



Apprentissage sur corpus: théories méthodes, applications, perspectives

Henry Tyne

► To cite this version:

Henry Tyne (Dir.). Apprentissage sur corpus: théories méthodes, applications, perspectives. 97, 2013, Bulletin suisse de linguistique appliquée. hal-01208236

HAL Id: hal-01208236

<https://hal.science/hal-01208236>

Submitted on 2 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.

numéro 97

bulletin vals-asla

Vereinigung für angewandte
Linguistik in der Schweiz
Associaziun svizra
da linguistica applitgada
Association suisse
de linguistique appliquée
Associazione svizzera
di linguistica applicata



Impr. Rapidoffset, Le Locle

- Numéro 93, été 2011
Eva Lia Wyss, Daniel Stotz, Aleksandra Gnach,
Jean-François de Pietro & Ingrid de Saint-Georges (Hg.)
Sprachkompetenzen in Ausbildung und Beruf –
Übergänge und Transformationen
Band 1
Les compétences langagières dans la formation et dans
la vie professionnelle – Transitions et transformations
Volume 1
- Numéro 94, hiver 2011
Eva Lia Wyss, Daniel Stotz, Aleksandra Gnach,
Jean-François de Pietro & Ingrid de Saint-Georges (Hg.)
Sprachkompetenzen in Ausbildung und Beruf –
Übergänge und Transformationen
Band 2
Les compétences langagières dans la formation et dans
la vie professionnelle – Transitions et transformations
Volume 2
- Numéro 95, printemps 2012
Georges Lüdi (éd.)
Représentations, gestion et pratiques du plurilinguisme
- Numéro 96, hiver 2012
Elwys De Stefani, Anne-Danièle Gazin,
Anna Claudia Ticca (éds.)
L'espace dans l'interaction sociale

Henry Tyne (ed.)

- Data-driven approaches to language learning :
theories, methods, applications and perspectives
- Apprentissage sur corpus : théories, méthodes,
applications et perspectives



bulletin vals-asla

numéro 97

Bulletin suisse
de linguistique appliquée
été 2013

Vereinigung für angewandte
Linguistik in der Schweiz
Associaziun svizra
da linguistica applitgada
Association suisse
de linguistique appliquée
Associazione svizzera
di linguistica applicata

Vereinigung für Angewandte
Linguistik in der Schweiz

Association Suisse de
Linguistique Appliquée

Associazione Svizzera di
Linguistica Applicata

Associaziun Svizra da
Linguistica Applitgada



Le *Bulletin suisse de linguistique appliquée* est l'organe de publication de l'Association Suisse de Linguistique Appliquée (VALS/ASLA).

Comité de relecture pour ce numéro

Kate Beeching (University of the West of England, Bristol), Alex Boulton (Université de Lorraine), Thomas Michael Cobb (Université du Québec à Montréal), Sylvain Detey (Waseda University), Daniel Elmiger (Université de Genève), Emmanuelle Guerin (Université d'Orléans), Alain Kamber (Université de Neuchâtel), Isabelle Racine (Université de Genève), Carine Skupien Dekens (Université de Neuchâtel), Christian Surcouf (Université de Lausanne).

Nous remercions vivement les membres du comité de lecture pour leur contribution précieuse à la qualité de ce volume.

Publié avec le soutien financier de l'Académie suisse des sciences humaines et sociales (ASSH / SAGW), le *Bulletin suisse de linguistique appliquée* paraît deux fois par an.

Chaque article paru dans le bulletin a été soumis à un processus de "peer review" anonyme.

Abonnement personnel	Suisse CHF 30.-	Etranger CHF 35.-
Abonnement institutionnel	Suisse CHF 50.-	Etranger CHF 55.-
Prix au numéro	Suisse CHF 25.-	Etranger CHF 25.-
Pas de réduction librairie		

Rédaction Prof. Alain Kamber (Université de Neuchâtel)

Responsable comptes rendus Prof. Laurent Fillietaz (Université de Genève)
E-mail: Laurent.Fillietaz@unige.ch

Administration Clelia König (Université de Neuchâtel)

Abonnements, commandes Institut des sciences du langage et de la communication
Centre de linguistique appliquée
Rue des Beaux-arts 28, CH-2000 Neuchâtel
E-mail: bulletin.valsasla@unine.ch
CCP: 20-7427-1

© Centre de linguistique appliquée de l'Université de Neuchâtel, 2013
Tous droits réservés.

Table des matières

Data-driven approaches to language learning: theories, methods, applications and perspectives

Apprentissage sur corpus: théories, méthodes, applications et perspectives

Jean-François DE PIETRO

Une pensée en mouvement et en contact. Hommage à
Bernard Py 1-6

Henry TYNE

Corpus et apprentissage-enseignement des langues 7-15

Betsy KERR

Grammatical description and classroom application.
Theory and practice in data-driven learning 17-39

Angela CHAMBERS

Learning and teaching the subjunctive in French: the
contribution of corpus data 41-58

Corinne LANDURE

Corpus et dictionnaires: prise de conscience linguistique
chez des apprenants LANSAD 59-76

Juliette DELAHAIE & Danièle FLAMENT-BOISTRANCOURT

Corpus et enseignants de français L2 en Flandre: une
histoire réussie 77-96

Alex BOULTON & Henry TYNE

Corpus linguistics and data-driven learning: a critical
overview..... 97-118

Comptes rendus

Berruto, G. (2012).
Saggi di sociolinguistica e linguistica.
Alessandria: Edizioni dell'Orso 119-122

- Pugnière-Saavreda, F., Sitri, F. & Veinard, M. (2012).
*L'analyse du discours dans la société. Engagement du
chercheur et demande sociale.*
Paris: Honoré Champion 123-125
- Ferrari, A., Cignetti, L., De Cesare, A.-M., Lala, L.,
Mandelli, M., Ricci, C. & Roggia, C.E. (2008).
*L'interfaccia lingua-testo. Natura e funzioni
dell'articolazione informativa dell'enunciato.*
Alessandria: Edizioni dell'Orso 127-136

Une pensée en mouvement et en contact

Hommage à Bernard Py¹

Jean-François DE PIETRO

IRDP, Neuchâtel

Faubourg de l'Hôpital 43 / CP 556, CH-2000 Neuchâtel

jean-françois.depietro@irdp.ch

Lorsque nous prenons la parole, nous créons un microcosme et nous le dotons d'une structure dont nous tenons la clé dans la mesure où nous en sommes les auteurs. Mais nos paroles sont destinées à être appréhendées par autrui, c'est-à-dire à être saisies à partir d'un point de vue qui est forcément différent du nôtre.

(Bernard Py 1995: 82)

Après plusieurs années d'une maladie pernicieuse, Bernard Py nous a quittés le 1^{er} octobre 2012, à l'âge de 73 ans. Nous tenons à lui rendre hommage dans le cadre de notre *Bulletin* tant il a joué un rôle important, décisif, dans la constitution et la consolidation – scientifiques et institutionnelles – de notre discipline.

Né à La Chaux-de-Fonds en 1939, Bernard Py a étudié à l'Université de Neuchâtel. C'est peut-être sa passion pour l'Espagne et l'espagnol qui va le conduire à s'intéresser à la linguistique. D'abord professeur d'espagnol au lycée, il s'est en effet progressivement tourné vers l'étude de l'apprentissage de cette langue puis du français et des langues en général. Sa thèse et ses premiers travaux portaient sur l'interrogation en espagnol, dans une perspective encore largement influencée par les théories structuralistes et générativistes (cf. Py 1971 et 1973; Arcaini *et al.* 1979). Toutefois, Bernard perçoit rapidement les limites de ces théories, de l'analyse contrastive par exemple, et s'oriente vers des approches plus dynamiques plaçant au cœur de ses travaux l'activité du sujet et la notion d'interlangue (Py 1980).

Mais c'est au tournant des années 80 qu'il va commencer à développer sa propre conception et que naît sa collaboration, ô combien fructueuse, avec G. Lüdi et ce qui allait devenir "l'équipe Bâle-Neuchâtel". Tous les deux jeunes professeurs, ils lancent leur premier projet de recherche sur le bilinguisme et les langues en contact en centrant leur étude sur "la langue que parlent des migrants établis dans le canton suisse francophone de Neuchâtel"

¹ Merci à Laurent Gajo, Alain Kamber, Marinette Matthey, Cecilia Serra et Geneviève de Weck pour leur collaboration à la rédaction de ce texte.

(1986/2003: 5)². Ce projet les conduira à la rédaction du fameux ouvrage *Être bilingue* (1986/2003³; 1984 pour la version allemande), qui en est désormais à sa 3^e édition et reste une référence en la matière. D'une certaine manière, tout ce qui allait suivre était déjà là: l'intérêt pour le *bilinguisme*, bien sûr, mais non pas un bilinguisme idéalisé, "parfait", le bilinguisme "réel", celui que pratiquent et développent les sujets bilingues, ces "apprenants expérimentés" (Py 1997); la *langue* qu'ils parlent, mais non pas pour en mesurer la correction: la, ou plutôt LES langues qu'ils (re)structurent dans leur nouveau contexte (Grosjean & Py 1991; Py & Grosjean 2002), qu'ils utilisent pour communiquer et s'intégrer, pour apprendre aussi, dans le cadre de "séquences potentiellement acquisitionnelles" (de Pietro *et al.* 1989); les *migrants*, auxquels Bernard n'a cessé de s'intéresser, pour comprendre comment ils (re)construisent leur environnement et leur identité au moyen de nouvelles pratiques langagières, alternant et mêlant les codes, et au moyen de représentations, ni "justes" ni "fausses" mais pratiques, utiles (Py 1992); les *contacts*, enfin, les *interactions* entre les langues et entre les sujets, tout à la fois le lieu et l'enjeu du bilinguisme, le moyen et le produit de l'acquisition (Py 1989).

Les nombreux travaux qui suivront, et que nous ne pourrons tous présenter ici³, s'inscriront dans ce programme – qui consiste finalement à relier ses deux thèmes de prédilection, l'acquisition et le bilinguisme: "D'une certaine manière, l'acquisition est un cas particulier de bilinguisme. Considérer l'apprenant comme un bilingue revient à reconnaître la pertinence de l'enracinement de l'alloglotte dans un contexte social et discursif" (Py, 1991, cité d'après Gajo *et al.* 2004: 138). Ces travaux viendront enrichir ce programme, l'élargir, l'infléchir parfois, mais sans jamais s'en détourner dans l'esprit, sans jamais perdre ce qui fait sa force. On se contentera ici d'en citer quelques jalons: *Vers un modèle exolingue de la communication interculturelle* (Alber & Py 1986), première touche de ce qui allait devenir l'étude de la conversation et des stratégies exolingues en situation d'interaction; *L'acquisition vue dans la perspective de l'interaction* (Py 1989) et *Acquisition et contrat didactique: les séquences potentiellement acquisitionnelles dans la conversation exolingue* (de Pietro *et al.* 1989), textes qui fondent le lien entre l'étude des interactions et celle de l'acquisition; *Acquisition d'une langue étrangère et altérité* (1992) et *L'apprenant et son territoire: système, norme et tâche* (1993), textes dans

² Outre ce premier projet sur le bilinguisme, cette collaboration va encore donner lieu à d'autres projets conduits dans le cadre du FNRS: *Aspects de la communication entre Suisses romands et alémaniques: structuration de la conversation exolingue* (1984-1986) – *Migration interne, contacts linguistiques et conversation* (1986-1989; Lüdi *et al.* 1995). Bernard n'a jamais cessé par la suite de conduire des projets en collaboration, avec l'Université de Lausanne (Gajo & Mondada 2000; Porquier & Py 2004), avec la Vallée d'Aoste et Andorre, avec Iwar Werlen, etc.

³ L'idéal pour avoir une idée plus large de ces travaux est de lire l'ouvrage édité par L. Gajo, M. Matthey, D. Moore & C. Serra (2004), qui offre un magnifique choix des principaux textes de Bernard complétés par les commentaires de collègues et ami-e-s. On y trouve notamment une bibliographie plus complète de ses travaux.

lesquels Bernard développe sa conception du sujet et de son travail de construction en contexte; *Dynamique des représentations dans des situations de migration. Étude de quelques stéréotypes* ([Oesch-]Serra & Py 1993), *Le discours comme médiation: exemple de l'apprentissage et des représentations sociales* (2000), *Pour une approche linguistique des représentations sociales* (2004), trois textes (parmi d'autres encore) dans lesquels il redéfinit la place et le rôle des représentations dans l'ensemble des processus interactionnels, acquisitionnels et identitaires en jeu; *Bilinguisme et plurilinguisme* (Py & Gajo 2013), enfin, un dernier texte écrit en collaboration, paru après son décès, qui s'inscrit parfaitement dans l'évolution actuelle des recherches, du *bi-* au *plurilinguisme*.

Les éléments mentionnés ici, parmi d'autres encore (alternance codique / marques transcodiques (Lüdi éd. 1987), conversation exolingue / bilingue (Alber & Py 1986; Py 1991), interaction (Py 1990), etc.), témoignent d'une personnalité toujours plus affirmée et de l'importance de ses apports au développement de la linguistique appliquée et à son orientation vers les pratiques sociales, réelles, des acteurs, loin de toute idéalisation et de toute idéologisation.

Ce faisant, c'est pour la constitution de la discipline *linguistique appliquée* que Bernard Py a joué un rôle décisif. C'est en effet grâce à la qualité et à la pertinence de ces travaux qu'une véritable chaire de linguistique appliquée – sauf erreur la première en Suisse, en tout cas dans la partie francophone – a été créée à l'Université de Neuchâtel. Lors de sa leçon inaugurale, magistrale, Bernard avait justifié l'existence de cette nouvelle discipline en la démarquant de la linguistique générale: l'étude de l'acquisition et de l'apprentissage, en effet, n'est pas un simple sous-produit des théories linguistiques mais bien un domaine en soi qui nécessite des notions (néocodage, alloglotte, alternance codique, bifocalisation de la communication, variation, restructuration, etc.) et des démarches nouvelles (prise en compte de l'articulation entre pratiques et représentations, inscription forte des travaux dans leur contexte, etc.) – un domaine qui interroge à son tour les concepts et théories de la linguistique générale (synchronie / diachronie, système et norme, etc.). Il y avait démontré par exemple, de manière convaincante, que le fameux modèle de Jakobson devait être revu à l'aune du sujet parlant, que tout n'était pas précodé, que le système n'était pas donné d'avance ni nécessairement partagé par les interlocuteurs, nous rappelant que la compréhension n'est souvent qu'un cas particulier du malentendu. Pour Bernard, "L'étude de l'acquisition (ou de tout autre phénomène langagier socialement ou psychologiquement situé) est une voie d'accès à l'étude du langage considéré de manière générale et globale" (2000; cité d'après Gajo *et al.* 2004: 248). La dénomination même de linguistique *appliquée* pouvait de ce fait apparaître comme quelque peu paradoxale...

Les années 80 furent donc celles des grands projets d'équipe et de la reconnaissance institutionnelle. Bernard a été jusqu'à sa retraite, en 2002, professeur de linguistique appliquée et directeur du Centre de linguistique appliquée (CLA). Au cours de ces années, et jusqu'en 1996, il a également pris la direction du Cours pour la formation d'orthophonistes de l'Université de Neuchâtel, contribuant largement à la réorientation de celui-ci vers des approches interactives et à la mise en place progressive d'un Institut et d'une première chaire qui sera occupée par Geneviève de Weck.

L'air de rien – car sa modestie était légendaire –, Bernard a ainsi largement contribué à faire de la linguistique appliquée ce qu'elle est aujourd'hui. C'était un esprit libre, jamais enfermé dans quelque cadre théorique que ce soit, sans cesse en mouvement, un créateur. Ce sont ces traits de caractère qui l'ont conduit de l'analyse contrastive à la notion d'interlangue, à une conception dynamique et originale du *sujet parlant* – constructeur de micro-systèmes éphémères, actif dans la communication et l'élaboration interactive du sens à travers des activités langagières nées de son intelligence, de son inventivité – que s'efforçait de décrire et comprendre le linguiste qu'il était. Un linguiste *au service des locuteurs* – migrants, enseignants (Berthoud & Py 1993), élèves... – intéressé avant tout par leur créativité et leur ingéniosité. Sensible aussi à leurs représentations et à leur inscription dans le tissu social, à leur recherche identitaire d'un *territoire* de l'apprenant (Py 1993). C'était ce *sujet parlant* qui l'intéressait, plutôt que le système, le parcours individuel plutôt que la norme (Baggioni & Py 1987), la ruse plutôt que la soumission à la règle.

Il était un théoricien du mouvement, un peu iconoclaste, mais pas du changement brutal: du mouvement discret, du bricolage, tout en subtilité, en nuance. Pourtant, toujours *l'air de rien*, il a progressivement abordé des questions cruciales pour la théorie linguistique et, plus encore, pour la constitution d'une linguistique appliquée qui ne soit pas inféodée à la linguistique générale.

Bernard était aussi un homme de contact. Il concevait la recherche comme une entreprise collective, et il n'a cessé de s'engager dans des projets d'équipe. Et – peut-être encore plus important pour lui – il est toujours resté, avant tout, un enseignant, marquant des générations d'étudiants, les fascinant par son mélange de rigueur, de *pensée en action* et de charme. Comme me l'a écrit à son propos un autre Bernard – Schneuwly – dans un mot émouvant, il "était une belle personne, généreuse, toujours à l'écoute, humble, avec toujours cette touche de différence qui lui permettait de voir les choses un peu autrement et de nous faire voir des choses que nous ne percevions pas."

Depuis quelques années déjà, en raison de la maladie, il avait dû restreindre son activité, mais il restait pourtant en éveil, intéressé à tout ce qui concernait le langage, mais aussi sa maladie – qu'il parvenait quelque peu à garder à distance en la comprenant mieux. Il en avait d'ailleurs parlé avec acuité et

lucidité dans le cadre d'un émouvant reportage radiophonique réalisé par Francine Del Coso, une ancienne collaboratrice de l'équipe.

Bernard nous a quittés, mais son souvenir demeure bien vivant, ainsi que ses écrits et la marque profonde qu'il a imprégnée à la linguistique appliquée. Par ce bref hommage, nous lui exprimons toute notre affection et notre reconnaissance, et disons notre profonde sympathie à Thérèse Jeanneret, sa compagne, à ses enfants ainsi qu'à toute sa famille et ses proches.

BIBLIOGRAPHIE

- Alber, J.-L. & Py, B. (1986). Vers un modèle exolingue de la communication interculturelle. *Etudes de Linguistique Appliquée*, 61, 78-90.
- Arcaïni, E., Py, B. & Favretti Rossini, R. (1979). *Analyse contrastive et apprentissage des langues: la syntaxe de l'interrogation en espagnol, français, italien et anglais*. Bologne: Pàtron.
- Baggioni, D. & Py, B. (1987). Conversation exolingue et normes. In: H. Blanc, M. Le Douaron & D. Véronique (éds.), *Colloque international: acquisition d'une langue étrangère: perspectives et recherches* (pp. 72-81). Paris: Didier.
- Berthoud, A.-Cl. & Py, B. (1993). *Des linguistes et des enseignants*. Berne: Peter Lang.
- De Pietro, J.-F., Matthey, M. & Py, B. (1989). Acquisition et contrat didactique: les séquences potentiellement acquisitionnelles de la conversation exolingue. In: D. Weil & H. Fugier (éds.), *Actes du troisième colloque régional de linguistique* (pp. 99-124). Strasbourg: Université (Réédité en 2004, in L. Gajo et al. (éds.), pp. 79-93).
- Gajo, L., Matthey, M., Moore, D. & Serra, C. (éds.) (2004). *Un parcours au contact des langues. Textes de Bernard Py commentés*. Collection LAL (Langues et apprentissage des langues). Paris: Crédif-Didier.
- Gajo, L. & Mondada, L. (2000). *Interactions et acquisitions en contexte. Modes d'appropriation de compétences discursives plurilingues par de jeunes immigrés*. Fribourg: Editions universitaires (préface de B. Py).
- Grosjean, F. & Py, B. (1991). La restructuration d'une première langue: l'intégration de variantes de contact dans la compétence de migrants bilingues. *La Linguistique*, 27 (2), 35-60.
- Lüdi, G. & Py, B. (1984). *Zweisprachig durch Migration. Einführung in die Erforschung der Mehrsprachigkeit am Beispiel zweier Zuwanderergruppen in Neuenburg (Schweiz)*. Tübingue: Niemeyer.
- Lüdi, G. & Py, B. (2003). *Etre bilingue* (3^e édition revue). Berne: Peter Lang. (1^{ère} éd. 1986; éd. allemande: 1984).
- Lüdi, G. (Ed.). (1987). *Devenir bilingue, parler bilingue*. Tübingue: Niemeyer.
- Lüdi, G., Py, B., de Pietro, J.-F., Franceschini, R., Matthey, M., Oesch-Serra, C. & Quiroga, C. (1995). *Changement de langage et langage du changement*. Lausanne: L'Age d'Homme.
- [Oesch]-Serra, C. & Py, B. (1993). Dynamique des représentations dans des situations de migration. Étude de quelques stéréotypes. *Bulletin CILA*, 57, 71-83.
- Porquier, R. & Py, B. (2004). *Apprentissage d'une langue étrangère: contextes et discours*. Paris: Didier (coll. Essais).
- Py, B. (1971). *La interrogación en el español hablado de Madrid*. Bruxelles: Aimav.
- Py, B. (1973). Analyse des erreurs et acquisition des structures interrogatives du français. *Bulletin CILA*, 17, 21-29.

