres approches, de disparaître et ne représenter qu'une curiosité dans l'histoire de la didactique des langues.

## 5 RÉFÉRENCES

- Allan R. (2006). « Data-driven learning and vocabulary : investigating the use of concordances with advanced learners of English », *Trinity college Dublin centre for language and communication studies : occasional paper 66*.
- Aston G. (1996). « The British National Corpus as a language learner resource », in : Botley S. Glass J., McEnery A. & Wilson A. (dir.) *Proceedings of TALC 1996. UCREL technical papers 9*, p. 178-191.
- Aston G. (1997). « Involving learners in developing learning methods : exploiting text corpora in self-access », in : Benson P. & Voller P. (dir.) *Autonomy and independence in language learning*, London : Longman, p. 204-263.
- Bernardini S. (2000). « Systematising serendipity : proposals for concordancing large corpora with language learners », in : Burnard L. & McEnery T. (dir.) *Rethinking language pedagogy from a corpus perspective*, Frankfurt : Peter Lang, p. 225-234.
- Bernardini S. (2002). « Exploring new directions for discovery learning », in : Kettemann B. & Marko G. (dir.) *Teaching and learning by doing corpus analysis*, Amsterdam : Rodopi, p. 165-182.
- Boulton A. (2006a). « Bringing corpora to the masses : free and easy tools for language learning », in : *7<sup>th</sup> Teaching and language corpora conference*, Paris : BNF / Université Paris 7, 1-4 juillet.
- Boulton A. (2006b). « Tricky to teach, easier to learn : empirical evidence for corpus use in the language classroom », *American association of applied corpus linguistics*, Flagstaff AZ : University of Northern Arizona, 20-22 octobre.
- Boulton A. (2007). « DDL is in the details... and in the big themes », in : *4<sup>th</sup> Corpus linguistics conference*. Birmingham : University of Birmingham centre for corpus research, 27-30 juillet.
- Boulton A. (à paraître). « Looking (for) empirical evidence for DDL at lower levels », in : Lewandowska-Tomaszczyk B. (dir.) *Practical applications of language and computers*.
- Boulton A. & Wilhelm S. (2006). « Habeant corpus – they *should* have the body : tools learners have the right to use », *Asp*, n°49-50, p. 155-170.
- Bowker L. (1998). « Using specialised monolingual native-language corpora as a translation resource : a pilot study », *Meta*, Vol. 43/4, p. 631-651.
- Bowker L. (1999). 'Exploring the potential of corpora for raising language awareness in student translators' *Language awareness* 8/3-4, p. 160-173.
- Braun S. (2005). « From pedagogically relevant corpora to authentic language learning contents », *ReCALL*, Vol. 17/1, p. 47-64.
- Chambers A. (2005). « Integrating corpus consultation in language studies », *Language learning & technology*, vol. 9/2, p. 111-125.

- Chambers A. (2007). « Popularising corpus consultation by language learners and teachers », in : Hidalgo E., Quereda L. & Santana J. (dir.) *Corpora in the foreign language classroom*, Amsterdam : Rodopi, p. 3-16.
- Chambers A. & O'Sullivan I. (2004). « Corpus consultation and advanced learners », *ReCALL*, Vol. 16/1, p. 158-172.
- Cheng W., Warren M. & Xun-feng X. (2003). « The language learner as language researcher : putting corpus linguistics on the timetable », *System*, Vol. 31/2, p. 173-186.
- Ciezielska-Ciupek M. (2001). « Teaching with the internet and corpus materials : preparation of ELT materials using the internet and corpus resources », in : Lewandowska-Tomaszczyk B. (dir.) *PALC 2001 : Practical applications in language corpora. Lodz studies in language 7.*, Frankfurt : Peter Lang, p. 521-531.
- Cobb T. (1997a). *From concord to lexicon : development and test of a corpus-based lexical tutor*, Thèse de doctorat, Montréal : Concordia University.
- Cobb T. (1997b). « Is there any measurable learning from hands-on concordancing ? », *System*, Vol. 25/3, p. 301-315.
- Cobb T. (1999). « Breadth and depth of lexical acquisition with hands-on concordancing », *CALL*, Vol. 12/4, p. 345-360.
- Cobb T., Greaves C. & Horst M. (2001). « Can the rate of lexical acquisition from reading be increased ? An experiment in reading French with a suite of on-line resources », (trad.) in : Raymond P. & Cornaire C. (dir.) *Regards sur la didactique des langues secondes*, Montréal : Editions Logiques, p. 133-153.
- Cresswell A. (2007). « Getting to “know” connectors ? Evaluating data-driven learning in a writing skills course », in : Hidalgo E., Quereda L. & Santana J. (dir.) *Corpora in the foreign language classroom*, Amsterdam : Rodopi, p. 267-287.
- Curado Fuentes A. (2002). « Exploitation and assessment of a business English corpus through language learning tasks », *ICAME journal*, Vol. 26, p. 5-32.
- Curado Fuentes A. (2003). « The use of corpora and IT in a comparative evaluation approach for oral business English », *ReCALL*, Vol. 15/2, p. 189-201.
- Curado Fuentes A. (2007). « A corpus-based assessment of reading comprehension in English », in : Hidalgo E., Quereda L. & Santana J. (dir.) *Corpora in the foreign language classroom*, Amsterdam : Rodopi, p. 309-326.
- Fan M. & Xun-feng X. (2002). « An evaluation of an online bilingual corpus for the self-learning of legal English », *System*, Vol. 30/1, p. 47-63.
- Fligelstone S. (1993). « Some reflections on the question of teaching, from a corpus linguistics perspective », *ICAME journal*, Vol. 17, p. 97-109.
- Frankenberg-Garcia A. (2005). « A peek into what today's language learners as researchers actually do », *International journal of lexicography*, Vol. 18/3, p. 335-355.
- Gaskell D. & Cobb T. (2004). « Can learners use concordance feedback for writing errors ? », *System*, Vol. 32/3, p. 301-319.
- Hadley G. (2002). « Sensing the winds of change : an introduction to data-driven learning », *RELC journal*, Vol. 33/2, p. 99-124.
- Horst M., Cobb T. & Nicolae I. (2005). « Expanding academic vocabulary with an interactive on-line database », *Language learning & technology*, Vol. 9/2, p. 90-110.
- Ilse W-R. (1991). « Concordancing in vocational training », in : Johns T. & King P. (dir.) *Classroom concordancing. English language research journal*, Vol. 4, p. 103-113.
- Johns, T. (1991a). « Should you be persuaded : Two examples of data-driven learning », in : Johns T. & King P. (dir.) *Classroom concordancing. English language research journal*, Vol. 4, p. 1-16.
- Johns T. (1991b). « From printout to handout : grammar and vocabulary teaching in the context of data-driven learning », in : Johns T. & King P. (dir.) *Classroom concordancing. English language research journal*, Vol. 4, p. 27-45.
- Johns T. (1997). « Contexts : the background, development and trialling of a concordance-based CALL program », in : Wichmann A., Fligelstone S., McEnery T. & Knowles G. (dir.) *Teaching and language corpora*, Harlow : Addison Wesley Longman, p. 100-115.
- Johns T. & King P. (dir.) (1991). *Classroom concordancing. English language research journal* 4.