- Py, B. (1980). Quelques réflexions sur la notion d'interlangue. *TRANEL*, 1, 31-54.
- Py, B. (1989). L'acquisition vue dans la perspective de l'interaction. *DRLAV*, 41, 83-100.
- Py, B. (1990). Les stratégies d'acquisition en situation d'interaction. In: D. Gaonac'h (éd.), *Acquisition et utilisation d'une langue étrangère. L'approche cognitive* (pp. 81-88). Numéro spécial de *Le français dans le Monde*.
- Py, B. (1991). Bilinguisme, exolinguisme et acquisition. *TRANEL*, 17, 147-161.
- Py, B. (1992). Acquisition d'une langue étrangère et altérité. *Cahiers de l'Institut de linguistique et des sciences du langage*, 2, 113-126.
- Py, B. (1993). L'apprenant et son territoire: système, norme et tâche. *AILE*, 2, 9-24.
- Py, B. (1995). Quelques remarques sur les notions d'exolinguisme et de bilinguisme. *Cahiers de praxématique*, 25, 79-95.
- Py, B. (1997). Pour une perspective bilingue sur l'enseignement et l'apprentissage des langues. *Etudes de linguistique appliquée*, 108, 495-503.
- Py, B. (2000). Le discours comme médiation. In: A.-C. Berthoud & L. Mondada (éds.), *Modèles du discours en confrontation* (pp. 117-130). Berne: Peter Lang.
- Py, B. (2004). Pour une approche linguistique des représentations sociales. *Langages*, 154, 6-19.
- Py, B. & Gajo, L. (2013). Bilinguisme et plurilinguisme. In: J. Simonin & S. Wharton (éds.), *Sociolinguistique du contact, modèles, théories. Dictionnaire encyclopédique des termes et concepts*. Lyon: ENS Editions.
- Py, B. & Grosjean, F. (2002). Variantes de contact, restructuration et compétence bilingue: approche expérimentale. In: V. Castellotti & B. Py (éds.), *La notion de compétence en langue* (pp.19-27). Lyon: ENS éditions. (Notions en questions n° 6).

Corpus et apprentissage-enseignement des langues

Henry TYNE

Université de Perpignan Via Domitia et équipe VECT (EA 2983)
52 avenue Paul Alduy, 66000 Perpignan, France
henry.tyne@univ-perp.fr

1. Introduction

Un corpus est une collection de textes (oraux ou écrits) réunis selon un certain nombre de principes et permettant d'observer des faits de langue. Les corpus sont présents, sous différentes formes, en didactique des langues depuis longtemps. L'atout des corpus que l'on met souvent en avant est celui de l'authenticité des données (même si un corpus n'est pas forcément constitué de données vraiment 'authentiques'¹): par exemple, le Corpus d'Orléans fournit dans les années 1970 des instances véritables d'utilisation de la langue cible pour l'enseignement du Français Langue Etrangère (FLE) (Biggs & Dalwood 1976; voir aussi Chambers 2009); les corpus permettent également de s'informer sur la situation sociale (comment sont réalisés certains actes de parole – voir le corpus LANCOM: Flament-Boistrancourt & Cornette 1999; Delahaie & Flament-Boistrancourt ce volume). Les corpus nous renseignent aussi sur le fonctionnement de la langue comme système, sur les tendances, les fréquences, etc.: c'est ainsi que des dictionnaires (à commencer par le *Dictionnaire fondamental* – voir Gougenheim 1958; voir aussi le projet plus récent de constitution du dictionnaire *COBUILD* pour l'anglais – voir Sinclair 1987), des grammaires (pour l'anglais – par ex. Biber *et al.* 1999), et plus récemment des manuels (pour l'anglais – par ex. McCarthy *et al.* 2005) ou des guides pour enseignants (pour l'anglais – par ex. O'Keefe *et al.* 2007; pour le français un ouvrage assez isolé, traitant davantage de la présentation des données-ressources que du travail sur corpus – Detey *et al.*

¹ Par exemple, Allan (2009) se sert d'un corpus de romans 'simplifiés' en didactique de l'anglais langue seconde (L2). Par ailleurs, Bilger & Campione (2002), dans un travail descriptif (français langue première ou L1) portant sur le découpage syntaxique en séquences fonctionnelles, ont recours à un corpus de "textes fabriqués". Je ne reviendrai pas ici sur le fond de la question de l'authenticité des documents dans l'apprentissage-enseignement des L2 qui a déjà été amplement traitée par ailleurs (par ex. voir *Mélanges CRAPEL*, 31; *Bulletin suisse de linguistique appliquée*, 92). Je me contenterai de rappeler la place aujourd'hui incontournable des documents authentiques en didactique des langues (malgré les débats, anciens ou en cours – voir différents articles dans les numéros 30 et 31 de la revue *Applied Linguistics*). Enfin, les termes 'L1' et 'L2' sont utilisés ici de façon générale, d'après la distinction classique des acquisitionnistes (cf. Klein 1989), et la 'L2' n'est pas censée se démarquer de la 'LE' qui est ici son synonyme.

2010), ont pris le parti de présenter la langue telle qu'elle se manifeste à travers les données attestées, en mettant en avant des aspects tels que la fréquence d'utilisation, mais aussi les liens qu'entretiennent les mots avec leur 'voisinage' ou contexte d'utilisation (collocations), la distribution de la grammaire en fonction du registre ou de la situation d'utilisation, etc.

Mais dans ce qui vient d'être évoqué, le corpus, c'est-à-dire la masse de données, n'est pas touché directement par l'apprenant. Il s'agit d'utilisations indirectes. Les utilisations directes (voire semi-directes avec l'enseignant qui joue un rôle de médiateur – voir Kerr ce volume), où l'apprenant travaille sur les données, se développent plus tardivement, en partie à cause du retard de la mise en place des technologies nécessaires, mais aussi en partie à cause du défi que représente l'apprentissage sur corpus (Conrad 2000). Par ailleurs, une majorité écrasante des études existantes sur l'utilisation de corpus en didactique (ainsi que des manuels et autres outils, sans parler des corpus eux-mêmes) concerne l'anglais.

Ce numéro thématique se propose de réunir une série d'articles dédiés plus spécifiquement à l'apprentissage sur corpus (pris au sens le plus large, allant de l'utilisation d'un concordancier à l'observation de données dans la formation d'enseignants) en contextes francophones, c'est-à-dire que les contributions portent soit sur l'apprentissage du français, langue typiquement peu visée par les recherches existantes dans ce domaine, soit sur des publics d'apprenants francophones confrontés à d'autres langues. Ce numéro sera également l'occasion de faire le point sur ce que nous entendons par 'apprentissage sur corpus'.

2. 'Data-driven learning'

L'apprentissage 'conduit ou porté par les données' (traduction littérale de '*data-driven learning*' – voir la question de la définition ci-dessous) s'est développé au cours des vingt et quelques dernières années (depuis Johns [1990] 1991) et en est venu à s'imposer en tant que véritable nouvelle approche et en tant que paradigme de recherche en didactique des langues (surtout dans les sphères anglo-saxonnes). Selon cette approche, qui prend les données comme point de départ pour l'apprentissage, les apprenants sont amenés à découvrir et à cerner la langue, ainsi que les différents usages, à travers le contact avec les données généralement réunies sous forme de corpus (Chambers 2010). Par exemple, un exercice simple peut porter sur l'observation des occurrences d'un mot ou d'une expression sous forme de lignes dans ce qu'on appelle une 'concordance' (Fig. 1 ci-dessous). Voir Kerr (ce volume) pour une présentation d'exercices à partir de concordances. Mais l'accès aux données peut être autre et l'étude de concordances est loin d'être la seule façon de concevoir la rencontre entre apprenants et corpus (voir Boulton & Tyne ce volume).

un édifice dont on doit voir les dimensions d' un coup d'	oeil	, après l' avoir examiné par parties & en détail. Que l'
bon ... Par contre il y a une scène qui mérite un coup d'	oeil	, c' est lorsque le raz de marée détruit la côte Est des
de toutes ses forces à cette tentation de l' Oeil pour	oeil	, dent pour dent vengeur, engendrant la haine) pour les
toi-même » du Nouveau Testament mettant fin à « Oeil pour	oeil	, dent pour dent » de l' Ancien Testament. Suis -je dans le
que je m' attends au pire puisqu'il dépérissait à vue d'	oeil	, déshydraté et intoxiqué. Je n' ai jamais voulu perdre
. Le tapage du meurtre l' avait attirée. D' un large coup d'	oeil	, elle comprit tout, et s' enfuyant d' horreur laissa tomber
élégance qui brillait dans chaque objet que rencontrait son	oeil	, et plus encore dans la disposition et l' arrangement des
horloge d' échecs est utile au bénéficiaire, attrape le L'	oeil	, et s' assure qu' un plus de joueur dans votre événement
bien proportionné, & qu' il ne passe pas la portée de l'	oeil	, il n' importe. Voilà la règle de la nature telle qu'
: le plafond très bas et le lit fort haut. D' un coup d'	oeil	, j' en explorai l' intérieur, puis je me glissai près de la

Fig. 1: Concordance pour *œil* (source: Leeds Internet Corpus)

Le terme '*data-driven learning*' (dont la forme abrégée anglais est 'DDL' – à ne pas confondre avec la 'Didactique Des Langues'), s'il pose déjà problème pour la traduction en français (nous optons ici pour 'apprentissage sur corpus' ou 'ASC', ce qui ne rend pas très bien le '*data-driven*' de l'original, mais qui a au moins le mérite d'englober à peu près toutes les acceptions du terme²), n'a pas de définition exacte en anglais (Boulton 2011a): pour les uns, dans une version "hard" (Gabrielatos 2005), il est avant tout synonyme d'autonomie, et l'apprenant est un chercheur, voire même un "détective" (Johns 1997) ou un enquêteur enregistrant et transcrivant les données (Tyne 2009); pour les autres, dans une forme plus "soft", il s'agit plutôt de profiter (directement ou indirectement) de ce qu'offre l'approche sur données, tout en maintenant un cadre plus ou moins 'classique' (création d'exercices reconnus – voir par ex. les textes à trous chez Römer 2006; voir aussi Kerr ce volume).

Si les approches les plus autonomisantes, voire les plus sérendipiteuses semblent parfois incompatibles avec les contextes ordinaires d'enseignement des langues (Mukherjee 2006), ou semblent aller au-delà des compétences et habitudes dont disposent enseignants et apprenants (au risque même de trop réduire le regard que porte l'apprenant sur la langue – Kamber 2011: 200), il convient toutefois de rappeler que l'ASC n'est pas forcément une pratique (ou ensemble de pratiques) qui est censée remplacer l'existant et n'a pas la prétention de changer à tout jamais le rôle de l'enseignant. En revanche, comme le prétendent Boulton et Tyne (à paraître), l'ASC (et l'utilisation de corpus en général en didactique des langues) mérite d'avoir une place dans les dispositifs dans la mesure où il présente des avantages sur le plan pédagogique mais aussi acquisitionnel et s'apparente à bien des pratiques qui existent déjà à la fois dans le domaine de l'éducation et dans nos vies ordinaires (voir la section 3 ci-dessous).

² Il est à noter d'ailleurs que la différence que l'on identifie parfois en linguistique de corpus ("*corpus-based approach*" quand on se sert du corpus afin de tester une théorie existante; "*corpus-driven approach*" quand on part des seules données sans théorie *a priori* – voir Tognini-Bonelli 2001) ne pose pas problème avec une appellation large et englobante. Voir par ailleurs McEnery *et al.* (2006: 8) pour qui cette opposition n'a pas forcément lieu d'exister.

L'application des pratiques et des méthodes d'observation qui ont été développées en linguistique de corpus n'est pas nécessairement très aisée pour la didactique des langues; si bon nombre de travaux descriptifs sur corpus restent à ce jour peu ou pas utilisés par les didacticiens, qu'en est-il de leur utilisation de façon plus ou moins directe par des enseignants ou par des apprenants? Car le propre du '*data-driven learning*', dans son état le plus 'pur', est de faire de l'apprenant un chercheur, capable de s'informer sur le fonctionnement de la langue cible à partir des données brutes. Tout en allant dans le sens des arguments cognitivo-constructivistes (O'Sullivan 2007), une telle approche soulève néanmoins des questions d'ordre méthodologique (comment la mettre en place?) mais aussi pratique et culturel (respect des programmes et des habitudes, contraintes horaires, devoirs, outils 'conventionnels', etc.). Elle conduit également à une réflexion sur la description de l'objet langue et sur l'application des observations pour l'apprentissage de la langue: tendances et régularités plus que règles; fréquences plus qu'exemples uniques, etc.

3. Révolution ou évolution?

Les grands changements méthodologiques ou philosophiques en didactique des langues surviennent de manière cyclique (tous les 20-25 ans, disons) et ne sont pas toujours bien accueillis par les différents acteurs concernés (Decoo 2001). Que certains vivent l'arrivée des corpus comme une "révolution" (Rundell & Stock 1992; McCarthy 2008), ou voient leur rôle dans la vulgarisation des méthodes de l'ASC comme une "mission" (Römer 2009), est certainement justifié dans la mesure où, comme le constate Sinclair (2004), l'écart qui existe entre la description des données attestées et la manière dont on a pris l'habitude de présenter la langue depuis des siècles est très important. Et comme le souligne également Sinclair (2004), le fait de recourir aux données attestées peut poser problème: les observations sur données attestées mettent à mal l'approche dichotomique en grammaire qui voit les choses comme étant 'correctes' ou 'incorrectes' (Conrad 2010), et toutes les formes rencontrées ne rentrent pas facilement dans les cases existantes (cf. Chambers ce volume).

Cependant, ce qui constitue indéniablement un changement très important en termes de description de la langue ne doit pas forcément être vécu comme une révolution lors de son application à la didactique des langues. En termes des pratiques et des méthodes mises en avant par le travail sur corpus, il y a un continuum qui fait que l'accueil de l'ASC peut se faire en douceur à certains égards. En effet, le développement des pratiques de découverte dans le cadre de la perspective actionnelle va tout à fait dans le sens de l'apprenant comme chercheur; la mise en avant de pratiques autonomisantes y est également propice (O'Sullivan 2007). Par ailleurs, le rôle incontournable que jouent les

TIC dans les dispositifs en général aujourd'hui est un atout pour développer l'ASC, tout comme l'est le développement de pratiques ordinaires d'observation et de recherche via les moteurs de recherche comme Google, par exemple (voir ci-dessous).

Au-delà de ces constats, il y a également aujourd'hui une volonté générale de procéder à la mise en place de savoirs via la construction, plus ou moins en autonomie, qui résulte d'une approche comparative privilégiant ce qui émerge des données, quelles qu'elles soient, et cela au-delà même de la didactique des langues. Par exemple, en discutant récemment avec une amie maîtresse formatrice, j'ai été frappé par son discours en ce qui concerne le travail en arts plastiques à l'école et par la manière dont ses propos rappelaient, sur le fond, des arguments de l'ASC. En effet, elle critiquait un collègue qui avait travaillé sur des tableaux afin que les élèves produisent des 'copies' dans le style de Dalí, Picasso, Herbin, Arcimboldo, etc. Elle m'expliquait la différence entre le "faire à la manière de" et le "observer des exemples pour". Ainsi, plutôt que de présenter un tableau pouvant être étudié de fond en comble avant de servir de modèle ou de simple mise en route, on peut aussi présenter plusieurs tableaux différents (mais qui traitent par exemple d'un même thème ou qui répondraient à la même consigne). Dans ses explications, cette amie mettait en avant le fait que les élèves parviennent à mieux saisir les caractéristiques importantes, définitoires lorsque celles-ci se répètent. Par exemple, dans une collection de portraits, on verrait à la fois les éléments récurrents (qui permettent de voir qu'il s'agit bien de portraits) mais aussi les touches individuelles, propres à chaque tableau ou à chaque artiste. Bref, elle soutenait l'idée selon laquelle il paraît aujourd'hui plus intéressant, pour un certain nombre de raisons à la fois pédagogiques et artistiques, de demander à un enfant de comprendre (mais aussi de produire) à partir de l'étude d'un ensemble plutôt que de faire un exercice plus ou moins de copiage (fût-il même très bien réussi au niveau du produit final) à partir d'une seule œuvre.

Cette anecdote est utile car elle sert à illustrer deux points importants. Tout d'abord, elle sert à comprendre la logique de base de ce que constitue l'ASC (pris au sens large), et ce face à l'approche qui consiste à travailler certes sur des documents authentiques mais où chaque document constitue un tout, un modèle unique. Deuxièmement, cette anecdote permet de mettre en avant un point que soulignent Boulton et Tyne (ce volume; voir aussi à paraître) quant à l'importance des pratiques existantes, parfois très simples, mais qui s'appuient au fond sur une même façon d'envisager l'objet d'étude. Par exemple, dans l'utilisation ordinaire de l'ordinateur sur Internet l'idée est de rechercher, d'être confronté à une masse de résultats qu'on devra ensuite trier avant de décider ce qu'on retient comme information pertinente par rapport à la requête originelle: on navigue, on cherche, on consulte, on fait défiler des listes, etc. Le principe du 'web comme corpus' qui, au départ, ne suscitait pas

l'engouement des linguistes (voir Sinclair 2005: sn³, pour qui le "*World Wide Web is not a corpus*") ne peut être ignoré aujourd'hui: les possibilités de recherche actuelles sont assez performantes (avec les fonctions avancées dans Google, par exemple, ou en se servant d'Internet pour confectionner un corpus via BootCat⁴), la recherche par mot-clé (que l'on trouve parfois intuitivement, par exemple, en entrant le simple nom d'un produit ou quelque autre mot associé) est devenue une pratique courante. Par ailleurs, la quantité de données qui est consultable en direct sur Internet (depuis le Web 2.0) est telle qu'elle représente aujourd'hui une masse importante, inégalable par d'autres moyens, et facilement accessible (Boulton 2011b).

4. Présentation du numéro

Les cinq articles qui constituent ce numéro concernent, de près ou de loin, l'application du l'ASC dans des contextes différents mettant en jeu l'apprentissage du français L2 ou l'apprentissage de l'anglais par des apprenants francophones. Le numéro s'organise en fonction du contenu et des considérations des différentes contributions: apports théoriques et applications, études empiriques et description des données pour l'apprentissage-enseignement, bilan et perspectives.

Dans le premier article, **Betsy Kerr**, partant d'un principe important mis en avant par l'étude de corpus et théorisé notamment par Sinclair, à savoir la prise en compte du lexique par la 'lexico-grammaire', travaille sur les applications des corpus pour l'enseignement du français L2. On voit comment les exercices proposés par Kerr épousent à la fois des considérations théoriques et linguistiques et des considérations davantage liées à l'appropriation de la langue, le tout dans un contexte d'ASC médié par l'enseignant.

Ensuite, **Angela Chambers** montre dans son article comment le corpus permet de faire émerger des occurrences du subjonctif afin d'offrir une source riche en exemples aux apprenants pour qui le subjonctif en français est souvent pris pour une espèce d'épouvantail. Non seulement le corpus permet de compléter de manière pratique la présentation classique des 'règles' du subjonctif, mais il permet aussi de voir en quoi le subjonctif (en dehors des explications sur la possibilité) est une affaire de formules en *que*, ce qui débouche sur la découverte inévitable d'occurrences autres dont la forme *après que* + subjonctif... Sont ainsi traitées des questions de méthodologie de travail sur les données et des questions de gestion des formes qui ne rentrent pas forcément dans le moule des explications plus 'classiques'.

³ Sans numérotation.

⁴ <http://bootcat.sslmit.unibo.it/> (15.05.13).

L'expérience présentée par **Corinne Landure**, travaillant dans le cadre de l'anglais LANSAD⁵, prétend démontrer comment la consultation de corpus par les apprenants permet de développer la prise de conscience linguistique (le fait de pouvoir prendre du recul, de repérer et choisir les bonnes formules, etc.) en L2; le corpus serait même plus bénéfique que le dictionnaire à certains égards. Toutefois, elle rappelle qu'il est important de préparer les apprenants à la consultation de corpus et que cette préparation serait bénéfique comparée à celle ciblant l'utilisation d'outils plus 'conventionnels' (dictionnaires).

L'article de **Juliette Delahaie** et **Danièle Flament-Boistrancourt** revient sur les 'effets' de la formation dispensée aux enseignants de FLE en Belgique flamande suite à l'élaboration du corpus LANCOM. Se basant sur l'analyse des productions d'une enseignante formée à partir du corpus LANCOM dans les années 1994-1995, ces auteures prétendent que, douze ans après la formation initiale, l'expérience a été concluante car elle enseigne un français oral proche de ce que l'on observe dans les données attestées.

Enfin, l'article d'**Alex Boulton** et **Henry Tyne** dresse un bilan de l'ASC depuis ses débuts en tentant de comprendre les possibles évolutions d'une approche qui peine à s'imposer mais qui semble tout à fait en adéquation avec des pratiques existantes en didactique des langues. Ces auteurs plaident pour une application large de l'ASC en insistant sur la possibilité de travailler sur des activités simples qui ne requièrent pas nécessairement de corpus 'fini' (par ex. le travail sur la collecte de données, sur la transcription; l'utilisation de techniques *ad hoc* comme les concordances au tableau, etc.).

BIBLIOGRAPHIE

- Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal*, 63, 23-32.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Londres: Pearson.
- Biggs, P. & Dalwood, M. (1976). *Les Orléanais ont la parole*. Londres: Longman.
- Bilger, M. & Campione, E. (2002). Propositions pour un étiquetage en 'séquences fonctionnelles'. *Recherches sur le français parlé*, 17, 11-37.
- Boulton, A. (2011a). Data-driven learning: the perpetual enigma. In: S. Gozdz-Roszkowski (éd.), *Explorations across languages and corpora* (pp. 563-580). Francfort: Peter Lang.
- Boulton, A. (2011b). Bringing corpora to the masses: free and easy tools for language learning. In: N. Kübler (éd.), *Corpora, language, teaching, and resources: from theory to practice* (pp. 69-96). Berne: Peter Lang.

⁵ 'LANGues pour Spécialistes d'Autres Disciplines', c'est-à-dire la provision de cours de langues (souvent de l'anglais) pour des étudiants spécialistes dans d'autres domaines (par ex. des cours d'anglais pour les étudiants en architecture).

- Boulton, A. & Tyne, H. (à paraître). Corpus-based study of language and teacher education. In: M. Bigelow & J. Ennser-Kananen (éds.), *Handbook of educational linguistics*. New York: Routledge.
- Bulletin suisse de linguistique appliquée*, 92 (2010). *L'exploitation didactique des documents authentiques audio et vidéo dans l'enseignement des langues étrangères*, M. Dubois, A. Kamber & C. Skupien Dekens (éds.)
- Chambers, A. (2009). Les corpus oraux en français langue étrangère: authenticité et pédagogie. *Mélanges CRAPEL*, 31, 15-33. http://www.atilf.fr/IMG/pdf/melanges/01_Chambers.pdf (15.05.2013).
- Chambers, A. (2010). What is data-driven learning? In: A. O'Keeffe & M. McCarthy (éds.), *Routledge handbook of corpus linguistics* (pp. 345-358). Londres: Routledge.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st Century? *TESOL Quarterly*, 34, 3, 548-560.
- Conrad, S. (2010). What can a corpus tell us about grammar? In: A. O'Keeffe & M. McCarthy (éds.), *Routledge handbook of corpus linguistics* (pp. 227-240). Londres: Routledge.
- Decoo, W. (2001). *On the mortality of language learning methods*. Communication faite à Brigham Young University, 8 novembre 2001. <http://www.didascalie.be/mortality.htm> (08.07.2011 via Internet Archive).
- Detey, S., Durand, J., Laks, B. & C. Lyche (éds.) (2010). *Les variétés du français parlé dans l'espace francophone. Ressources pour l'enseignement*. Paris: Ophrys.
- Flament-Boistrancourt, D. & Cornette, G. (1999). Bon français ou vrai français? Une étude de l'acte de question menée à partir d'un extrait du corpus LANCOM: les scènes dites du baby-sitting. *Travaux de linguistique*, 38, 119-153.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal*, 8, 4, 1-35. <http://tesl-ej.org/ej32/a1.html> (18.12.2012).
- Gougenheim, G. (1958). *Dictionnaire fondamental de la langue française*. Paris: Didier.
- Johns, T. ([1990] 1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. [Version remaniée: T. Johns & P. King, (1991). Classroom Concordancing. *English Language Research Journal*, 4, 27-45].
- Johns, T. (1997). Contexts: the background, development and trialling of a concordance-based CALL program. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (éds.), *Teaching and language corpora* (pp- 110-115). Londres: Longman.
- Kamber, A. (2011). Contexte et sens: utilisation d'un corpus écrit dans l'enseignement/apprentissage du FLE. *Travaux neuchâtelois de linguistique*, 55, 199-218.
- Klein, W. (1989) [traduction]. *L'acquisition de langue étrangère*. Paris: Armand Colin.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41, 4, 563-574.
- McCarthy, M., McCarten, J. & Sandiford, H. (2005). *Touchstone*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies*. Londres: Routledge.
- Mélanges CRAPEL*, 31 (2009). *Des documents authentiques oraux aux corpus: questions d'apprentissage en didactique des langues*, A. Boulton (éd.)
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: the state of the art – and beyond. In: S. Braun, K. Kohn & J. Mukherjee (éds.), *Corpus technology and language pedagogy: new resources, new tools, new methods* (pp. 5-24). Francfort: Peter Lang.

- O'Keefe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL*, 19, 3, 269-286.
- 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*, 54, 121-134.
- Römer, U. (2009). Corpus research and practice: what help do teachers need and what can we offer? In: K. Aijmer (éd.), *Corpora and language teaching* (pp. 83-98). Amsterdam: John Benjamins.
- Rundell, M. & P. Stock (1992). The corpus revolution. *English Today*, 8, 4, 41-55.
- Sinclair, J. (1987). *Looking up: an account of the Cobuild project in lexical computing*. London: Collins.
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. In: J. Sinclair (éd.), *How to use corpora in language teaching* (pp. 271-299). Amsterdam: John Benjamins.
- Sinclair, J. (2005). Corpus and text-basic principles. In: M. Wynne (éd.), *Developing linguistic corpora: a guide to good practice* (pp. 1-16). Oxford: Oxbow Books. <http://ota.ahds.ac.uk/documents/creating/dlc/index.htm> (16.05.2013).
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tyne, H. (2009). Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL*, 31, 91-111. http://www.atilf.fr/IMG/pdf/melanges/05_Tyne.pdf (16.05.2013).

Grammatical description and classroom application. Theory and practice in data-driven learning

Betsy KERR

University of Minnesota

9 Pleasant St. S.E., Minneapolis MN 55422, USA

bjkerr@umn.edu

Partant de certains concepts intéressants du domaine de la linguistique sur corpus, cet article défend l'idée de faire une place plus importante à l'acquisition du vocabulaire et de mieux intégrer les aspects lexicaux et grammaticaux dans l'enseignement de la langue. Nous nous intéressons tout particulièrement à l'idée de la centralité des séquences lexicalisées dans les productions orales, telle qu'elles sont identifiées dans la grammaire lexicale de Sinclair puis appliquées à la didactique des langues avec l'approche de Lewis. De plus, cet article montre, à partir d'activités types, comment l'apprentissage à partir de corpus est propice aux activités métacognitives centrées sur l'apprenant, pointant des aspects importants concernant le lexique et la grammaire du français. Finalement, nous passons en revue plusieurs outils qui se prêtent à la préparation de ce type d'activité par l'enseignant.

Mots-clés: Concordance, corpus, approche lexicale, didactique des langues, linguistique sur corpus, FLE, concordanciers en ligne

1. Introduction

Many proponents of data-driven learning (DDL) or the use of concordance-generated data in language teaching and learning have lamented the fact that second language educators, especially outside of the field of English L2 instruction, have not embraced the use of DDL in significant numbers (see for example Cheng 2010: 327-328). My purpose here is to encourage L2 French teachers in particular (and those responsible for L2 French teacher education) to consider trying out concordance-based activities in their classrooms. I will begin by seconding others' suggestions for a progressive approach to developing learners' abilities to utilize DDL activities, as well as enumerating some of the holistic benefits that can accrue from the regular practice of DDL. The next section will sketch out Sinclair's idiom principle, a view of language originating in corpus linguistics which sees lexis and grammar as intimately related. A pedagogical application of this view, Lewis's Lexical Approach, will be briefly described as a model for a pedagogical approach in which DDL plays a central role in facilitating learner discovery of patterns of language use that integrate lexical and grammatical properties.

The third and most substantial section of the article will be devoted to practical questions relating to teachers' use of concordance-based or DDL activities. Following a presentation of types of language features that can be addressed by means of DDL, concrete suggestions are given for transforming

concordance data into classroom activities, and a number of examples of such activities are provided. Finally, four online French corpora with built-in concordancers that lend themselves to teacher and student use are described and compared.

2. Benefits of teacher-mediated DDL activities

Boulton (2011) proposes a prototype definition of DDL situating pedagogical uses of corpora consultation on a continuum of weak to strong versions of DDL, according to the number of prototype criteria exhibited by the given application¹. Pursuant to this framework, activities other than hands-on use of raw concordance data, such as teacher-devised classroom activities based on edited concordances, constitute 'weaker' versions. Nevertheless, in Boulton's opinion and as demonstrated by explicit recommendations of the originator of the term (Johns [1990] 1991, 1993), such activities have a rightful place and play an important role in the panoply of corpus-based learner activities (see Boulton & Tyne this volume).

It is precisely activities at this end of the continuum, i.e. teacher-mediated encounters with corpus data, that will form the object of this discussion. The motivation for this choice is the belief that teacher-mediated activities offer language teachers the easiest entry into DDL and, in turn, offer the best chance for the expansion of the application of DDL in L2 instruction. A typical activity of this type engages learners in a teacher-structured process of examining and analyzing concordance data which has been assembled and edited by the teacher with a particular group of learners and particular learning goals in mind. Though such teacher interventions may diminish certain aspects of the learner's experience that have been touted as beneficial (e.g. the encounter with totally authentic language and the autonomous nature of the learner's role in formulating hypotheses), these modifications have the advantage of making corpus-based work accessible to a larger student audience than would otherwise be possible. Obviously, teacher-produced concordances do not require the same technological resources as hands-on work, and, more importantly, if well executed, the teacher's adaptations of the data will facilitate its interpretation by learners with less advanced linguistic skills than are required to decipher unedited concordancer output, with its partial sentences, limited context, and unpredictable and ever-changing topics. Though these features will remain in any concordance data, judicious selection of concordance lines, as well as the provision of guidance in the process, can make these challenges considerably more manageable for learners of less advanced proficiency levels.

¹ Gabrielatos (2005: 11) provides a similar characterization and examples of concordance activities of a "soft" version (teacher-controlled) and a "hard" version (learner-controlled).

Teacher mediation of corpus investigation thus has the potential to blunt one of the major criticisms of DDL, namely, that it can be effective only with learners of relatively well-developed linguistic skills and /or high levels of motivation (Cheng 2010: 328). Another common complaint (Chambers 2010: 348), that corpus consultation is time-consuming and laborious, is countered by the streamlining afforded by the teacher's crafting of both the data for analysis and the task itself. Several proponents of DDL (Boulton 2011; Chambers 2010) have put forth a progressive approach, starting from maximally teacher-structured and teacher-aided activities and moving gradually toward independent, unaided use of concordances. Ideally, this would be a long-term process extending over as much as two years or more of study and coinciding with the learners' advancement in degree of competence in the language. Chambers (2010: 349) proposes a three-phase model as follows: (1) working with teacher-prepared concordance printouts to become familiar with the inductive process, (2) consultation of user-friendly online resources, and (3) independent corpus consultation selecting from a range of corpora. The hope is that this kind of long-term, recurrent practice will result in increased voluntary recourse to concordance data as a reference and writing aid on the part of advanced learners. Moreover, this approach provides a "gentle 'way in'" (Johns 1993, as cited in Boulton 2011: 573) not only for learners, but also for teachers, who have so far not embraced DDL activities in significant numbers.

Proponents of DDL argue that investigation of corpus data is a more effective learning tool than the conventional deductive approach of presenting rules with a few examples, thanks primarily to two features of DDL, namely, the engagement of the learner in discovering patterns for themselves (see 3.3 below) and the provision of a larger number of examples. In addition, the authenticity of the examples is believed to enhance learning. However, in addition to affording a more effective means of acquiring particular patterns of language use, DDL activities, when practiced with some regularity, offer more holistic benefits that may be just as important, or even more important, than the knowledge gained of specific points of vocabulary or grammar. O'Sullivan (2007: 278) cites numerous researchers who claim that engaging in DDL activities has the potential to bring general cognitive benefits which are vital to language learning such as the development of analytical skills, the ability to notice linguistic features, and problem-solving abilities (see also Landure this volume). Stevens (1991: 48) adds to these "the ability to extrapolate holistically from fragmentary evidence". My own observation is that most learners do not come to language study with a syntactic mind-set; that is, L2 learners, especially in the case where L1 and L2 share a great deal of common syntactic structure and lexis, naturally tend to focus, in their efforts to extract meaning from L2 texts, on meaning rather than on form. Consequently, when learners then come to attempt to express meaning in writing or

speaking, their awareness of morphological and syntactic features is insufficient to allow them to formulate language which is native-like in its formal features. By engaging learners in activities that require attention to formal features such as grammatical functions of sentence constituents, part of speech of lexical units, and function words (prepositions) introducing complements, teachers can lead learners to acquire the habit of paying attention to these features, while stressing the importance of awareness of such features in producing effective and accurate speech and writing. In addition, we will see below examples of the generalizability of certain patterns from a particular syntactic structure or lexical item to a larger category of linguistic items.

3. Theory and application: the relationship between lexis and grammar

3.1 Sinclair's idiom principle

Second language instruction, regardless of the context or the approach adopted (and even in the context of widely adopted communicative approaches), has over recent decades privileged explicit instruction in grammar, i.e. morphological and syntactic patterns of the L2. On the other hand, acquisition of lexis has generally been left to learning of the incidental type, believed to take place more or less naturally, through reading or other encounters with the L2, thanks to redundancy and contextual clues present in natural texts and communicative interactions. While many studies (e.g. Hulstijn *et al.* 1996, Pigada & Schmitt 2006) have demonstrated positive albeit somewhat limited effects of extensive reading on vocabulary acquisition, incidental learning of lexis is usually restricted by the limited time learners actually spend in reading or other naturalistic interactions in the target language (Sckett 2011).

Insights gained over the past few decades from studies employing corpus-based methods, as well as from other fields of investigation, have created a new interest in the structure of the lexis and its role in L2 acquisition. In particular, corpus data have revealed the extent to which certain combinations of words occur with greater frequency, even where other combinations would be just as valid from a syntactic and semantic point of view. Much work has been devoted to these frequent combinations or collocations, which have now become the basis for the leading learner-oriented dictionaries as well as some course materials, first for English and then for the other most-studied European languages².

² Pioneers in this field were the Cobuild group led by John Sinclair which published the first *Collins COBUILD English Language Dictionary* in 1987 (5th edition 2006) and the *Collins COBUILD English course* (Willis & Willis 1989).

John Sinclair, a pioneer of corpus linguistics with an interest in its application to second language acquisition, was so impressed by the prevalence of collocations and other multi-word units in actual native-speaker language that he called into question the very distinction between lexis and grammar. Accordingly, he proposed the notion of a "lexical grammar" where the grammar and lexis are built together, with each on an equal basis (Sinclair 2000: 191). To better understand the kind of patterns motivating this proposal, let us consider an example of lexical dependencies revealed by corpus data. As reported by Hunston and Francis (2000: 24-25), in a 1994 paper³, Sinclair develops the example of the phrase *naked eye*. Inspection of even a short concordance⁴ leads to the following description of the phrase's typical contexts:

Typically, *naked eye* occurs at the end of a clause and is preceded by *the*. Furthermore, *the naked eye* is preceded by *to*, or, less frequently, *with*. The prepositional phrase *to / with the naked eye* follows a range of words related to sight, most frequently *visible* or a form of the verb *see*. Prior to that is an indication of something that might be too small to be seen. About half the instances of this typical usage are negative (as against a general figure of 10% of all clauses being negative, see Halliday 1993b); in other words, something is described as being *invisible to the naked eye*. Of the positive instances, over half include modification, such as *easily*, *actually*, *might be* or *the first*.

Sinclair concludes from these observations that *naked eye* is part of a larger 'meaning unit', which has a consistent meaning, even as it displays a range of possible lexical realizations.

It is this kind of observation that leads Sinclair to articulate the idiom principle: semi-preconstructed phrases such as collocations actually represent a single choice made by the language user, despite their appearing to be analyzable into segments (Sinclair 1991, cited in Hunston & Francis 2000: 21). This process of speech production through the choice of language chunks is thus opposed to the conventional 'slot-and-filler' model, where "the syntactic structures form a series of slots, and these are filled with choices from the dictionary", the lexical items being chosen individually, with relatively little reference to the surrounding text (Sinclair 2000: 195).

Sinclair's work in corpus linguistics is paralleled by research in related fields such as second language acquisition and teaching, and psycholinguistics. At the same time, the object of study has been expanded to include various types of multi-word units, generally distinguished by their relative degree of fixedness. Schmitt (2000: 99-102) provides a summary of some of the proposed taxonomies, singling out as more common the categories of "compound words, phrasal verbs, fixed phrases, idioms, proverbs, and lexical

³ "A search for meaningful units of language", paper read at the International Symposium on Phraseology, University of Leeds, April 1994, as reported in Hunston and Francis (2000: 24-25).

⁴ I recommend the Time Magazine Corpus created by Mark Davies at Brigham Young University: <http://corpus.byu.edu/time/> (17.05.2013).

phrases". Terms employed by other researchers for lexical phrases include routine formulae, gambits and lexical chunks (Schmitt 2000: 101), and the study of these units is generally referred to as 'phraseology'. Among these researchers, the lexicographical perspective is represented by Mel'čuk's work on a taxonomy of "non-free phrases" (Hunston & Francis 2000: 8-10). In psycholinguistics, lexical phrases have figured prominently in debates about L1 and L2 acquisition, as well as about how language is encoded in the brain, whether as single lexical items to be combined with the help of grammatical rules, or as multi-word units. For example, Peters (1983) argues that L1 acquisition proceeds first through the acquisition of lexical phrases that are later analyzed into a system of words and structure (Hunston & Francis 2000: 11-12). Conklin and Schmitt (2008) provide support for this hypothesis by demonstrating a processing advantage in the reading of formulaic vs. non-formulaic sequences on the part of both native and non-native readers. In the field of language teaching, Pawley and Syder (1983) and Nattinger and DeCarrico (1989, 1992) highlight the role of the acquisition of lexical phrases in the achievement of fluency: given that lexical phrases are stored as whole units, their retrieval saves much-needed processing time, as compared to that required for on-line processing of several items which must be individually selected and sequenced according to grammatical constraints.

3.2 Pedagogical application: Lewis's Lexical Approach

Michael Lewis's Lexical Approach (1993) and Lewis (1997) describe an approach to second language teaching which Lewis believes to be consistent with a lexically based view of language such as that outlined above. In what follows, I will highlight some of Lewis's basic principles and important methodological recommendations, especially as they pertain to grammar instruction and to the use of corpora and concordancing. The following are among the more distinctive "key principles" laid out by Lewis (1993: vi-vii):

- Language consists of grammaticalized lexis, not lexicalized grammar.
- The grammar / vocabulary dichotomy is invalid; much language consists of multi-word 'chunks'.
- A central element of language teaching is raising students' awareness of multi-word units, and developing their ability to 'chunk' language successfully.
- It is the co-textual rather than situational elements of context which are of primary importance for language teaching.
- Grammar as a receptive skill, involving the perception of similarity and difference, is prioritized.
- The Present-Practice-Produce paradigm is rejected, in favor of a paradigm based on the Observe-Hypothesize-Experiment cycle.

- ... the materials and methods appropriate to beginner or elementary students are radically different from those employed for upper-intermediate or advanced students.

The Lexical Approach seeks to direct learners' attention to what are "highly probable" utterances, not simply to all possible utterances, as Lewis suggests is the case with typical grammar books, following the lead of generative grammar (1993: 12).

Moudraia (2001: 2) gives the following summary of the kinds of activities proposed by Lewis to develop learners' knowledge of multi-word units:

- Intensive and extensive listening and reading in the target language.
- First and second language comparisons and translation – carried out chunk-for-chunk, rather than word-for-word – aimed at raising language awareness.
- Repetition and recycling of activities, such as summarizing a text orally one day and again a few days later to keep words and expressions that have been learned active.
- Guessing the meaning of vocabulary items from context.
- Noticing and recording language patterns and collocations.
- Working with dictionaries and other reference tools.
- Working with language corpora created by the teacher for use in the classroom or accessible on the Internet ...

Despite Lewis's de-emphasis of grammar in favor of lexis, he does acknowledge a place in the language syllabus for grammar, rightly understood. Seeing grammar as "the search for powerful patterns" (1993: 137), Lewis suggests certain aspects of grammar which deserve increased attention. Two of these are word grammar, the set of patterns in which the word occurs, and de-lexicalized words. For words with relatively high meaning-content (such as *deposit*, *house*, *baby*), Lewis suggests the following pedagogical strategies: identifying high-frequency uses and distinguishing among them; identifying cognates, that is, derived or morphologically related forms, and their collocations; identifying and recording high frequency collocations, any natural noun phrase – verb phrase combinations, and any other lexical phrases in which the word occurs (1993: 142). De-lexicalized words are words of low meaning-content, which allows them to serve as "powerful pattern generators" (1993: 143). Examples for French are the verbs *faire*, *avoir*, *prendre* and function words such as the prepositions *de* (in at least certain uses) and *en*. Clearly, these are the kinds of patterns that can be explored through corpora and concordances.

3.3 *Constructivist learning theory*

The notion of learner as scientist, implicit in Lewis's Observe-Hypothesize-Experiment cycle, draws support from the branch of learning theory known as constructivism. This model has recently attracted increased interest from educational technologists, as developments in informational technology have greatly expanded the quantity and ease of use of data sources available for data-driven learning activities. Cobb (1999: 2)⁵ contrasts the two types of knowledge believed to be created by the traditional transmission of knowledge vs. the construction of knowledge by the learner:

One kind is inert, easily forgotten, and untransferable beyond its initial context of learning because it was "pre-emptively encoded" (Spiro, Coulson, Feltovich & Anderson 1988: 377) by an expert for transmission to a learner. The other kind is memorable and transferable to novel contexts because learners have encoded it for themselves out of raw data, or at least raised it from a lower to a higher level of organization, by forming and testing hypotheses as professional scientists do (Resnick 1987).

In the same article, Cobb (1999: 1) reports on a study in which he devised and carried out an empirical test of the prediction that knowledge of the second type would be more "flexible, transferable, and useful" than knowledge of the first type. Subjects in the study were students learning English for academic purposes who were engaged in two different sets of activities for vocabulary expansion: learning dictionary definitions of words for the control group vs. constructing their own definitions of words based on observation of the word in several sentence contexts drawn from concordance data. These latter activities were facilitated by the use of a specially designed 'dictionary toolkit' modeled on a lexicographer's workstation, which linked together a teacher-generated wordlist and a learner-created database with simplified versions of a corpus and concordance engine. The study results showed that the group using concordance data to construct definitions attained better results than the control group in tasks requiring the transfer of their word knowledge to novel contexts.