- Kennedy C. & Miceli T. (2001). « An evaluation of intermediate students, approaches to corpus investigation », *Language learning & technology*, Vol. 5/3, p. 77-90.
- Kennedy C. & Miceli T. (2002). « The CWIC project : developing and using a corpus for intermediate Italian students », in Kettemann B. & Marko G. (dir.) *Teaching and learning by doing corpus analysis*, Amsterdam : Rodopi, p. 183-192.
- Koosha M. & Jafarpour A. (2006). « Data-driven learning and teaching collocation of prepositions : the case of Iranian EFL adult learners », *Asian EFL journal quarterly*, Vol. 8/4, p. 192-209.
- Lavid J. (2007). « Contrastive patterns of mental transitivity in English and Spanish : a student-centred corpus-based study », in : Hidalgo E., Quereda L. & Santana J. (dir.) *Corpora in the foreign language classroom*, Amsterdam : Rodopi, p. 237-252.
- Lee D. & Swales J. (2006). « A corpus-based EAP course for NNS doctoral students : moving from available specialized corpora to self-compiled corpora », *English for specific purposes*, Vol. 25, p. 56-75.
- Leech G. (1997). « Teaching and language corpora: a convergence », in : Wichmann A., Fligelstone S., McEnery T. & Knowles G. (dir.) *Teaching and language corpora*, Harlow : Addison Wesley Longman, p. 1-23.
- Maia B. (1997). « Making corpora : a learning process », in : Aston, G. Gavioli L. & Zanetti F. (dir.) *Proceedings of corpus use and learning to translate*.  
<http://www.sslmit.unibo.it/cultpaps/paps.htm>, accessible via <http://web.archive.org/> April 2006]
- Mparutsa C., Love A. & Morrison A. (1991). « Bringing concord to the ESP classroom », in : Johns T. & King P. (dir.) *Classroom concordancing. English language research journal* 4, p. 115-134.
- Nunan D. (2007). « The nineteen-eighties », in : Brown, H. *Forty years of language teaching Language teaching* 40, p. 1-15.
- Seidlhofer B. (2000). « Operationalizing intertextuality : using learner corpora for learning », in : Burnard L. & McEnery T. (dir.) *Rethinking language pedagogy from a corpus perspective*, Frankfurt : Peter Lang, p. 207-223.
- 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, p. 213-234.
- St John E. (2001). « A case for using a parallel corpus and concordancer for beginners of a foreign language », *Language learning & technology* 5/3, p. 185-203.
- Stevens V. (1991). « Concordance-based vocabulary exercises : a viable alternative to gap-filling », in : Johns T. & King P. (dir.) *Classroom concordancing. English language research journal* 4, p. 47-61.
- Sun Y-C. & Wang L-Y. (2003). « Concordancers in the EFL classroom : cognitive approaches and collocation difficulty », *CALL*, vol. 16/1, p. 83-94.
- Todd R. (2001). « Induction from self-selected concordances and self-correction », *System*, vol. 29/1, p. 91-102.
- Tribble C. & Jones G. (1997). *Concordances in the classroom*, Houston : Athelstan.
- Turnbull J. & Burston J. (1998). « Towards independent concordance work for students : lessons from a case study », *ON-CALL*, vol. 12/2, p. 10-21.
- Willis D. (2003). *Rules, patterns and words*, Cambridge : Cambridge University Press.

## Sites Web :

- <http://cisad.adc.education.fr/reva/pdf/assessmentofenglish.pdf> page consultée décembre 2007.
- European Commission / Eurobarometer. (2006) Europeans and their Languages. [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_243\\_sum\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_sum_en.pdf) page consultée décembre 2007.
- European Commission / Europa. (2000) Languages in Europe. <http://www.europa.eu.int/comm/education/languages/lang/europeanlanguages.html> page consultée décembre 2006.