4. **Pedagogical uses of concordancing**

4.1 *Teacher manipulation of the concordance and the task*

We turn now to instructional activities where the teacher mediates between the raw data and the learner, reducing the complexity of the task but retaining the essential elements of DDL. Chambers (2010: 348-349) cites two main objections voiced by learners who have tried out DDL activities: first, the "time-consuming, laborious and tedious" nature of the activities, and second, the limited nature of the context supplied by concordance lines. The latter difficulty

⁵ Citations are from the author's pre-publication version:
<http://www.lexutor.ca/cv/Constructivism.html> (01.09.2007).

can be abated to some extent by choosing from available corpora those with text-types most likely to be familiar to the given group of students; however, in most cases, learners will have to grapple with limited context. To address the first objection, teachers will reduce the raw concordance to a manageable size by selecting a limited number of lines that lend themselves to the objective of the activity. In addition, learners' interpretation of concordance lines can be significantly aided by the teacher's judicious selection of concordance lines according to their anticipated readability by the given group of learners (Gilquin & Granger 2010). Finally, the activity can be made less tedious and more efficient by the teacher's construction of the task and provision of clear guidelines for the learners to follow.

4.2 Taxonomy of linguistic features for corpus-based study

Before looking at some examples, it will be useful to review the principal kinds of patterns that can be studied with the aid of a corpus and concordancer. The following inventory of features is loosely based on Aston's discussion of pedagogical concordancing (2001: 16-19). Of course, the particular features that teachers single out for special attention should ideally be selected from texts with which learners are working, or based on recurrent errors or problem areas for the given group of learners. Also, the scope of any activity will be determined by the level and needs of the learners.

1. Lexical items (words)
 - polysemous words: different meanings of the word and the contextual patternings associated with each of them. Examples: different uses of *tirer*, *projeter*, *abattu*, *passer*.
 - lexico-grammatical features, or word grammar: different syntactic contexts and the meanings associated with them. Examples: *réussir* + noun vs. *réussir à* + noun / infinitive; reflexive vs. Non-reflexive uses of *sentir*; *penser à* + noun / infinitive vs. *penser* + infinitive, etc.
2. Multi-word units (set expressions, collocations, etc.). Examples:
 - phrases with *même*, e.g. *même si*, *quand même*, *tout de même*, *quand bien même que*, *de même (que)*.
 - phrasal verbs such as *mettre en* + NP (*œuvre*, *doute*, *garde*, *place*, etc.).
 - idiomatic expressions, especially those which may be absent or difficult to locate in dictionaries: *montrer patte blanche*, *chercher midi à quatorze heures*, *tenir la dragée haute à quelqu'un*.
3. Derivational morphology: prefixes and suffixes. Examples:
 - the prefix *anti-*: to what type(s) of word does it attach, what types of words does it form?

- the suffix *-ment*: to what two types of words does it attach, forming what two types of words?
4. Grammatical patterns. Unless one is working with a tagged corpus (one in which the syntactic category of each word has been indicated in the text so that it can be read by the concordancer), this type of search will be limited to grammatical structures which are consistently represented by, or which regularly co-occur with, a small set of lexical items, such as the following. Examples:
 - relative pronouns *qui, que; dont; le quel* and contractions (in untagged corpora, concordances will need to be edited to include only the relevant occurrences of these forms).
 - the adverbial expression *ne ... que* (most concordancers allow searches for discontinuous features, often through the use of regular expressions).
 - subjunctive verb forms: grammatical forms having many different realizations can sometimes be located by searching for associated items. For example, a good number of subjunctive forms may be found by searching for *il faut que, à moins que, or bien que*. The converse approach, if investigating the various contexts that trigger subjunctive use, would be to search for very frequent subjunctive forms such as the third-person forms of *être* (cf. Chambers this volume).
 5. Comparison of items which are distinct but which present some similarities. The differences may be semantic or primarily syntactic. Examples:
 - homonyms. Examples: grammatical forms *leur* (pronoun vs. determiner), *lui* (dative pronoun vs. disjunctive pronoun). Similarly, referential vs. non-referential pronouns *il* and *ce*.
 - synonyms (or so-called 'translation synonyms'). Examples: *amener, apporter, ramener, rapporter, faire venir*, all related to Eng. 'to bring'; *briser, casser, rompre*, related to Eng. 'to break'; prepositions *en, dans* ('in').
 - paronyms (items of similar form but different meaning, whether related or not). Examples: *accroissement* ('increase') and *croissance* ('growth'); *jouir* and *jouer*; *avoir la chance de* + infinitive vs. *avoir des chances de* + infinitive.
 - prepositions vs. corresponding conjunctions. Examples: *après* and *après que, avant* and *avant que, depuis* and *depuis que*, etc.
 - *passé composé* and *imparfait* forms of verbs such as *vouloir, connaître, avoir*, where the aspectual distinction between the two tenses occasions differing translations in English.

6. Comparison of the same feature in different text types.
Examples: comparison of an oral and a written corpus with respect to use of subjunctive, negation, or interrogative forms. In addition, all of the corpora of written French described below allow choices of various genres such as literary, journalistic, scientific, political and administrative; patterns of lexical usage are often quite different depending on the genre examined.

4.3 Designing concordance-based learning tasks

Concordance-based activities lend themselves to a wide variety of task types. Most often, learners are asked to study the concordance data prepared by the teacher and to draw generalizations about either structure or meaning of the targeted item or items. A simple case would employ a short concordance of one lexical item or lexical phrase and ask the students to infer the meaning of the expression. This can be done rather quickly (with the use of a PowerPoint presentation or, if technology is more limited, an overhead transparency) and is an effective means of developing learners' ability to guess meaning from context. This can be a useful technique for clarifying the meaning of unfamiliar expressions in a text that the class is reading together, as the single context of the text in question may not be sufficient to give learners confidence in their guessing. Once the students have formulated a hypothesis on the basis of several contexts, the hypothesis can be verified by reference to a dictionary definition. This kind of activity is facilitated by the use of Cobb's Compleat Lexical Tutor site (see below) which allows the uploading of a text and creates a hypertext providing access to a concordance and dictionary entry for any word of the text.

More complex tasks will require learners to draw generalizations about structure or meaning where the concordance data contains multiple patterns, either structural or semantic or, more often, both. In such cases, the task can be simplified by presorting the data, or alternatively, learners can be asked to sort the concordance lines into categories which are given to them or which they will define. These tasks work well with pairs or small groups of students, who may be asked to write out their findings for comparison with their peers' results in the follow-up discussion, which is of course essential. If the concordance data is somewhat lengthy, different parts of the concordance can be assigned to different groups; in this case, the parts of the concordance may be mixed, displaying various patterns, or they may be designed so as to separate the different patterns, so that each group focuses on one of the patterns only. A pattern-discerning activity can also be developed into a more or less extended worksheet, where the learners respond to a series of questions that lead them through various observations or more complex reasoning.

Another approach to the use of a concordance is to have learners carry out some operation that will engage them in meaningful processing of the concordance lines, thus exposing them to multiple instances of the target item. A typical task of this sort is a matching exercise where the concordance lines have been split at the keyword and one side of the contexts reordered, as in Example 4 of the section below, which requires processing of clauses conjoined by *de sorte que*. This type of exercise is particularly useful for providing practice in making use of contextual clues to construct meaning from a written text. Other exercise types have as their purpose the reinforcement or assessment of the learner's acquisition of a distinction or set of patterns which have previously been the focus of a concordance-based discovery activity. These are typically of the gap-filling type, but employ only authentic concordance data. Römer (2006: 125) presents a gap-fill exercise of this type for collocations with *speak* and *talk*. More creative sentence-completion tasks, where the part of the sentence that is supplied closely resembles part of (one of) the target pattern(s), may also be utilized for this purpose. Another possible approach is to have learners produce their own exercises based on what they find in a set of corpus data.

4.4 Sample concordance-based activities

The following are examples of concordance-based activities that I have used in teaching advanced-level post-secondary American students of French. The use of concordancing is virtually unknown in the U.S. French L2 educational context, which I attribute to two primary causes: the relatively low levels of proficiency attained by most American learners given their late start at L2 education, and the prevailing literary orientation of French L2 curricula at the postsecondary level. These two factors make it difficult to incorporate significant work with corpora in the regular curriculum. However, I have created and used activities such as these to focus on particular points of lexis and grammar that pose difficulties for large numbers of students on a recurring basis. Students have responded positively to these teacher-mediated activities; occasionally, the more linguistically curious among them will ask questions about linguistic aspects not in the focus of the lesson. Activities such as these can be completed in as little as fifteen or twenty minutes of class time. Though this is more time than is required to present an explanation and examples in the traditional fashion, I believe that these activities are more effective because of the fact that each learner is required to grapple with the language samples and formulate their own generalizations.

Due to space limitations, the concordances in the following sample activities are reduced to a bare minimum. Typically, such concordances would include ten to twelve concordance lines or more, but it is of course the teacher's prerogative to choose the optimal length, depending on variables such as the

nature of the task and the number of different usages, meanings or constructions to be included.

Example 1. Single-word polysemous lexical unit: *projeter*.

Task: semantic classification.

Concordance source: Lexiquum (La Presse – see section 5 below).

1. probablement inclus en partie dans les résultats passés de l'équipe. Ceux dont nous nous servons pour **projeter** l'avenir.
2. la robe! Pour le prêt-à-porter, les prix sont plus modestes: de \$1000 à \$2500. Chez Balmain on veut **projeter** l'image de la femme moderne, élégante. On ne présente pas de robe folle, pas de robe marrante.
3. films, 13 fois par mois en moyenne. Chez les jeunes de 18 à 24 ans, le magnétoscope sert surtout à **projeter** des films loués au club vidéo.
4. Une camionnette circulant en direction ouest, sur cette rue, aurait alors happée la voiture pour la **projeter** contre l'autobus de la STCUM qui circulait dans la direction opposée.
5. ment Blake Edwards. Le meilleur moyen de surmonter son trouble, c'est encore de le décrire ou de le **projeter** sur un écran.
6. architecte, Mme Lambert a dit à La Presse, hier soir, tout l'espoir qu'elle a de voir Montréal mieux **projeter** l'image de sa personnalité sur la scène mondiale et mieux se représenter comme collectivité

Discussion: students can be asked to identify the various meanings of the verb exemplified in the data, or they can be given four meanings and asked to classify the examples and identify collocations: 'to throw' (+ locative preposition), the concrete sense 'to project' (*un film, des images*), the figurative sense 'to project' (*une image*), and another figurative extension, 'to expect'. Particularly at the advanced level, discussion can bring attention to the nature of the relationships between the various meanings, i.e. concrete and figurative, relationships which are of course generalizable to many other lexical items.

Example 2. Lexico-grammatical features: *sentir* and *se sentir*.

Task: distinguish and describe two grammatical patterns.

Instructions to student: complete the table below the concordance by writing in the appropriate column the complement that follows each occurrence of *sentir* (non-reflexive use) or *se sentir* (reflexive use). (A *complement* is the word or phrase which accompanies a verb, usually following the verb, and which completes the meaning of the verb.) Then answer the questions following the table.

Concordance source: Lexiquum (Éditions Leméac 1991-1993).

1. Je ne pensais pas qu'elle pouvait **sentir** ma colère.
2. J'ai fini par apprendre qu'elle venait ici pour ne pas **se sentir** trop seule, qu'elle aimait les livres et le silence, ...
3. Des exercices, de la marche, des séances de méditation m'aideraient à **me sentir** plus à l'aise.
4. J'aime à **sentir**, sous mes pieds, la butte de pierre et de terre à la gloire du dieu, pleine et dure : pas de creux, ...
5. Cela donne à la pièce un air de fête, de désordre joyeux où l'on pourrait presque **se sentir** bien, ...
6. ... et ma venue au monde m'a traumatisé à jamais, si bien que, depuis, je ne cesse de dépérir et de **me sentir** nu, comme un arbre à qui on aurait, d'un seul tour de main, enlevé son écorce.
7. nous couvrons de crème aux agrumes pour **sentir** le pamplemousse: cette odeur nous donne toujours un air bronzé.

Non-reflexive uses	Reflexive uses
<i>le pamplemousse</i>	<i>trop seule</i>
...	...

What generalizations can you draw about the types of complement (noun, adjective, or adverb) which follow the non-reflexive and reflexive uses? Write two original sentences illustrating this difference.

Example 3. Phrasal verb: (*locution verbale*): *mettre en* + noun

Task: discern the meaning of one or more phrasal verbs from a set of Key-Word-In-Context (KWIC) lines.

Procedure: divide the class into small groups, give each group a small set of concordance lines for two or three phrasal verbs (*mettre en œuvre, en doute, en garde, en place, en pratique, en péril, en valeur, en évidence, en lumière*). The group must propose an appropriate translation or paraphrase for each expression, note its typical complements, and write an original sentence for each. Groups present their findings to the class. Follow-up activity: gap-filling exercise based on concordance lines for all expressions studied.

Sample concordance source: Sharoff, Leeds Internet Corpus (see section 5 below).

1. qui se posent dans les salles d'urgence. Nous allons aussi **mettre en oeuvre** la Stratégie ontarienne de prévention des accidents
2. Coût: le coût comprend le coût réel actuel pour **mettre en oeuvre** la mesure d'adaptation ainsi que tous les coûts
3. " Nous sommes déterminés à **mettre en oeuvre** la prochaine génération de mesures qui assurera
4. endroit au bon moment, avec l'équipement nécessaire pour **mettre en oeuvre** le plan élaboré par l'officier.
5. un incident grave, et à déterminer le moment et la façon de **mettre en oeuvre** le programme.

Example 4. Multi-word unit: the phrasal conjunction *de sorte que*

Task: match the initial main clause and the appropriate subsequent dependent clause introduced by *de sorte que*.

Instructions to student: trouvez la bonne terminaison de chaque phrase.

Concordance source: Lexiquum (La Presse).

1. Ces modèles nous ont donné une très bonne idée de l'endroit où allait frapper Hugo,
 2. Ce que les médecins appellent l'obésité abdominale est un facteur de risque très important
 3. Les assaillantes ont pris la fuite avant l'arrivée de la police,
 4. Tout ce temps, Toronto prenait les devants en matière de design et de bureaux de ventes;
 5. Nous savions tous qu'il y avait eu une enquête en 1975 et qu'elle n'avait rien donné,
 6. ... ont toujours eu la faveur des écoliers. Si vous prévoyez une orange, entaillez-la au préalable
-
- a. ... **de sorte que** personne n'a été arrêté relativement à cette affaire.
 - b. ... **de sorte que** j'ai pensé qu'il était absolument inutile d'aller à la police quelques années plus tard.
 - c. ... **de sorte que** l'industrie montréalaise de la mode se trouve en situation périlleuse...
 - d. ... **de sorte que** l'enfant puisse la peler facilement.
 - e. ... **de sorte que** les gens qui ont un tablier c'est-à-dire une bedaine pendante sont plus sujets aux crises cardiaques
 - f. ... **de sorte que** les autorités ont pu intervenir à temps et prendre les mesures d'évacuation qui s'imposaient,

Example 5. Discourse function of *il s'agit de* ...

Task: describe the discourse function of *il s'agit de* ... on the basis of several examples in extended contexts.

Sample concordance: various websites

Instructions to student: study the following examples of the use of the verb *s'agir (de)* and answer these questions:

1. What do you notice about the subject of this verb? To what does this subject refer?

2. Try to describe what the expression *il s'agit de* is doing in each context. What other French expression could replace it? What is the relationship between what precedes the expression and what follows it? Do you think there is any significance to the fact that all three examples involve direct quotes from an apparent 'expert'?

1. "Je suis très surpris. Je crois que c'est là la réaction d'une personne qui n'a pas résisté au stress. **Il s'agit d'**un employé qui disjoncte", avait répondu M. Nicolaï, ...
2. WASHINGTON (AFP) - Des scientifiques ont utilisé en laboratoire la thérapie génique pour aider des cellules du coeur endommagées à fonctionner de nouveau normalement, selon une étude publiée lundi dans la revue de l'American Heart Association. **Il s'agit d'**"une étape vers d'éventuels essais cliniques sur des humains", a déclaré l'équipe dirigée par le Dr Walter Koch, du Jefferson Medical College, à l'Université Thomas Jefferson de Philadelphie (Pennsylvanie, nord-est).
3. MONTREAL (AP) - Un incendie d'origine criminelle a détruit dans la nuit de dimanche à lundi la bibliothèque de l'école primaire juive United Talmud Torah à Montréal: pour la police **il s'agit d'**un "crime à caractère haineux" et antisémite qui semble être revendiqué par une organisation inconnue.

Discussion: a small number of occurrences in their extended context provide the basis for reflection on the function of the expression with respect to the preceding discourse. Students can be led to notice, in addition to the non-referential nature of *il*, that what is introduced by *il s'agit de* is a new description of an event described in the preceding discourse, and that this description adds some new information which usually entails some evaluative judgment of the event or situation.

Example 6. Syntactic contrast: *il est* + Adj + *de* + Inf vs. NP *est* + Adj + *à* + Inf

Task: distinguish two similar syntactic patterns that frequently occur with *facile* / *difficile* and supply the appropriate preposition before the infinitival complement.

Discussion: this example involves study of the word grammar of *facile* and *difficile* with infinitival complements, the one construction involving non-referential *il* and the preposition *de*, the other a referential subject and the preposition *à*. This pair of confusable constructions is of course generalizable to other adjectives that occur with a non-referential subject pronoun (*possible*, *impossible*, *bon*). The sample activity is a gap-filling verification exercise which would be preceded by an activity involving study of a similar concordance without gaps, where learners would be led to observe the differences between the two syntactic patterns and their semantic correlates.

Quelle préposition faut-il? *à* ou *de*?

Concordance source: Lexiquum (various subcorpora).

1. Il a saisi l'image. Notre premier sujet de conversation est donc facile ____ trouver : l'absence de Maria, son père et tout ce qui s'ensuit. Il se détourne un peu
2. Je compte sur elle pour prévenir les femmes de l'état de santé de Tibé. La vérité est plus facile ____ entendre dans le recueillement d'un cloître et celui des révérendes
3. Il est toujours plus facile ____ suivre des grands idéaux que d'accepter d'être un peu ridicule.
4. le climat a un peu la douceur de celui de la Floride. J'espère y amener les parents. Ça ne sera pas facile ____ persuader papa qui, depuis son accident à la jambe, a perdu l'habitude des voyages.
5. selon un rapport de la DEA publié cet été. Ce puissant dérivé de la cocaïne, facile ____ se procurer et d'un prix modique - une dizaine de dollars la dose - est actuellement disponible

Example 7. Translation synonyms *en* and *dans*.

Task: examine a somewhat lengthy concordance of each preposition to observe a syntactic difference in their objects, and to note common collocations and differences in collocating behavior.

Discussion: while an edited concordance prepared and printed by the teacher can be used for this purpose, a more effective approach would be to have students work with one of the user-friendly online concordancers (see section 5), or at least a lengthy print-out for each preposition. An initial comparison of the two concordances will reveal an important syntactic generalization: while *dans* is always followed by a determiner, *en* is almost never followed by one. Next, students can be asked to look for common collocations of each (assuming the concordancer allows sorting according to the right-hand context) and to indicate, if possible, with what kinds of words each one commonly occurs. At lower levels, this will lead to noticing the use of *en* with months, years, and certain geographical names. In addition, learners can be asked, after some familiarization with the concept of set phrases, to determine if there are any set expressions among the collocations. This question should lead to the observation that *en* occurs in very many fixed expressions, whereas *dans* occurs in relatively fewer such expressions. Finally, learners should be led to observe that *en* in fact represents two different lexical items, the pronoun and the preposition. In turn, within the category of prepositional uses, the use of *en* with the present participle to form the gerund can be distinguished from the other more numerous cases.

5. Comparison of online corpora / concordancers

Because websites that combine corpora and easy-to-use concordancers offer teachers the easiest approach to creating their own concordance-based learning materials, I will describe here what I have found to be the most useful among such sites that are not password-protected. The table in Fig. 1 (see below) compares the features of the following four sites:

- Compleat Lexical Tutor:
http://www.lextutor.ca/concordancers/concord_f.html (17.05.2013)
- Lexiquum: <http://atour.iro.umontreal.ca/cgi-bin/lexiquum> (17.05.2013)
- CorpusEye: <http://corp.hum.sdu.dk/cqp.fr.html> (17.05.2013)
- Leeds Internet Corpora: <http://corpus.leeds.ac.uk/internet.html> (17.05.2013)

Designed by Tom Cobb for language teachers and their students, the Compleat Lexical Tutor (CLT) is the most user-friendly of the four sites, as shown by its use of ordinary English terms in the interface, as in the choice of 'Equals', 'Starts', 'Ends', or 'Contains' to search for variant forms of a lexical item or all lexical items with certain endings or base forms⁶. Also helpful is the ease of exporting selected lines to a Word document which even preserves the KWIC formatting. However, the small size of the CLT corpus is limiting, especially if one prefers to use more contemporary texts. In selecting corpora, one should keep in mind that learners' success at interpreting concordance lines will be facilitated when the corpus searched consists of text-types and kinds of topics which are familiar to the learners.

The Lexiquum concordancer is also very user-friendly, though it offers the least versatility, lacking both the capability of sorting KWIC lines by words in the immediate context of the keyword, and the provision of any information about collocations. However, desirable features are the large size of the corpus and the variety of text-types offered. I have found the literary corpus (Editions Leméac: novels, plays and poetry by 15 authors, published 1991-1993) to be very useful due to the relative accessibility of its language and its inclusion of personal narratives and dialogue.

⁶ It should be noted that this site contains many other software programs for researchers, teachers and learners of French and English, such as vocabulary profilers and hypertext and cloze builders. See Sevier (2004) for a review of the site.

	Compleat Lexical Tutor	Lexiquem	CorpusEye	Leeds Internet Corpora
<i>CORPORA</i>				
<i>Text types</i>	Le monde 1998 (1.1 M words), 3 19th-c. authors, 2 nonfiction	Largest: journalistic, leg. debates & commissions; small lit. & scient.	Parliamentary debates (29 M), Wikipedia 2005 (37.8 M)	General- purpose Internet corpus (open source development)
<i>Size (total)</i>	2.8 M words	229.3 M words	66.8 M words	200 M words
<i>OUTPUT</i>				
<i>No. of hits</i>	Variable, < 5,000	Variable, < 500 per corpus	Unlimited	Variable, < 1,000
<i>Sort options</i>	1-3 words L / R	None	None	1 word L / R
<i>KWIC line length</i>	Variable, <140 char.	Variable, <200 char.	Fixed, <100 char.	Variable
<i>Extended context</i>	Paragraph-length	Leg. debates & commissions only	Varies, 1-3 sentences	Yes, variable length
<i>Identified source</i>	No	Yes	Wikipedia only	Inconsistent, URL
<i>VERSATILITY</i>				
<i>Phrasal search</i>	Yes	Yes	Yes	Yes
<i>Variant forms</i>	'Equals / starts / ends / contains'	Add '+' to end of lemma	Reg. express. (. * = wildcard)	Reg. express. (. * = wildcard)
<i>Discontinuous forms</i>	No	Yes ('..' = up to 25 characters)	Yes: 'refined' or CQP ⁷ search	Yes: CQP search
<i>Collocations</i>	Yes	No	Yes, multi- option menu	Yes, multi- option menu
<i>Export lines</i>	Very easy: all, selected, or random, in 1 click	Manual only	Single lines or all, with 1 click	Manual only

Fig. 1. Comparison of selected online concordancers

Both CorpusEye (CE) and the Leeds Internet French Corpus (LIC) introduce a greater degree of complexity in their interfaces, due to their greater versatility. Namely, they both offer the option to include syntactic indicators in the search expression. LIC provides a simple means of adding syntactic labels to lexical search terms, which can be helpful, for instance, in eliminating homonymous forms of a different syntactic category. Both CE and LIC offer very sophisticated search options through the use of regular expressions and the special CQP-search language, but both concordancers can also be used for ordinary lexical searches without the use of these more advanced tools. CE's

⁷ Corpus Query Processor.

sizable Wikipedia 2005 corpus offers a text-type that is certainly familiar to most learners and covers a wide variety of topics. LIC presents considerable flexibility and ease of use, for instance in its multiple sorting options, but it also has certain disadvantages. For instance, though the concordancer can be set to search for up to 1,000 hits, the contexts generated are always the first ones encountered in the corpus, rather than a sampling throughout, which can result in a large proportion of occurrences from the same source. Both CE and LIC offer extensive information about frequencies of collocations, and LIC includes a helpful option to access concordance lines corresponding to each of the collocations listed.

Though I have described above what I believe to be the most readily usable concordancing sites, there are of course other online databases that allow similar searches. For instance, the FRANTEXT database (<http://www.frantext.fr/>, 17.05.2013), a collection of almost 3,000 literary works, predominantly of the last three centuries, now offers a very user-friendly interface with multiple search capabilities. Though conceived and developed primarily for literary and linguistic research, FRANTEXT is quite usable for pedagogical purposes, as it allows the limiting of searches by author, title, period and / or genre. Output is variable from 500 to 10,000 occurrences, with various sorting options and collocation tables. A possible drawback, however, is the limitation of access to institutional subscribers. Also of interest is the Leipzig University's Wortschatz site that includes two search engines of interest. The first (<http://corpora.informatik.uni-leipzig.de/>, 17.05.2013) offers searches of 230 corpus-based monolingual dictionaries and provides information on frequency and collocations, as well as an extensive list of complete-sentence examples with URLs of their source sites. The second (http://wortschatz.uni-leipzig.de/ws_fra/index.php, 17.05.2013) searches a corpus of 700 million words drawn from Francophone newspapers, websites and Wikipedia articles (see Kamber 2011). A disadvantage of both is that the examples are not presented in KWIC format, so they occur in no particular order, making it impossible to abstract patterns of occurrence from the data. Also, multi-word searches are not possible.

All of the foregoing are sites whose corpora consist of written texts. Online corpora of spoken French are still considerably less numerous than the latter. There is however the SACODEYL site (<http://www.um.es/sacodeyl/>, 17.05.2013), a small pedagogic corpus focusing on European youth language, which includes searchable transcriptions of about 4 hours of interviews with French young people, as well access to the video recordings. The ELICOP site (<http://bach.arts.kuleuven.be/elicop/>, 17.05.2013) offers searchable access to several large spoken corpora, but it lends itself better to research than to pedagogical purposes, due to the linguistic nature of the transcriptions.

Of course, ordinary web browsers may be used as concordancers and are quite usable for a quick check of spelling or grammatical construction. However, for more sophisticated questions of usage or collocations, the output of such search engines is often problematic in both content and format. For this reason, computational linguists have devised software to improve upon the output of web searches (Kilgariff & Grefenstette 2003), one example of which is the online WebCorp concordancer (<http://www.webcorp.org.uk/live/>, 17.05.2013).

Finally, there is also the possibility to download corpora such as the Chambers-Rostand Corpus of Journalistic French and the Chambers-Le Baron Corpus of Research Articles in French, both available from the Oxford Text Archive (<http://ota.ahds.ac.uk/>, 17.05.2013). Corpora such as these can be searched using software such as AntConc, a fairly user-friendly program available at Laurence Anthony's AntLab site (http://www.antlab.sci.waseda.ac.jp/antconc_index.html, 17.05.2013).

6. Conclusion

As suggested above, concordance-based activities can facilitate the acquisition of both lexis and grammatical forms. With respect to grammatical items, observation of data and hypothesis formulation can help to focus on the role of grammatical forms in producing differences of meaning, thus conveying the fundamental principle that communicating effectively is not primarily a matter of following rules but of making appropriate choices corresponding to alternative meanings. Shelly (1993) suggests a similar kind of approach to complex grammatical distinctions; the kind of teacher-guided, meaning-focused inductive dialogue modeled in her article would be perfectly complemented by a concordance-based observation activity.

Some proponents of DDL have pointed to the fact that a teacher's implementation of DDL requires a certain acceptance on the teacher's part of a reduced degree of control. Gilquin & Granger (2010: 366-367) describe it this way:

Although [...] DDL activities may be situated along a continuum from teacher-led to learner-led, it is nonetheless true that teachers have a less central role in DDL than in traditional teaching. They tend to have relatively little control over what happens during the lesson, which may be 'incompatible with the "minimum risk" scenario which can be found in many teaching cultures' (Boulton 2009: 93). In addition, since the computer becomes the main source of knowledge, this may be experienced as a 'loss of expertise' by the teacher (Hunston 2002: 171). DDL, therefore, requires that teachers take risks, and agree to 'let go' and let the student take pride of place in the classroom.

Though the kinds of activities I have described clearly fall toward the teacher-led end of the continuum, as students progress in their use of DDL, this aspect does need to be recognized by the teacher. At the same time, however, the use of concordance-based activities does not lessen the teacher's workload

when it comes to planning and creating activities. In the words of Lewis (1993: 149):

Far from being a soft option for teachers, this imposes considerable demands on them. Their skill lies in perceiving actual or potential confusions in their student's perception of a system of English [or other target language], and devising a task accompanied by language data focusing on the particular language problem. Teachers must choose when to intervene and provide appropriate tasks, questions and data. That is difficult enough, but then they must resist the temptation to interfere, and allow students the time to discuss, formulate, and if necessary reformulate their perceptions.

While the factor of the extra time required for DDL activities will have to be reckoned with by the teacher, if we are serious about creating effective learning conditions, we should not be satisfied with reliance on 'imposed' explanation. It is therefore my hope that more second language educators will accept the challenge to explore the potential benefits of data-driven learning.

REFERENCES

- Aston, G. (2001). Learning with corpora: an overview. In: G. Aston (ed.), *Learning with corpora* (pp. 7-45). Houston: Athelstan.
- Boulton, A. (2011). Data-driven learning: the perpetual enigma. In: S. Gozdz-Roszkowski (ed.), *Explorations across languages and corpora* (pp. 563-580). Frankfurt: Peter Lang.
- Chambers, A. (2010). What is data-driven learning? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 345-358). London: Routledge.
- Cheng, W. (2010). What can a corpus tell us about language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 319-332). London: Routledge.
- Cobb, T. (1999). Applying constructivism: a test for the learner as scientist. *Educational Technology Research & Development*, 47, 15-31.
- Collins COBUILD (2006). *Advanced learner's English dictionary*. New York: HarperCollins.
- Conklin, K. & Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 1, 72-89.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal*, 8, 4, 1-35. <http://tesl-ej.org/ej32/a1.html> (15.05.2013).
- Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 359-370). London: Routledge.
- Hulstijn, J., Hollander, M. & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327-339.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Philadelphia: Benjamins.
- Johns, T. ([1990] 1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. [Revised version: T. Johns & P. King (eds.), *Classroom concordancing*. *English Language Research Journal*, 4, 27-45].
- Johns, T. (1993). Data-driven learning: an update. *TELL&CALL*, 2, 4-10.

- Kamber, A. (2011). Contexte et sens: utilisation d'un corpus écrit dans l'enseignement/apprentissage du FLE. *Travaux neuchâtelois de linguistique*, 55, 199-218.
- Kilgariff, A., & Grefenstette, G. (eds.) (2003). Special Issue on web as Corpus. *Computational Linguistics*, 29.
- Lewis, M. (1993). *The lexical approach: the state of ELT and the way forward*. Hove, England: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the lexical approach: putting theory into practice*. Hove, England: Language Teaching Publications.
- Moudraia, O. (2001). *Lexical approach to second language learning*. Washington, D.C.: ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics. EDO-FL-01-02. <http://www.cal.org/resources/digest/0102lexical.html> (09.08.2007 from CAL database).
- Nattinger, J. & DeCarrico, J. (1989). Lexical phrases, speech acts and teaching conversation. In: P. Nation & R. Carter (eds.), *Vocabulary acquisition*. *AILA Review*, 6, 118-39.
- Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL*, 19, 3, 269-286.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In: J. Richards & R. Schmidt (eds.), *Language and communication* (pp. 191-226). London: Longman.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language*, 18, 1-28.
- 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*, 54, 121-134.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Sevier, M. (2004). Review of the compleat lexical tutor. *Teaching English as a Second Language – Electronic Journal*, 8, 3, 1-10. <http://tesl-ej.org/ej31/m2.html> (15.05.2013)
- Shelly, S. (1993). Rule reformulation at the advanced level. *French Review*, 66, 760-768.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2000). Lexical grammar. *Naujoji metodologija*, 24, 191-202. <http://donelaitis.vdu.lt/publikacijos/sinclair.pdf> (09.23.2007)
- Sockett, G. (2011). Autonomie de l'apprenant et environnement personnel d'apprentissage. *Les Langues modernes*, 105, 22-31.
- Stevens, V. (1991). Concordance-based vocabulary exercises: a viable alternative to gap-fillers. In: T. Johns & P. King (eds.), *Classroom concordancing*. *English Language Research Journal*, 4, 47-63.
- Willis, J., & Willis, D. (1989). *Collins COBUILD English course*. London: Collins COBUILD.

Learning and teaching the subjunctive in French: the contribution of corpus data

Angela CHAMBERS

University of Limerick

Plassey Technological Park, Limerick, Ireland

angela.chambers@ul.ie

Depuis le milieu des années 1980, l'exploitation de corpus linguistiques dans le cadre de l'apprentissage des langues est en croissance constante. Cette étude contribue à la recherche consacrée à cette évolution en examinant l'apport potentiel des corpus de langue française accessibles sur Internet pour l'apprentissage et l'enseignement du subjonctif. Nous examinerons dans une première partie quelques publications scientifiques sur le rôle des corpus dans l'apprentissage et l'enseignement des langues, et nous nous pencherons ensuite sur trois exemples: l'emploi du subjonctif avec *falloir* , l'emploi moins systématique du subjonctif pour exprimer une dimension subjective, et les temps plus rarement employés du subjonctif, l'imparfait et le plus-que-parfait. Le but de l'analyse est de savoir si les données consultées permettent d'enrichir le nombre limité d'exemples présentés dans les manuels et les livres de grammaire en fournissant un accès à des emplois concrets récurrents. L'étude soulève également des questions, telles que le rôle de la répétition dans l'apprentissage des langues, la présence ou l'absence de contexte, la façon dont les données peuvent être exploitées par les enseignants, et les défis posés par la variation et les erreurs.

Mots-clés: Corpus, enseignement des langues, apprentissage des langues, variation, le subjonctif en français

1. Introduction

Since the publication in the 1970s of one of the earliest corpus-based course books, *Les Orléanais ont la parole* (Biggs & Dalwood 1976), corpora have contributed to language learning and teaching in a number of ways. Dictionaries and grammars (in particular for English) are increasingly based on corpus data instead of invented examples. In addition, a substantial body of research publications are devoted to the use of corpora with learners, either by using concordances prepared by the teacher as language learning materials (see Kerr this volume), or by giving the learners direct access to corpora and providing training in the use of concordancing software. Boulton (2010) identified more than 70 studies in the context of learning outcomes, and in subsequent publications has studied other aspects of teachers' use of corpus data with learners (see, for example, Boulton 2012¹). Despite developments such as these, researchers in the area of corpus consultation by learners are well aware that the potential of corpus data to enhance the language learning and teaching environment has yet to be realised.

¹ Alex Boulton has identified more than 116 relevant publications in all (personal communication, 10 October 2012). See <http://bit.ly/STZegS> ('autres utilisateurs' > 'connexion anonyme', 16.05.2013).

This article aims to contribute to this development by investigating the potential of a number of freely and easily available small corpora of French for learning and teaching the subjunctive. The hypothesis is that the data will provide a rich source of occurrences of naturally occurring discourse to complement the necessarily limited number of examples provided in course books and grammars. After a brief account of the research context in relation to theoretical and pedagogical investigations of the role of corpora in language learning and teaching, the paper focuses on three areas: the common and uncontroversial use of the subjunctive with *falloir*²; the less clear and less systematic use of the subjunctive to express a subjective dimension, in this case with *sembler*; and the more rarely used tenses of the subjunctive, namely the imperfect and pluperfect. Analysis of these data enables us to investigate the hypothesis above, and also raises a number of theoretical and pedagogical issues in areas such as the role of repetition in language learning, the presence or absence of context in these data, their appropriateness for different levels of learner, and the challenges which the presence of variation and errors poses for the teacher.

2. Corpora and language learning: the research and teaching context

2.1 *The corpus as a source of multiple genuine examples*

Given that several renowned researchers in applied linguistics have criticised the reliance on invented examples in language learning materials (Sweet [1899] 1964: 73; Firth 1957: 24; Sinclair 1988: 6; all cited in Cook 2001: 366-367. See also McCarthy & Carter 1995), it is not surprising that one of the attractions of corpora for language teachers and researchers lies in the fact that they are a source of multiple genuine examples. The term 'genuine' is used here in preference to 'authentic', in keeping with Widdowson's (1983: 30) use of the term, reserving 'authentic' to refer to the use to which texts are put in a classroom rather than as an inherent characteristic which they possess. Indeed, in several studies of corpus consultation by learners, they appreciate this access to examples which are not invented (Bernardini 2002: 179; Chambers 2005: 120; Cheng *et al.* 2003: 181; Yoon & Hirvela 2004: 275; Wen-Ming & Hsien-Chin 2008: 42). Cobb (1997: 303) stresses the importance of multiple examples, referring to this aspect of corpus consultation as "multicontextual learning". Ellis (2002: 177) describes the list of occurrences in a concordance as "mindful repetition in an engaging communicative context", seeing this new use of repetition as a modern reinvention of behaviourist repetition as a way of mastering aspects of language use. While the underlying reasons for the use of repetition are not at all the same, this

² *Vouloir* could also have been used here.

reappearance of repetition in the context of language learning and teaching in the late twentieth and in the twenty-first century is nonetheless of considerable interest and deserves more attention in research in corpora and language learning and teaching. In relation to Cobb's use of the term "multicontextual learning", it is important to note that context is not necessarily absent in this form of learning, as the learner can be presented with examples to complement a single occurrence of the aspect of the language in a text which is being studied in class. Furthermore, a concordancer such as Wordsmith Tools (Scott 2004) gives access to the whole text by a simple click of the mouse.

This is not to suggest that the inclusion of multiple genuine examples presents no problems for the learner and teacher. Learners who participated in empirical studies where they had the opportunity to search a corpus and analyse the results, while giving generally positive reactions, nevertheless described the process of sorting and analysing large numbers of concordance lines as time-consuming, laborious and tedious (Yoon & Hirvela 2004: 274; Chambers 2005: 120). In addition, teachers using concordance data can also encounter difficulties. As we shall see, the rarer uses of the subjunctive in this study provide examples of performance errors in an area where they are well known in French, the imperfect and pluperfect subjunctive, which presents a challenge to teachers of learners at an advanced level. Furthermore, an example of taboo language, which can be a problem of varying levels of difficulty for teachers in corpus data, also occurs in these data.

Finally, while genuine examples are now commonly used, and promoted in documents such as the *Common European Framework of Reference for Languages* (Council of Europe 2001), it is important to note that this is not universally accepted in language learning research. Cook (2001) makes the case for invented examples, claiming that they have four advantages: as a means of making the lesson more personal and spontaneous; illustration; promoting noticing; and as a mnemonic device (2001: 379-380). Without disagreeing with Cook, it can be argued that corpora also have the potential to fulfil some of these functions. For example, several researchers have pointed out that the multiple examples provided in a concordance promote noticing, in the sense in which Schmidt (1993) uses the term. As Schmidt (1995: 20) observes: "The 'noticing hypothesis', states that what learners notice in input is what becomes intake for learning". For Gabrielatos (2005), the sight of a concordance, with repeated occurrences of the search word surrounded by co-text, even encourages a new way of reading texts for learners, condensed reading. In conclusion, while corpora clearly make it possible to give learners access to multiple genuine examples to support their learning, there is still a need for specific investigations of what can be learned from consulting corpus data, of the extent to which learning actually occurs, and of the reactions of

learners and teachers to their use. This article deals with the first of these issues.

2.2 Finding corpora

Teachers wishing to integrate corpora into their teaching must first identify easily and, ideally, freely available resources. They may also wish to create a corpus, although this clearly requires much more time and effort (Timmis 2010). Cappeau and Seijido (2005) provide an online directory of corpora in French, and the French governmental body, the *Délégation générale de la langue française et des langues de France* (DGLFLF) hosts a site giving details of corpora which are available online³. In the teaching context, however, a large reference corpus is unlikely to be of immediate use to a teacher interested in producing teaching materials based on corpus data but lacking training in corpus consultation and analysis. Chambers *et al.* (2011) suggest a graded approach to this problem, beginning with the use of corpora which are available online and have built-in concordancers. A resource such as Lextutor⁴, for example, provides access to substantial corpora in French, as well as a concordancer which can be used with little or no prior training (cf. Kerr this volume). At the time of writing the French corpora, 2.8 million words in all, include a spoken corpus, as well as written texts in literary, journalistic and other genres. In addition, spoken language resources suitable for learning and teaching have been developed in several languages, including French, in two projects funded by the European Commission, SACODEYL and Backbone⁵. These corpora include interviews with native speakers of French, adolescents in the case of SACODEYL and adults in the case of Backbone, annotated in ways intended to be of practical help to language teachers. These, and other resources, are examples of pedagogic corpora, which can be defined as:

A large enough and representative example of the language, spoken and written, a learner has been or is likely to be exposed to via teaching material, either in the classroom or during self-study activities. (Meunier & Gouverneur 2009: 186)

While such resources make it easy for teachers to integrate corpus data in their teaching, it is important to note that easily and freely available resources are still few in number, particularly for languages other than English. Corpora such as SACODEYL and Backbone can be classified as small corpora. While Aston (1997: 54) defined small corpora as typically consisting of approximately 20,000-200,000 words, it has since become common to find corpora of one

³ www.dglflf.culture.gouv.fr/recherche/corpus_parole/corpus_en_ligne.htm (14.05.2013). At the time of writing this database appears to have been last updated in 2006.

⁴ www.lexutor.ca (07.05.2012).

⁵ www.um.es/sacodeyl/ and www.um.es/backbone/ (14.05.2013).

million words created for language learning purposes (see, for example, Chambers & Rostand 2005; Chambers & Le Baron 2007).

A number of corpora are used for this study. The *Corpus Chambers-Rostand de français journalistique* (Chambers & Rostand 2005, henceforth referred to as CRFJ) consists of approximately one million words and includes 1,723 articles from the French newspapers *Le Monde*, *L'Humanité*, *La Dépêche du Midi*, in 2002 and 2003. The rubrics represented are editorials, culture, sport, national news, international news, and finance. As we shall see, this produces a substantial number of occurrences of the subjunctive with *falloir*. As it includes fewer examples of *sembler*, Lextutor is also used. Finally, for the rarer uses of the imperfect and pluperfect tenses of the subjunctive, where the number of examples in CRFJ is very small, a one-million word corpus of research articles in French is also used (Chambers & Le Baron 2007, henceforth referred to as CLARF), containing 159 research articles in 20 peer-reviewed, freely available online journals. The ten disciplines featured in the corpus are: media / culture, literature, linguistics and language learning, social anthropology, law, economics, sociology and social sciences, philosophy, history, and communication studies. Small corpora such as these provide fewer examples than large reference corpora and two of the three are also limited in being genre-specific. They do, however, present a number of advantages for the teacher with little or no expertise in corpus linguistics, or for the learner with direct access to the corpora. As Aston (1997: 55) notes, they are easier to use and to become familiar with than their larger counterparts, and can serve as a useful starting point for students who may later wish to move on to the analysis of larger corpora in a research context. Aston (2001: 75-76) also points out that, while such corpora are useful for finding examples in the language-learning context, they cannot be used as the basis for generalisations about language use in the context of linguistic research.

3. Investigating the subjunctive in small corpora

3.1 *Falloir*

Even a corpus of one million words will provide significant numbers of occurrences of common uses of the subjunctive. In a search for forms of the lemma *falloir*⁶ in CRFJ, for example, the 1,054 occurrences are distributed as follows: *faut* (728), *faudra* (99), *va falloir* (12), *fallait* (92), *faudrait* (66), *fallu* (51), *fallut* (2), *faillie* (4). While the majority of these involve the infinitive or a noun rather than the subjunctive (which is in itself already quite an interesting finding), there are nonetheless enough examples to provide useful teaching and learning materials. A simple search for *faut* followed by *que* or *qu'*, for example, produces 35 occurrences of the subjunctive, while additional uses of

⁶ Wordsmith Tools (Scott 2004) is the concordancer which has been used for this study.

faut with the subjunctive can be found including the negative or an inverted form (*encore faut-il que*), or where *faut* is followed by an adverbial or adverbial phrase such as *bien*, *bien entendu*, *d'abord*, or *néanmoins*. In the case of the other forms of the lemma *falloir* the occurrences of the subjunctive are as follows: *faudra* (10), *va falloir* (2), *fallait* (11), *faudrait* (15), *fallu* (6), *fallut* (0), *faillie* (0). The next step would involve finding expressions suitable for learners at the appropriate level. Examples⁷ are provided below, which could be used with learners at intermediate or advanced level:

- Un contrôleur lui dit alors qu'avec une poussette ce n'est pas possible, qu'il **faut** qu'elle attende le suivant.
- Il **faut** que cette équipe aille aux Jeux, insiste Alain Weisz.
- Demain, il **faut** qu'on revienne dans cette putain de rue.
- j'ai fini par me dire qu'il y avait un problème. Il **fallait** que je prenne du recul.
- pour la saison prochaine tout est possible. Il **faudra** pour cela que je sois à 100 %, ce qui n'est pas encore le cas
- s'il veut être champion du monde, il **faudra** qu'il se batte, et je suis sûr qu'il le fera
- Il ne suffira donc pas à Sharon de gagner. Encore **faudra-t-il** qu'il puisse gouverner.
- On craint tout, reconnaît Eric Mariottat. Mais il **va falloir** que l'Etat craigne aussi, car bientôt, il n'y aura plus de restaurateurs
- le couple franco-allemand peut lui apporter cette colonne vertébrale. Encore **faudrait-il** qu'il fasse preuve de dynamisme.

As the above occurrences illustrate, the teacher can easily find straightforward examples at different levels of difficulty which, while not fully situated in context, nonetheless make it easy in some cases for the learners to relate them to the world in which they live. For teachers and learners with direct access to the corpus, Wordsmith Tools (Scott 2004) and other concordancers, such as the freely available AntConc⁸ make it possible to access the full text with a simple click of the mouse, so that, in the last example below, for example, additional text could make it clear that *le couple franco-allemand* refers to the two governments. The unpredictability of concordance lines also presents the teacher with challenges which a course book will not normally include, such as the use of taboo language in the quotation in the third example above. Thus, rather than unquestioningly using the examples provided by the authors of course books, teachers are encouraged to reflect on the type of language use which they present to their learners and the advice which they give.

⁷ Bold print is used in the examples to draw attention to items under discussion.

⁸ <http://www.antlab.sci.waseda.ac.jp/software.html> (14.05.2013).

Paper or electronic resources created by a teacher could clearly include many more examples of the subjunctive with *falloir* than the small number listed here, thus encouraging the "noticing" (Schmidt 1993) and "condensed reading" (Gabrielatos 2005) referred to above. In addition, for less advanced learners, simple gap fill exercises could easily be created by providing the infinitive and asking the learners to include the verb in the subjunctive. At a more advanced level, learners could carry out the searches themselves, and study occurrences of more advanced uses, such as the uses of *falloir* where the subject and verb are inverted. It is also worth noting here that studying concordances can often result in what Johns (1988) terms serendipitous discoveries, i.e. uses which one notices but which were not the object of the search. It is clear from even a superficial examination of occurrences of *falloir* that there is also scope for producing examples of uses with numbers and temporal expressions, which could be of use to the teacher or learner in another context, or which could provide the learners with examples of incidental learning.

3.2 The subjunctive with *sembler*

A useful function of corpus data in the context of the subjunctive is the easy availability of examples of uses where native speakers will use the subjunctive with a particular verb, *sembler* for example, in certain situations but not in others. As we shall see, however, this can also present the teacher with a challenge. Grammars can provide a clear (though potentially questionable) explanation of such phenomena, but, understandably, they can only include a few examples. In the case of *sembler*, L'Huillier (1999: 165) provides the following explanation and examples:

Il me semble que means 'I think that', hence the indicative, while *il semble que* means 'it looks as if', i.e. there is a high degree of subjectivity, hence the subjunctive.

Ex: *Il me semble que vous avez changé d'avis*

It seems to me that you have changed your mind

But:

Il semble que vous ayez changé d'avis

It looks as if you have changed your mind.

The very small number of examples given in the corpus-based Oxford-Hachette French Dictionary follow this rule. A teacher searching for expressions with *il semble que* and *il me semble que*, will easily find a significant number of occurrences in CRFJ and in the French corpora in Lextutor, 59 in all for *il semble que* and 10 for *il me semble que* and *il nous semble que*. In expressions with *il me semble que* and *il nous semble que*, only the indicative is used. In the case of *il semble que* it quickly becomes evident, however, that the subjunctive is used in the majority of cases but not in all. The distribution in the two corpora is given in Table 1. "Same form"

indicates that it is not possible to decide if the indicative or subjunctive mood is used, as they have the same form.

Expression	CRFJ	Lextutor corpora
<i>Il semble</i> + subjunctive	9	21
<i>Il semble</i> + indicative	6	10
<i>Il semble</i> + same form	5	8
<i>Il me semble</i> + indicative	4	4
<i>Il me semble</i> + subjunctive	0	0
<i>Il me semble</i> + same form	1	0
<i>Il nous semble</i> + same form	1	0
Total	26	43

Table 1: The indicative and subjunctive moods with *sembler*

As Table 1 shows, carrying out simple corpus searches has the advantage of providing the teacher and learner with many genuine occurrences of *il semble que*, but the variation from the generally accepted rule presents a challenge.

As the number of uses of *il semble que* with the indicative (16), while fewer than the subjunctive (30), is not insignificant, it is tempting to examine the context to see if a pattern can be observed. Given the small size of the corpora (3.8 million words in all) and the limited number of occurrences, this could not produce a reliable finding, but perhaps an indication that a hypothesis could be tested in a much larger sample. At a first glance a teacher or learner might get the impression that the indicative is used when the degree of subjectivity is relatively low, as in the examples below (the source is CRFJ unless otherwise stated):

- Certains bureaux sont même restés fermés. "**Il semble qu'il** y a beaucoup de gens en Géorgie qui ne pourront pas voter [...]"
- le président uruguayen implore l'aide extérieure. **Il semble que**, pas plus que les grands organismes prêteurs, FMI, Banque Mondiale, BIB, les gouvernements argentins et uruguayens ne peuvent admettre l'échec de l'ultralibéralisme
- c'est la difficulté qu'il y a en Afrique à voir nos films. De plus en plus, **il semble que** nous les faisons pour un public européen.
- Comment les Agenais sont-ils parvenus à vaincre Toulouse? En étant tout simplement plus stupides que nous sur tous ces points de fixation pourris. **Il semble que** c'était une des solutions cette année pour nous battre.
- "Affaires de police! s'écria le roi, relevant les paroles de M. de Tréville affaires de police! et qu'en savez-vous, Monsieur? Mêlez-vous de vos mousquetaires, et ne me rompez pas la tête. **Il semble**, à vous entendre, **que**, si par malheur on arrête un mousquetaire, la France est en danger. Eh! que de bruit pour un mousquetaire ! (Lextutor: Dumas, les Trois Mousquetaires)

In the last example the number of parenthetical comments which separate *il semble que* from *en danger* might derail all but advanced learners, but it is

nevertheless interesting to show that this use of *il semble que* with the indicative is not a new phenomenon. Other examples of Dumas's use of it are also present in searches of *Les trois Mousquetaires*. Consulting the occurrences of *sembler* with the subjunctive, however, one also finds cases where the level of subjectivity is even lower, and where the writer adds an adverb after the verb to strengthen the point being made:

- Mais **il semble** bien **que** Jacques Chirac soit enclin à laisser moins d'autonomie à Jean-Pierre Raffarin qu'il n'en avait accordée à Alain Juppé
- Non seulement l'administration américaine n'a pas l'intention de renoncer devant les oppositions que sa stratégie de toute puissance provoque, mais **il semble** même **qu'**elle les ait intégrées.
- **Il semble** en effet **qu'**Henri Darré, directeur de l'usine depuis septembre 2001, ait mis un point d'honneur à se signaler par une attitude le plus souvent qualifiée de "provocatrice": fouilles, interdiction de se parler, de s'asseoir
- Mondial 2002 Pour les retransmissions des matchs de la Coupe du monde, " **il semble**, a priori, **que** nous ayons trouvé un accord avec toutes les radios nationales
- Des discussions ont en effet eu lieu et **il semble que** nous soyons suivi sur tous les points, détaille Thierry Jarlet,

Are the uses of the indicative then a case of performance errors, or an illustration of variation where usage is changing? A much larger corpus would be required to provide data for a serious study of this phenomenon. The question which arises in the context of this study, however, is how the teacher can deal with such occurrences. For less advanced learners it may be tempting to omit the examples which do not conform to the rules in order to avoid confusion. For more advanced learners, however, they provide a richer and more genuine learning experience than the sanitised presentation of rules which course books and grammars understandably present as unproblematic. Conrad (2004: 80) makes the case for including variation in language teaching, pointing out that "teaching practices that are informed by variation are the exception rather than the rule". As Sinclair (2004: 274) so aptly sums up the situation, "Variation is a nuisance".

3.3 *The imperfect and pluperfect subjunctive*

Teaching the imperfect and the pluperfect subjunctive to advanced learners presents an interesting challenge for the teacher of French. As these forms are rarely used in French, it is not surprising that they do not feature prominently in course books, grammars and language learning web sites. The web site french.about.com gives the following advice:

The imperfect subjunctive is a literary verb form used in formal writing such as literature, journalism, and history. Like all literary verb forms, you really only need to be able to recognize it, not use it. ([http:// french.about.com](http://french.about.com), 14.05.2013)

Similarly L'Huillier (1999: 159) does not include these forms in her grammar, giving the following advice: "The imperfect is rarely used and conjugations should be checked in a dictionary or book of conjugations". She also notes that the imperfect and pluperfect subjunctive "are almost never used in conversation or everyday writing, particularly as their forms tend to sound very strange and even sometimes ridiculous, particularly in the 1st and 2nd persons" (p.160). She cites two examples of the use of the tenses for comic effect (pp.147 and 161), a literary text and a text from France-Inter. As we shall see, a third example of this is found in the corpus data. Grevisse, however, challenges the accepted ideas of the comic potential of the tenses in that they sound peculiar, and of the fact that they are difficult to learn:

Sauf cas particuliers, [he has just quoted a text where the author explicitly refuses to use "embrassasse"], l'idée selon laquelle le subjonctif imparfait où plus-que-parfait serait contraire à l'euphonie est subjective: c'est seulement la rareté de ces formes qui les rend surprenantes, et *fascinassions* n'est pas moins harmonieux que *fascination*. - C'est aussi leur rareté qui les rend difficiles, et non leur difficulté qui les rend rares. (1993: 1271)

The imperfect and pluperfect subjunctive are thus neglected by grammarians, considered a subject of ridicule by literary authors and the media, and rarely used. The implication is that they therefore present no challenge to the teacher and the advanced learner of French; they can simply be ignored.

However, reading the comments of grammarians and researchers about how rare these tenses are, one finds comments to the effect that, when they are used in contemporary French, certain patterns can be observed. L'Huillier (1999: 160-161), for example, notes that the third person is "slightly more 'acceptable'" than the first and second person. Girodet (1986: 865) points out that they mainly occur with the verbs *avoir* and *être*. Hanse (1991) provides a broader range of verbs with which they are used, namely *être*, *avoir*, *pouvoir*, *devoir* and *vouloir*. It is thus tempting to revise the advice of french.about.com as follows for advanced learners: you need to be able to recognise it and to be aware of the limited number of contexts where it is still currently used. It is here that corpus data can be particularly useful, providing the teacher and learner with recent examples of usage. Given the rarity of the uses of the tense, two one-million word corpora have been consulted here, CRFJ and CLARF. These corpora are consulted as a source of examples, and not as a means of comparing language use in the two genres. As we shall see, the data confirm the observations noted above and, more importantly, they provide the teacher and learner with genuine examples of recent usage which are not found in other language teaching resources.

In the two corpora of journalistic French and research articles, two million words in all, there are 116 occurrences of the imperfect and the pluperfect subjunctive. It is important to note that in untagged corpora such as these, finding these occurrences involves searching for words with all the possible endings, and then deleting those which are not relevant (words such as

passions, connaît, russes etc.). The use of the first, second and third person is distributed as in Table 2.

	CRFJ	CLARF
First person	0	4
Second person	0	0
Third person	34	78
Total	34	82

Table 2: Occurrences of the imperfect and pluperfect subjunctive in the first, second and third person

The four occurrences of the first person do not contradict L'Huillier's comment, as they consist of two historical quotations and two occurrences in a single article which make fun of the tenses.

The occurrences in the third person, of more practical interest to the teacher and learner, are distributed as in Table 3.

	CRFJ	CLARF
Singular <i>être</i>	16	29
Singular <i>avoir</i>	9	12
Singular <i>devoir</i>	0	1
Singular <i>pouvoir</i>	0	5
Singular others	6	18
Plural <i>être</i>	3	6
Plural <i>avoir</i>	0	4
Plural others	0	3
Total	34	78

Table 3: Occurrences in the third person

The 79 occurrences of the verbs *être* and *avoir* are of particular interest here, as they provide examples of language use which are documented by researchers and grammarians as still existing, but where advanced learners will not find significant numbers of examples to guide them in course books. Examples of these and other verbs are provided in the Appendix. As the concordance lines in the Appendix illustrate, in the occurrences of *avoir*, the use of the pluperfect subjunctive as a conditional is well represented (see Grevisse 1993: 1267):

Le plus-que-parfait du subjonctif s'emploie dans la langue soignée avec la valeur du conditionnel passé [...] cet emploi est particulièrement fréquent à la troisième personne du singulier.

This is also the case with the verb *être*, where inversion of subject and verb is particularly evident (12 occurrences out of 16).

While these examples of use can be helpful to learners, they nonetheless pose once again the challenge of containing performance errors and variation. Firstly the form of the past historic and the imperfect subjunctive are sometimes confused:

des services de mise en relation en ligne disparaissaient corps et biens. **Ce fût le cas** notamment du Deuxième Monde de Canal Plus, gigantesque chat room

Secondly, in both corpora there are a number of examples of the use of the subjunctive with *après que*, possibly influenced by the fact that the subjunctive is used with *avant que*:

CRFJ

- **après que** Thierry Henry, bousculé par le défenseur Ivan Cordoba, **se mît** dans l'idée de se venger.
- **Après que** Michel Gomez **ait lancé** un souhait
- **Après que** le Président de la République **ait confirmé** le cap de la baisse d'impôt
- Il a été abattu par des gardes du corps **après que** son complice, Bouraoui Al-Ouaer, un autre Tunisien, **eût fait** exploser
- à terminer cette rencontre virile, qui aurait pu dégénérer en première mi-temps **après que** Thierry Henry, bousculé par le défenseur Ivan Cordoba, **se mît** dans
- La liaison ferroviaire avec Nîmes a ainsi été interrompue **après qu'un** TGV **ait été** frappé par la foudre, avant que la SNCF ne suspende tout

CLARF

- contaminées auraient été écoulés en Europe continentale pendant plus d'un an **après que** les autorités britanniques les **aient proscrites** de leur territoire.
- Le Parlement européen récompense ici l'action d'un leader charismatique **après que** le Comité organisateur du Prix Nobel de la Paix **ait décerné** en 1996 ce
- 'objectif des médias, qui consistait à informer le plus vite possible le public **après que survienne** un événement, fut remplacé par cette exigence inimaginable
- autres sous-cultures musicales existantes à l'époque : découverte soudaine, **après que** chaque nouvelle " mode " **ait déjà su** mettre en place une forme d'esthé
- affectés de phénomènes de survivance dus à la permanence de toute ville, même **après qu'aient disparu** les facteurs qui furent à l'origine de sa création.

When confronted by these concordance lines the teacher has to decide how to present the results of the corpus search to the learners. For learners at intermediate level it may be tempting to omit the uses of the subjunctive with *après que* as performance errors, but this could be ignoring the reality of language change (see, for example Leeman-Bouix 1994; Ball 2000).

Après que is a particularly common example of how variation can exist in the language of educated writers such as novelists and journalists. Desrosiers (2000: 14) cites De Beauvoir and Mauriac using the subjunctive, finding examples of both the indicative and the subjunctive in the latter. Summarising the explanations put forward by linguists, he cites the commonly held view that the subjunctive is used by analogy with its use with *avant que*, but also refers to Leeman-Bouix's observation that prepositions which are followed by the infinitive, such as *avant* and *après*, are used with the subjunctive when used

as conjunctions. In the context of this study it is interesting to note that he also makes the point that, if the indicative were used with *après que*, it would be in the past anterior, a tense which he describes as:

un temps que l'usage courant tend à délaisser, le trouvant vieilli, très soutenu, sinon littéraire. Sa décadence est accentuée par le fait qu'il est en principe employé dans la subordonnée en corrélation avec un passé simple dans la principale, lequel est tombé en désuétude dans la langue parlée. En insistant en faveur de l'indicatif, les grammairiens donnent ainsi l'impression de demander aux locuteurs de s'adonner à une sorte d'acharnement linguistique pour faire survivre un temps de verbe qui, aux yeux d'un bon nombre, a quelque chose de guindé aujourd'hui.

The teacher or learner consulting these easily available corpora may not have the time available to follow up the cases of variation encountered, but will at least be able to pass on to the learners the fact that they exist. The amount of data available in the corpora consulted here is small, but it nonetheless provides examples, many of them relatively recent, of occurrences of the imperfect and pluperfect subjunctive, thus introducing advanced learners to an aspect of native speaker language use which they will encounter in newspapers and literary texts. By using these in class, perhaps in support of a single example occurring in a text, the teacher can give the learners attested examples of actual language use. While there is insufficient data to allow generalisations about which verbs need to be taught, the teacher can at least be confident that the concordance lines presented to the learners represent recent occurrences of these rare forms.

4. Conclusion

This study confirms the hypothesis that the data on the subjunctive are easily available and small corpora provide a rich source of occurrences of naturally occurring discourse to complement the necessarily limited number of examples provided in course books and grammars. It is arguable that this choice of the aspect of the language to study, namely the subjunctive, provides a view of corpus consultation as overly complicated, and does not present a realistic view of the way that many aspects of the language can be presented to learners in concordances which clearly reveal patterns of use and where variation in native speaker language use is not an issue. The aim is to show, however, that the aspects of language use where there is variation pose challenges for the teacher and for the advanced learner which can be seen in a positive light. Firstly they clearly show that language use by native speakers, even educated speakers and well known writers like Alexandre Dumas, does not always conform to grammatical rules, and that a descriptive rather than a prescriptive approach can reveal aspects of use which cannot be learned from traditional resources such as course books, grammars and dictionaries. As McCarthy (1998: 78) notes, teachers and learners become slaves to the rules, forgetting that "discourse drives grammar, not the reverse". Secondly, by clearly illustrating the presence of variation, corpus data have

the potential to increase language awareness for both teachers and learners. Noticing the use of the subjunctive with *après que* in the corpus data, for example, has made me more aware of its existence in the spoken language of French native speakers. Finally, this increase in language awareness can encourage teachers to take greater control of the content of what they teach, in particular of the ways in which they present grammatical rules to advanced learners in areas where variation is not uncommon.

As mentioned above, studies of the use of corpora by learners, particularly through giving learners direct access to the corpus data, have elicited negative as well as positive responses from learners. This study has tried to mitigate such reactions by suggesting that teachers could use the data in search of additional teaching materials. Finally, while, from an applied linguistic or pedagogical point of view, the focus of this study may appear to fly in the face of more current communicative or task-based approaches, it is important to remember that it is not intended to be seen as an approach to language learning, but rather as a way in which grammatical points can be included in a class where a communicative approach is used. The corpus-based examples would thus have the advantage of providing the learners with examples of actual use from both journalistic and literary sources. The fact that the resources are freely available online and easy to consult with minimal training, particularly in the case of Lextutor, should make them attractive to teachers and learners, although empirical studies would be necessary to confirm this and to investigate the reactions of both teachers and learners to examples of variation.

REFERENCES

- Aston, G. (1997). Small and large corpora in language learning. In: B. Lewandowska-Tomaszczyk & J.P. Melia (eds.), *Practical applications in language corpora* (pp. 51-62). Lodz: Lodz University Press.
- Aston, G. (2001). Text categories and corpus users: a response to David Lee. *Language Learning & Technology*, 3, 1, 73-76.
- Ball, R. (2000). *Colloquial French grammar: a practical guide*. Oxford: Blackwell.
- Bernardini, S. (2002). Exploring new directions for discovery learning. In: B. Kettemann & G. Marko (eds.), *Teaching and learning by doing corpus analysis* (pp. 165-182). Amsterdam/New York: Rodopi.
- Biggs, P. & Dalwood, M. (1976). *Les orléanais ont la parole*. London: Longman.
- Boulton, A. (2010). Learning outcomes from corpus consultation. In: M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (eds.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching* (pp. 129-144). London: Equinox.
- Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In: A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (eds.), *Corpus-informed research and learning in ESP: issues and applications* (pp. 261-291). Amsterdam: John Benjamins.

- Cappeau, P. & Seijido, M. (2005). *Les corpus oraux en français (inventaire 2005 v.1.0)*. http://www.culture.gouv.fr/culture/dglf/recherche/corpus_parole/Inventaire.pdf (07.05.2012).
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9, 2, 111-125.
- Chambers, A., Farr, F. & O'Riordan, S. (2011). Language teachers with corpora in mind: from starting steps to walking tall. *Language Learning Journal*, 39, 1, 85-104.
- Chambers, A. & Le Baron, F. (éds.) (2007). *Le corpus Chambers-Le Baron d'articles de recherche en français*. <http://www.ota.ahds.ac.uk/headers/2527.xml> (14.05.2013).
- Chambers, A. & Rostand, S. (éds.) (2005). *Le corpus Chambers-Rostand de français journalistique*. Oxford (Oxford Text Archive). <http://www.ota.ahds.ac.uk/headers/2491.xml> (14.05.2013).
- Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System*, 31, 2, 173-186.
- Cobb, T. (1997). Is there any measurable learning from hands on concordancing? *System*, 25, 3, 301-315.
- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In: J. Sinclair (ed.), *How to use corpora in language teaching* (pp. 67-85). Amsterdam: John Benjamins.
- Cook, G. (2001). "The philosopher pulled the lower jaw of the hen". Ludicrous invented sentences in language teaching. *Applied Linguistics*, 22, 3, 366-387.
- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Desrosiers, J. (2000). *Après que et le subjonctif*. *L'Actualité langagière*, 33, 14.
- Ellis, N. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Firth, J.R. (1957). *Papers in linguistics. 1934-1951*. London: Oxford University Press [cited in Cook, 2001].
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal*, 8, 4, 1-35. <http://tesl-ej.org/ej32/a1.html> (14.05.2013).
- Girodet, J. (1986). *Pièges et difficultés de la langue française*. Paris: Bordas.
- Grevisse, M. (1936 [ed. Goosse, A. (1993)]). *Le bon usage. Grammaire française*. Paris, Louvain-la-Neuve: Duculot.
- Hanse, J. (1991). *Nouveau dictionnaire des difficultés de la langue française*. Paris: Duculot.
- L'Huillier, M. (1999). *Advanced French grammar*. Cambridge: Cambridge University Press.
- Johns, T. (1988). Whence and whither classroom concordancing? In: T. Bongaerts, P. De Haan, S. Lobbe & H. Wekker (eds.), *Computer applications in language learning* (pp. 9-27). Dordrecht: Foris.
- Leeman-Bouix, D. (1994). *Les fautes de français, existent-elles?* Paris: Seuil.
- McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & Carter, R. (1995). Spoken grammar: what is it and how can we teach it? *ELT Journal*, 49, 3, 207-218.
- Meunier, F. & Gouverneur, C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In: K. Aijmer (ed.), *Corpora and language teaching* (pp. 179-201). Amsterdam: John Benjamins.
- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206-226.

- Schmidt, R. (1995). Consciousness in foreign language learning: A tutorial on the role of attention and awareness in learning. In: R. Schmidt (ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Scott, M. (2004). *Wordsmith tools version 4.0*. Oxford: Oxford University Press.
- Sinclair, J. (1988). *New directions in English dictionaries*. Unpublished manuscript [cited in Cook, 2001].
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. In: J. Sinclair (ed.), *How to use corpora in language teaching* (pp. 271-299). Amsterdam: John Benjamins.
- Sweet, H. (1899 [ed. Mackin, R. (1964)]). *The practical study of languages: a guide for teachers and learners*. London/Oxford: Oxford University Press [cited in Cook, 2001].
- Timmis, I. (2010). Teachers telling tales: Exploring materials for teaching spoken language. In: F. Mishan & A. Chambers (eds.), *Perspectives on language learning materials development* (pp. 63-85). Oxford: Peter Lang.
- Wen-Ming, H. & Hsien-Chin L. (2008). A case study of corpus-informed online academic writing for EFL graduate students. *CALICO*, 26, 1, 28-47.
- Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford: Oxford University Press.
- Yoon, H. & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing*, 13, 257-283.

Appendix

Verbs in the imperfect and pluperfect subjunctive

Le Corpus Chambers-Rostand de français journalistique

Avoir and être

1. par la loi, pas par le dialogue social; par la norme, pas par le contrat. Or il eût été essentiel, et pour le coup véritablement novateur, de tout faire
 2. les Tchéchènes avaient eu le temps de mettre leurs menaces à exécution, le résultat eût été pire. Mais quoique les Russes aient eu l'initiative de l'attaque, on éprouve
 3. BADEN-BADEN (Allemagne) de notre envoyée spéciale Alphonse Allais eût été content : à Baden-Baden, la ville est bien à la campagne.
 4. Shéhérazade est de tous les temps et de tous les pays. Sans doute eût-elle reçu le Prix Nobel de littérature si le prix avait existé en ces temps
 5. proportion du SMIC, a-t-il précisé. Une déclaration qui laisse un goût amer. L'eût-il fait au moment où la gauche était aux affaires
-
1. Lionel Jospin a compris, avec retard, que la négociation, aussi incertaine fût-elle, était le seul chemin envisageable.
 2. Nicolas Philibert ne s'en est jamais caché, faisant valoir qu'un cinéaste - fût-il documentariste - n'est pas journaliste.
 3. lors de la saison 2001-2002, c'est peut-être trop pour un joueur de football, fût-il du calibre du milieu de terrain international d'Arsenal.
 4. il tombe, il attend longtemps la voiture de dépannage, s'en irrite fût-ce avec sa courtoisie naturelle,
 5. il n'existe pas d'oasis épargnées par les flux qui agitent la planète - fût-ce une salle de classe auvergnate. Pas plus qu'il n'existe d'êtres épargnés
 6. Il est moins lié à l'extrême droite qu'il ne le fût, et beaucoup plus à la situation internationale
 7. détermination du gouvernement à juguler l'immigration clandestine, fût-elle de transit vers l'Eden anglais, montre ici ses limites.
 8. contre l'Irak. On ne saura jamais quel a été le principal tourment moral qui fût assez intense pour conduire le scientifique britannique David Kelly
 9. L'impérialisme libéral, fût-il animé des meilleures intentions, risque d'alimenter un ressentiment
 10. " Le Tour, évidemment, ne ressemble en rien à une classique, fût-elle la plus complète de toutes. Mais Guimard n'a pas tort, car Lance Armstrong
 11. était réputée pour son sens de l'accueil de l'autre, aussi étrange et étranger fût-il.
 12. Reste qu'on est en droit d'attendre d'un chef de gouvernement, fût-ce au cœur de la trêve estivale, plus de détermination et de précision
 13. Encore faudrait-il que la République fût juste, généreuse, attentive à tous ses enfants. Elle ne l'est pas.
 14. Mais, pas de chance, le dernier de poule est un compétiteur coriace, fût-il italien. Trévise a quand même défait les London Wasps (32-12)
 15. Quoi qu'il en fût, cette découverte de l'homosexualité supposée de Molière a visiblement beaucoup
 16. Iran soit assez flexible pour changer du jour au lendemain à cause d'un prix, fût-il prestigieux comme le Nobel de la paix.

Other verbs

1. Et elle préférerait que la Maison Blanche se préoccupât un peu plus de la crise économique
2. Trois petits quarts d'heure suffirent pour que tout rentrât dans l'ordre.
3. Le simple bon sens aurait voulu que l'on s'attachât depuis vingt ans à réduire progressivement puis à arrêter cette pêche

4. ne vaudrait-il pas mieux que l'ONU accompagnât, pour " l'encadrer ", une intervention militaire considérée comme inévitable
5. a répondu à l'annonce d'une vieille dame du coin, qui souhaitait que quelqu'un mît de l'ordre dans ses archives.
6. après que Thierry Henry, bousculé par le défenseur Ivan Cordoba, se mît dans l'idée de se venger.

Corpus et dictionnaires: prise de conscience linguistique chez des apprenants LANSAD

Corinne LANDURE

Institut Universitaire de Technologie Hubert Curien
7 rue des fusillés de la Résistance, 88000 Epinal, France
corinne.landure@univ-lorraine.fr

Bilingual and monolingual dictionaries are mainly used by students as reference resources and are generally appreciated for their apparent simplicity of use and their efficiency, in particular concerning the immediacy of the results. Nevertheless, and maybe for these same reasons, they are very often underexploited by the users. With the democratization of innovative technologies, the consultation of corpora offers new possibilities and perspectives to second language learners. Numerous studies have shown the importance of corpus exploitation both at the level of the development of autonomy and of the linguistic awareness of the learners. Based on these reports, the present article describes a comparative study using a bilingual dictionary, a monolingual dictionary and a generic corpus to identify and understand certain items such as uncountable nouns in English. The experiment took place in a University Institute of Technology in France with the cooperation of students in the first year of Commercial Marketing. This work looks at the students' linguistic awareness according to the use of these three resources. The results of the study invite critical reflection on the means to be developed for effective practices of observation and intake of language by ESP learners.

Keywords: Data-driven learning, dictionaries, linguistic awareness, language events, corpus

1. Introduction

Quels que soient les formes et les moyens d'apprentissage utilisés, ceux-ci doivent correspondre aux besoins et aux spécificités des apprenants afin de faciliter leur apprentissage. Pour autant "la décision d'apprendre ne peut être prise que par l'apprenant" (Bogaards 1988: 100). Mais que signifie 'apprendre' pour un apprenant? Le fait de se poser la question n'est pas neutre. La forme de l'apprentissage et les moyens mis à disposition reposent bien souvent sur des choix faits par l'enseignant ou dépend de directives institutionnelles qui peuvent ne pas convenir à tous les étudiants pour différentes raisons. Pour l'apprentissage du lexique par exemple, l'apprenant dispose de manuels spécialisés qui concentrent le vocabulaire à acquérir sous forme de listes thématiques. Il peut également avoir recours à la lecture de divers ouvrages où les unités lexicales sont diluées dans le texte (Tournier 2004). Pourtant, comme le fait remarquer Cobb (1999), si ces manières d'apprendre présentent des avantages, elles ne sont pas garantes d'un apprentissage efficace. Dans le premier cas, l'apprentissage peut aboutir à des connaissances incomplètes et dans le second, l'acquisition par la lecture s'avère lente. Bien souvent, dans ces deux situations, les étudiants s'aident d'un dictionnaire bilingue, même parfois d'un dictionnaire unilingue, et comme le souligne Galisson (1991: 17)

"[les dictionnaires bilingues] se portent plutôt bien, faisant l'objet d'une prédilection des apprenants". Surcouf (2010: sn¹), quant à lui, remarque:

Qu'il soit plébiscité par les apprenants, et plus ou moins discrédité par les enseignants, le dictionnaire bilingue, en raison même de sa popularité, modèle incontestablement l'apprentissage de la langue étrangère.

Pour leur part, Schmitt *et al.* (1997) notent que les dictionnaires bilingues sont parmi les premiers outils utilisés par les étudiants quand ils lisent dans une langue étrangère. Le fait est que le dictionnaire bilingue répond de façon satisfaisante à un besoin ponctuel, notamment pour la recherche de la signification d'une unité lexicale, et les étudiants apprécient autant l'apparente facilité de son utilisation que l'immédiateté du résultat ainsi obtenu. Pourtant, comme le font remarquer Schmitt *et al.* (1997) dans leur étude, bien que le dictionnaire soit le premier choix de consultation, il n'est pas cité comme une ressource efficace d'apprentissage. Ce constat corrobore les résultats d'une enquête réalisée auprès de 160 de nos étudiants: 90% utilisent le dictionnaire uniquement pour la recherche d'une traduction ou d'une définition et 72% déclarent n'avoir jamais reçu de formation théorique et pratique concernant l'utilisation d'un dictionnaire bilingue durant leur scolarité (ce taux atteint 89% pour un dictionnaire unilingue). Le dictionnaire serait avant tout utilisé comme ressource de référence plutôt que ressource d'apprentissage pour preuve le faible taux de rétention lors de l'utilisation de dictionnaire à des fins d'apprentissage (Landure 2011). C'est aussi un support 'historique' (ce qu'utilisaient les enseignants eux-mêmes ainsi que leurs enseignants...) transmis de génération en génération. Dans ce cas, quelles ressources l'apprenant a-t-il à disposition pour repérer et traiter efficacement les phénomènes linguistiques?

Cette étude a un double objectif: tout d'abord évaluer la prise de conscience linguistique des apprenants dans leur façon d'appréhender les faits de langue (dans notre étude il s'agit des noms indénombrables) selon qu'ils utilisent un dictionnaire bilingue, unilingue ou un corpus; ensuite déterminer dans quelle mesure une formation préalable à l'utilisation de ces trois ressources influe sur la prise de conscience linguistique. Dans un premier temps, nous définirons ce que sont les indénombrables puis nous présenterons l'approche sur corpus avant de décrire nos deux expérimentations.

Nous supposons qu'un apprentissage efficace se reconnaît par une connaissance des éléments lexicaux, non pas uniquement en tant qu'unité de sens isolé mais en tant qu'unité de sens aux spécificités propres, régissant l'unité lexicale ainsi que sa dimension morphosyntaxique. Certaines distinctions linguistiques entre langue source et langue seconde (L2) peuvent représenter des écueils lors de l'apprentissage; la traduction littérale des items permet rarement d'en cerner les spécificités linguistiques, même si le mot fait

¹ Sans numérotation.

sens dans le contexte. Il en est ainsi des indénombrables en anglais qui, s'ils ne comportent pas de difficultés sémantiques particulières, constituent des obstacles difficilement repérables par l'apprenant, par exemple: *luggage* se traduit par un pluriel en français, ne prend pas de 's' à la fin, ne s'utilise jamais avec un article indéfini et le verbe qui le suit est au singulier. Ces difficultés concernent l'item dans sa morphosyntaxe. Le Priault, auteur de la *Grammaire progressive de l'anglais*, définit ainsi les indénombrables (1996: 8):

La distinction dénombrable/indénombrable a des conséquences grammaticales importantes, en particulier sur l'accord verbal, ainsi que sur le phénomène de détermination (les articles et la quantité).

Il souligne le caractère complexe des indénombrables pour des apprenants français en l'occurrence et conclut "que chaque langue découpe différemment la réalité" (1996: 11).

2. 'Data-driven learning' et apprentissage expérientiel

Dans le cadre de notre travail de recherche, nous procédons à une étude comparative impliquant la consultation de trois ressources pour évaluer la prise de conscience linguistique d'apprenants d'une L2: un dictionnaire bilingue anglais-français, *Le Grand Robert & Collins* (GR&C), un dictionnaire unilingue anglais, *Oxford Advanced Learner's Dictionary of Current English* (OALD), et un corpus générique, le British National Corpus (BNC). Si les deux premières ressources sont familières aux étudiants, l'exploitation de données de corpus est une nouveauté pour nos apprenants. En effet, la forme d'apprentissage sur corpus (ASC) baptisée par Johns ([1990] 1991) '*data-driven learning*' est peu connue ou reconnue dans les sphères enseignantes traditionnelles (Boulton 2008b; Lapaire 2009). Pourtant, de nombreuses études recensées par Boulton (2008a) soulignent l'importance d'une telle approche pour améliorer la prise de conscience linguistique des apprenants (cf. O'Sullivan 2007; voir aussi Chambers ce volume). L'ASC offre aux apprenants la possibilité de gérer et de déterminer leur propre apprentissage, d'appréhender et d'organiser de manière inductive leurs connaissances (Boulton 2008a). Aston (2001) souligne également l'importance d'une telle approche pour développer la prise de conscience des apprenants et pour améliorer leur compétence linguistique en découvrant par eux-mêmes des faits de langue qu'ils n'avaient pas forcément remarqués avant cela. Boulton (2010: sn) fait remarquer:

[qu'] avec des outils appropriés, l'approche permet l'étude des usages habituels et probables de façon empirique et non à partir de simples intuitions ou de quelques exemples isolés.

Afin d'introduire la notion de prise de conscience linguistique chez des apprenants LANSAD ('LANGues pour Spécialistes d'Autres Disciplines') et d'en identifier les postures, notre étude s'appuie sur le célèbre modèle d'apprentissage expérientiel de Kolb (1984). Bien qu'il existe d'autres cadres

de référence traitant ce domaine, le modèle de Kolb, largement employé dans la formation aux adultes, se trouve être le mieux adapté à notre public composé d'étudiants âgés de 19 à 21 ans, en première année de Diplôme Universitaire de Technologie (DUT) en Techniques de Commercialisation. Kolb, bien que ne traitant pas spécifiquement l'apprentissage des langues, considère l'apprentissage expérientiel en quatre étapes: l'expérience concrète, l'observation réfléchie, la conceptualisation abstraite et l'expérimentation. Ces différents stades mettent en opposition d'une part le concret et l'abstrait au niveau de la 'préhension' c'est-à-dire de l'action, et d'autre part le réflexif et l'actif au niveau de la 'transformation', c'est-à-dire de la réflexion (Chevrier & Charbonneau 2000; Wittorski 2001). Les associations de ces formules aboutissent à quatre différents styles d'apprentissage (cf. Boulton 2010) sur lesquels nous ne nous attarderons pas dans cette étude. Notre expérience privilégie plus particulièrement l'étude et l'analyse comparative des procédés et attitudes cognitives de nos apprenants à travers les quatre étapes de l'apprentissage expérientiel telles que définies par Kolb mais transposées et adaptées par nos soins dans le cadre de la consultation de dictionnaires bilingue, unilingue et d'un corpus pour l'appréhension de faits de langue. En rapprochant les quatre étapes de l'apprentissage expérientiel de Kolb (tableau 1 ci-dessous), on s'aperçoit qu'elles correspondent aux principales notions de l'apprentissage sur corpus, telles que décrites par Johns (1986: 159):

By concentrating and making it easy to compare the contexts within which a particular item occurs, [a concordance] organizes data in a way that encourages and facilitates inference and generalization.

Au mode exploratoire, c'est-à-dire l'expérience concrète où l'apprenant est directement acteur voire chercheur, correspond l'exploration des données sur corpus avec l'utilisation d'un concordancier par l'apprenant. L'observation réfléchie des données se rapporte à l'observation et à la comparaison des données en contexte et co-texte qui aboutissent à une réflexion que l'on peut éventuellement considérer comme étant une des phases essentielles de la prise de conscience linguistique de l'apprenant. Cette réflexion fondée sur l'observation va permettre de mettre en place une conceptualisation abstraite (repères, notions, hypothèses) et ainsi contribuer à une réflexion inductive qui participe également à une autre phase majeure de la prise de conscience langagière. Enfin, le mode vérificateur s'inscrivant dans l'expérimentation conduit à une généralisation des concepts en les confrontant à des applications actives.

Apprentissage expérientiel (Kolb 1984)	Data-driven learning (Johns 1986)
Mode exploratoire expérience concrète	Exploration corpus & 'hands-on' (<i>concordance</i>)
Mode réfléchi observation réfléchie	Observation & comparaison (<i>organize data, compare contexts</i>)
Mode abstraitif conceptualisation abstraite	Induction (<i>inference</i>)
Mode vérificateur expérimentation	Généralisation & application (<i>generalization</i>)

Tableau 1. Comparaison entre apprentissage expérientiel et 'data-driven learning'

3. Méthode

Comme nous l'avons déjà mentionné, l'objectif principal de cette étude est de situer et d'évaluer la prise de conscience linguistique d'apprenants d'une L2 selon qu'ils utilisent un dictionnaire bilingue, unilingue ou un corpus générique dans leur capacité à repérer et à traiter un fait de langue. Pour ce faire, nous avons procédé à deux expériences sur le terrain à une année d'intervalle. La première consiste en la participation de 78 étudiants en première année de DUT de Techniques de Commercialisation qui n'ont reçu, à dessein, qu'une formation théorique brève. Cette présentation d'une quinzaine de minutes, sous forme de diapositives, est assurée par l'enseignante pour chacune des quatre classes composées de vingt étudiants environ et porte sur la manière d'utiliser des dictionnaires et de consulter le corpus. La deuxième expérience concerne 82 nouveaux étudiants en première année de DUT de Techniques de Commercialisation qui ont bénéficié d'une formation préalable de quelques heures sur l'utilisation des dictionnaires bilingue, unilingue et d'un corpus. Le niveau d'anglais de ces deux groupes est moyen et correspond au niveau B1 du Cadre Européen Commun de Référence pour les Langues (CECRL – Conseil de l'Europe 2001), comme l'attestent les résultats à un test informel du TOEIC administré au début du semestre pour chaque année respective. Ces 160 étudiants sont des francophones natifs, à l'exception de quatre Chinois et d'un Marocain, et ont un parcours scolaire très différent: ils sont titulaires soit d'un Bac professionnel, soit d'un Bac technique ou encore général ce qui peut expliquer l'hétérogénéité de leur niveau en anglais. D'origine modeste, ces étudiants n'ont jamais séjourné plus de deux semaines dans un pays anglophone et sont tous domiciliés dans la même région.

Ne pouvant avoir accès à la salle informatique, les étudiants travaillent sur supports papier comportant les consignes qui se rapportent aux tâches à effectuer à partir des définitions d'items extraites du GR&C bilingue pour le premier couple d'items (*information / inquiry*), du dictionnaire unilingue OALD pour le deuxième couple (*advice / suggestion*), et du BNC dans lequel nous avons sélectionné quinze phrases pour l'item *luggage* et quinze autres phrases pour *suitcase*. Ces mots ont été choisis par nos soins pour plusieurs raisons. Tout d'abord, parce qu'ils avaient été vus précédemment en cours

dans divers contextes; ils n'étaient donc pas totalement inconnus de nos étudiants au niveau sémantique, ce qui n'est pas négligeable puisque nous voulions mettre l'accent sur la particularité morphosyntaxique d'une unité lexicale composant une paire plutôt que sur son sens. La raison de ce choix s'explique par un constat. Nous avons remarqué lors des cours précédents ces expériences que les étudiants s'attachaient à retenir davantage le sens des mots (cf. Kerr ce volume) au détriment de ses dimensions morphosyntaxiques malgré les efforts et l'insistance de l'enseignante. D'autre part, les mots composant chaque paire ont un sens proche bien qu'ils ne soient pas synonymes mais surtout l'un d'entre eux est indénombrable, ce qui va nous permettre, plus particulièrement dans l'exercice lacunaire, d'évaluer la prise de conscience des étudiants sur ce fait de langue et de la dimension morphosyntaxique de l'unité lexicale. L'impact du sens dans l'exercice lacunaire est très réduit puisque cet exercice a été conçu de telle sorte qu'il ne prend principalement en compte que l'unité lexicale dans sa fonction morphosyntaxique et non sémantique.

Afin de pouvoir déterminer les attitudes et comportements des étudiants lors de leurs diverses consultations, nous leur avons demandé de procéder à quatre tâches. La première activité consiste à consulter un couple de mots dans une ressource de référence et à relever le temps passé pour chaque consultation, grâce à une horloge numérique provenant d'un site Internet et projetée au tableau (<http://www.horlogeparlante.com>). La deuxième activité concerne la rédaction d'une phrase dans la langue cible intégrant l'item consulté afin de savoir si l'étudiant a repéré dans la ressource consultée le fait de langue et la dimension morphosyntaxique du mot recherché. La troisième porte sur la traduction dans la langue source de leur production (annexes 1 & 2) qui n'a d'autre intention que d'évaluer les interférences possibles avec le français et ainsi de confirmer ou d'infirmer si l'apprenant en a saisi et assimilé les spécificités dans la langue cible mais également comment il les traduit dans la langue source. Enfin, la quatrième activité consiste à remplir un exercice à trous pour chaque couple d'items; cet exercice a été conçu par nos soins à partir de lignes de concordance provenant du BNC (annexe 3) et de telle sorte que les choix à opérer par les apprenants portent davantage sur une réflexion ayant trait à la dimension morphosyntaxique de l'unité lexicale plutôt que sur sa fonction sémantique. Chacune des trois activités (consultation, rédaction et traduction) va permettre d'apprécier les cheminements et comportements des apprenants dans le traitement des données provenant des trois ressources dans la phase active d'exploration puis d'observation et enfin d'induction aboutissant à une production écrite. Ainsi, notre étude va s'attacher à évaluer l'item dans sa spécificité propre en tant que fait de langue, sa traduction dans le contexte et ses répercussions sur le co-texte selon la ressource consultée. La phase vérificatoire consiste en un exercice à trous pour évaluer la prise en compte du co-texte par les

apprenants lors du remplissage et ces résultats seront à rapprocher de ceux provenant de l'activité de production.

4. Résultats

4.1 Première étude comparative de la prise de conscience d'un fait de langue selon la ressource consultée

Le tableau 2 (voir ci-dessous) nous indique que tous les étudiants ont consulté les différentes ressources puisqu'ils ont tous indiqué sur une grille le temps passé pour la consultation de chaque item. Nous remarquons qu'ils ont passé plus de temps sur les quinze lignes de concordance provenant du corpus, deux à trois fois moins de temps pour le dictionnaire unilingue et encore moitié moins pour le dictionnaire bilingue. Il est intéressant de noter la différence de participation dans la phase productive; elle s'élève à 76% après consultation sur corpus, arrive à 95% après consultation du dictionnaire unilingue et atteint les 100% après consultation du GR&C bilingue. Concernant ces activités, aucune limite de temps n'a été fixée, et tous les étudiants ont déclaré avoir terminé ces trois tâches au bout d'une heure. Ceux qui ne sont pas parvenus à tout faire (24%) n'ont pas évoqué un manque de temps mais plutôt une incompréhension ou une incapacité à exécuter l'activité demandée. Ils mentionnent un problème de traduction du mot, d'incompréhension de l'ensemble des occurrences ou d'un manque d'inspiration.

Ressources & items	Consultation			Rédaction	Traduction
	nombre %	temps moyen par apprenant	écart- type	nombre%	nombre%
Dictionnaire bilingue					
<i>information</i>	100%	34 sec	07 sec	100%	100%
<i>inquiry</i>	100%	30 sec	08 sec	100%	100%
Dictionnaire unilingue					
<i>advice</i>	100%	58 sec	18 sec	96%	94%
<i>suggestion</i>	100%	50 sec	16 sec	94%	93%
Corpus					
<i>luggage</i>	100%	3 mn	1.58 mn	76%	75%
<i>suitcase</i>	100%	2 mn	1.30 mn	76%	76%

Tableau 2. Activités du 1^{er} groupe de 78 étudiants sans formation

Ressources utilisées	Items consultés	0 réponse	Contexte approprié	Traduction appropriée	Forme de l'item appr	Co-texte approprié
Dic biling	<i>information</i>	0%	100%	86%	81%	83%
Dic uniling	<i>advice</i>	4%	76%	28%	95%	56%
Corpus	<i>luggage</i>	24%	88%	54%	88%	76%

Tableau 3. Analyse des résultats des activités du 1^{er} groupe

Le tableau 3 renseigne quant aux indénombrables consultés respectivement dans les différentes ressources. L'unité lexicale *information* a été traitée complètement par les étudiants lors des activités de consultation, de rédaction et de traduction. L'item dans le contexte n'a pas posé de problème puisqu'en plus d'être un mot transparent, les étudiants ont pu s'assurer du sens dans le dictionnaire bilingue.

La traduction a été correcte (ou acceptable) dans 86% des cas, ce qui peut paraître peu mais qui s'explique par le fait que les étudiants l'ont traduit souvent par un singulier en français alors qu'il s'agissait d'un pluriel dans le contexte. La forme de l'unité lexicale est correcte à hauteur de 81%, la différence provient essentiellement de l'ajout d'un 's'. On relève peu d'erreurs au niveau du co-texte, celles-ci proviennent des quantifieurs (15%) ou de l'accord de la forme verbale (2%). Pour l'item *advice* consulté dans le OALD, 4% des participants n'ont pas exécuté les tâches demandées. 76% l'ont intégré dans un contexte approprié et seulement 28% ont donné une traduction adéquate; on relève un important nombre de traductions au singulier à la place d'un pluriel dans le contexte, beaucoup d'approximations ('opinion', 'avis'), de nombreux contre-sens ('ordre', 'ordonnance', 'notice', 'information') ainsi qu'un certain nombre de changements de classe grammaticale, par exemple: *to advice* (*sic*). La forme de l'item a été plutôt bien respectée, on ne relève que 5% de 's' à la fin de *advice*. Le co-texte a posé un certain nombre de problèmes, on note entre autres la présence du déterminant *an* et la confusion entre le verbe et le substantif (*I / you advice* (*sic*) et *advice*). Concernant l'item *luggage* que les étudiants devaient consulter dans les quinze lignes de concordance provenant du BNC, on relève une défection de presque un quart des étudiants concernant les activités de rédaction et de traduction. Cela pourrait s'expliquer par notre choix intentionnel d'avoir fait peu de formation sur l'exploitation de corpus, par le manque de motivation des apprenants ou de leur incompréhension face cette nouvelle ressource. A ce stade, nous ne pouvons qu'émettre des hypothèses. Le mot *luggage* dans le contexte (88%) n'a pas posé trop de problèmes aux étudiants qui ont plus souffert au niveau de la traduction. La plupart l'ont traduit par 'valise', 'objet', 'affaires', d'autres ont pensé qu'il pouvait s'agir d'une marque de voiture, d'une commode. 12% ont ajouté un 's' à l'item et 76% ont respecté le co-texte, les erreurs se situant au niveau des quantifieurs (57%) et de l'accord de la forme verbale (43%).

Ces résultats confirment en partie l'efficacité du dictionnaire bilingue pour la traduction de l'item et son sens dans le contexte. En revanche, la forme de l'item a été moins prise en compte au niveau de sa spécificité linguistique par rapport aux résultats obtenus avec les autres ressources. Les étudiants se focalisent certainement plus sur le sens du mot et sont moins attentifs à ces spécificités. Pourtant, les répercussions du fait de langue sur le co-texte ont été bien traitées; afin d'être en mesure de confirmer ou d'infirmer ces résultats, nous envisageons de les rapprocher de ceux provenant de l'exercice à trous portant sur le co-texte. Si le dictionnaire unilingue a généré de gros problèmes de traduction, nous notons un résultat mitigé pour le co-texte que les étudiants ont, semble-t-il, ignoré alors que la forme de l'item a été dans l'ensemble bien traitée. Mise à part la traduction, les résultats provenant de la consultation sur corpus sont plutôt satisfaisants, notamment au niveau de la forme de l'item en tant que fait de langue et des répercussions sur son co-texte. En revanche, on note une déstabilisation conséquente des étudiants puisqu'un quart d'entre eux n'ont pas participé aux activités de production ce que nous allons tenter de comprendre et d'expliquer dans la partie suivante.

La particularité de l'exercice à trous conçu à partir du BNC, est de confirmer ou d'infirmer les résultats obtenus lors de la phase de production écrite. Cet exercice concerne plus particulièrement les répercussions du fait de langue dans le co-texte. Nous estimons que s'il y a bien prise de conscience linguistique, les manières de procéder de l'apprenant se révéleront dans la façon de remplir les blancs en fonction du co-texte. Cette prise de conscience a pu commencer lors de la phase d'exploration, se poursuivre lors de l'observation pour aboutir à une réflexion inductive. Les résultats de ce cheminement vers une prise de conscience linguistique se traduisent dans la manière d'aborder et d'appréhender une activité, dans notre cas dans la façon de traiter un fait de langue dans son co-texte.

L'exercice lacunaire (annexe 3) comporte 24 phrases extraites du BNC. Les huit premières doivent être complétées avec *information* et / ou *inquiry*, les huit suivantes avec *advice* et / ou *suggestion* et les huit dernières avec *luggage* et / ou *suitcase* dans leur forme singulier ou pluriel. Les étudiants disposent de 45 minutes pour cette activité et aucun document n'est autorisé. Le tableau 4 présente les ressources utilisées pour chaque couple d'items. On note que les résultats provenant de la consultation des dictionnaires pour l'appréhension d'un fait de langue sur son co-texte atteignent respectivement 41% pour le dictionnaire bilingue et 40% pour l'OALD alors que la consultation sur corpus approche les 34%. Pourtant lors de l'activité de production, les résultats du co-texte suite à la consultation du dictionnaire bilingue atteignaient le double (83%), 56% après consultation du dictionnaire unilingue et 76% après la consultation sur corpus.

Ressources	items	approprié au co-texte 1 ^{er} groupe
Dictionnaire bilingue	<i>information and / or inquiry</i>	41%
Dictionnaire unilingue	<i>advice and / or suggestion</i>	40%
Corpus	<i>luggage and / or suitcase</i>	34%

Tableau 4. Résultats de l'exercice lacunaire pour le co-texte

Notre hypothèse de départ supposait qu'un apprentissage efficace allait se reconnaître par une connaissance des items non pas uniquement en tant qu'unité de sens isolé mais en tant qu'unité de sens aux spécificités propres, régissant l'unité lexicale ainsi que son entourage proche. Notre étude partant de cette hypothèse consistait à comparer trois ressources de référence pour évaluer la prise de conscience linguistique d'apprenants d'une L2. Les résultats sont surprenants voire décevants si nous nous fondons sur des études antérieures prêtant à l'ASC des qualités notamment pour l'amélioration de la prise de conscience linguistique et le développement de l'autonomie des apprenants. En parlant de l'autonomie, Holec (1979: 3) souligne que:

Cette capacité n'est pas innée, elle doit s'acquérir, soit de manière 'naturelle', soit (c'est le cas le plus fréquent) par un apprentissage formel, c'est-à-dire systématique et réfléchi.

Et si justement, comme l'autonomie, la prise de conscience linguistique n'était pas innée et nécessitait une formation? C'est ce que nous avons voulu vérifier avec la deuxième expérience se passant dans les mêmes conditions à un an d'intervalle. Les 82 participants sont des étudiants en première année de DUT en Techniques de Commercialisation. La seule différence réside dans une formation de deux heures (au lieu de 15 minutes pour la première expérience) dont ils ont bénéficié avant cette expérience pour l'utilisation des dictionnaires et de la consultation de corpus, et ciblant, bien entendu, d'autres phénomènes langagiers que les indénombrables.

4.2 Deuxième étude comparative de la prise de conscience d'un fait de langue selon la ressource consultée

Si nous comparons les résultats du tableau 5 aux résultats obtenus une année auparavant (tableau 2) avec des étudiants n'ayant bénéficié d'aucune formation préalable, on remarque que le temps moyen par étudiant pour les consultations du dictionnaire bilingue et unilingue varie entre 1 et 8 secondes supplémentaires, en revanche le temps passé pour la consultation sur corpus a triplé. On note également une augmentation sensible de participation aux activités de production après la consultation sur corpus.

ressources		consultation		rédaction	traduction
& items	nombre %	temps moyen / apprenant	écart-type	nombre%	nombre%
Dictionnaire bilingue					
<i>information</i>	100%	35 sec	05 sec	100%	100%
<i>inquiry</i>	100%	32 sec	06 sec	100%	100%
Dictionnaire unilingue					
<i>advice</i>	100%	65 sec	13 sec	100%	100%
<i>suggestion</i>	100%	58 sec	15 sec	98%	98%
Corpus					
<i>luggage</i>	100%	8 mn	1.02 mn	91%	89%
<i>suitcase</i>	100%	7 mn	0.50 mn	95%	95%

Tableau 5. Activités du 2^e groupe de 82 étudiants avec formation

Ressources utilisées	Items consultés	0 réponse	Contexte approprié	Traduction appropriée	Forme de l'item appr	co-texte approprié
Dic biling	<i>information</i>	0%	100%	96%	91%	85%
Dic uniling	<i>advice</i>	1%	72%	30%	90%	66%
Corpus	<i>luggage</i>	8%	91%	65%	95%	82%

Tableau 6. Analyse des résultats des activités du 2^e groupe

Le tableau 6 montre une augmentation des résultats dans toutes les rubriques, ce que nous pourrions imputer à la formation dont les étudiants ont bénéficié. En revanche, il est intéressant de comparer ces résultats avec ceux de l'exercice à trous (tableau 7). On remarque à nouveau une différence non négligeable entre l'activité de production et l'exercice à trous au niveau des résultats du co-texte qui pourrait s'expliquer par la teneur de l'exercice lui-même, trop complexe peut-être, ou au contraire par une minimalisation de la production écrite qui intègre l'item dans une phrase trop simple ou qui s'inspire d'une phrase trouvée dans les ressources consultées. En revanche, le tableau 7 indique une augmentation sensible de résultats appropriés au co-texte et cela quelles que soient les ressources utilisées, avec un résultat presque doublé pour l'utilisation de corpus.

Ressources	items	approprié au co-texte (1 ^{er} groupe)	approprié au co-texte (2 ^e groupe)
Dictionnaire bilingue	<i>information and / or inquiry</i>	41%	52%
Dictionnaire unilingue	<i>advice and / or suggestion</i>	40%	48%
Corpus	<i>luggage and / or suitcase</i>	34%	63%

Tableau 7. Etude comparative de l'exercice à trous

A ce stade, peut-on avancer que la formation pour l'utilisation de ressources a augmenté la capacité de prise de conscience linguistique d'apprenants d'une L2? Peut-on avancer que la consultation de corpus améliore efficacement la

prise de conscience et donne les moyens à l'apprenant de développer sa sensibilité linguistique en fonction de ses besoins? Les résultats montrent que la formation préalable portant sur l'utilisation des corpus est plus efficace pour une meilleure prise de conscience linguistique, alors que la formation préalable portant sur l'utilisation des dictionnaires bilingue et unilingue n'a pas vraiment amélioré la prise de conscience linguistique des apprenants.

5. Discussion

Afin de répondre à ces deux questions, nous nous appuyons sur les travaux de Wittorski (2001) qui, s'inspirant des recherches de Foerster (1960), rapporte qu'un individu se fabrique une représentation de la situation en fonction de ses expériences antérieures, et que par conséquent les modèles d'action mobilisés dépendront de ses expériences personnelles. Puis, Wittorski se fonde sur les travaux de Serre (1992) pour illustrer les processus qui vont guider l'action et que nous avons représentés schématiquement (tableau 8):

Modèle d'action par rapport aux expériences passées	Hypothèses émises en fonction de la situation	Action engagée suivant les hypothèses	Vérification de l'(in)efficacité de l'action
---	---	---------------------------------------	--

Tableau 8. Schéma inspiré des travaux de Serre pour les processus de mobilisation

Ces processus s'appliquent à notre deuxième expérience pour laquelle un modèle d'action ou des modèles d'actions ont été mis en place par les apprenants et qui découlent directement des expériences passées; dans notre cas, il s'agit de la formation dont ils ont bénéficié dans le cadre de la consultation des ressources. Des hypothèses ont été alors émises en fonction de la situation; en ce qui nous concerne, des hypothèses ont été exprimées par rapport à des situations de consultation intéressant diverses ressources de référence. Ces dernières ont contribué à engager une action en fonction des expériences antérieures et postulats émis en amont; dans notre cas, cela consiste en la production écrite faisant suite aux consultations des trois ressources. Enfin, l'exercice lacunaire a permis de vérifier l'efficacité ou l'inefficacité de la consultation des trois ressources utilisées pour la compréhension des faits de langue et ses répercussions sur le co-texte. Les résultats de l'exercice lacunaire montrent qu'une formation préalable ciblant l'utilisation de dictionnaires bilingue et unilingue n'a que peu d'impact sur la prise de conscience alors qu'une formation impliquant l'utilisation des corpus obtient une meilleure prise de conscience linguistique. Les apprenants peuvent ainsi reproduire des modèles d'action à mettre en œuvre selon les situations rencontrées, émettre des hypothèses par rapport aux données et procéder à une réflexion inductive surtout dans le cas de la consultation de corpus. Effectivement, on peut avancer que la formation a plus profité en

termes d'efficacité à la consultation de corpus, peut-être du fait de sa nouveauté et de la manière de l'aborder. Les corpus offrent aux apprenants la possibilité d'exploiter des données, d'observer le co-texte et le contexte et permet d'améliorer la sensibilité linguistique en fonction des besoins. Alors que les résultats moyens provenant de l'utilisation des dictionnaires accentuent leur efficacité médiocre dans la prise de conscience linguistique, Wittorski, citant les travaux d'Argyris *et al.* (1987), évoque l'inefficacité de l'action engagée en l'expliquant par la non-capacité de l'acteur à remettre en cause sa perception de la situation et à reproduire de manière récurrente des modèles d'action inefficaces. Deux solutions d'apprentissage sont alors proposées pour y remédier: la première consiste à adopter une stratégie comportementale différente pouvant être initiée par une autre personne mais sans remettre en question le cadre de la situation que les auteurs ont appelé "*single loop learning*"; l'autre alternative intervient au niveau du cadre de perception qui lui permet d'adopter de nouveaux modèles d'action reposant sur le questionnement et l'analyse de la situation. Ce deuxième choix, "*double loop learning*" serait un véritable changement en termes d'apprentissage, ce que Schön (1983) appelle "*reshaping the situation*". C'est effectivement ce que nous avons observé dans le cadre de notre expérience. Les dictionnaires sont des ressources très appréciées mais très mal utilisées et il est quelquefois plus difficile de changer de mauvaises habitudes, fossilisées par des années de pratique que d'apprendre à en utiliser de nouvelles efficacement. De plus, Surcouf (2010: sn), en évoquant les dictionnaires, déclare:

[qu'] en traduisant, l'apprenant se prive, entre autres, de la possibilité de construire des réseaux d'associations morphologiques, lexicaux et sémantiques en L2. Ainsi, en vient-il en partie à inhiber, à son insu, l'un des procédés essentiels de l'apprentissage de toute langue: l'organisation des savoirs en réseaux permettant un traitement regroupé des informations et une rétention en mémoire plus structurée, plus aisée et donc plus efficace.

6. Conclusion

Cette étude semble démontrer que la prise de conscience linguistique peut être favorisée par la consultation de corpus qui permet de repérer et d'appréhender des phénomènes linguistique de manière efficace. L'ASC s'apparente dans ses grandes lignes à l'apprentissage expérientiel (Kolb 1984) avec lequel il partage des notions essentielles. Nous pensons qu'une prise de conscience linguistique adopte la même ligne que l'apprentissage expérientiel et partage des attitudes et des comportements semblables. Pour autant, la prise de conscience linguistique n'est pas innée et nécessite une formation afin de pouvoir aborder différentes situations en se créant des modèles d'action (Serre 1992). Un aspect majeur de la prise de conscience s'inscrit dans la manière dont nous appréhendons l'apprentissage d'une L2. L'explication réside en partie dans les modalités cognitives de

construction et de réalisation associées à l'apprentissage ou au savoir-apprendre expérientiel. L'apprenant organise ses connaissances par rapport à la représentation de la tâche et également par rapport aux ressources utilisées, il adopte consciemment ou pas un cheminement qui a des répercussions sur sa façon d'apprendre. Il modéliserait mentalement un parcours d'apprentissage qui agira sur sa façon d'appréhender ses connaissances acquises ou en cours d'acquisition.

Les résultats linguistiques de la deuxième expérience indiquent que l'exploitation de données sur corpus augmente sensiblement la capacité de prise de conscience linguistique d'apprenants d'une L2 à repérer et à traiter des faits de langue au niveau de l'item et du co-texte. Cependant, cette prise de conscience n'est efficace que si elle est précédée d'une formation linguistique. La comparaison des résultats avec d'autres ressources de référence telles que le dictionnaire bilingue et unilingue accentuent la propension à mener une réflexion inductive lors de la consultation de corpus et comme le soulignent Chevrier et Charbonneau (2000: sn) "le savoir ainsi constitué est réutilisable pour vivre de nouvelles expériences".

BIBLIOGRAPHIE

- Aston, G. (2001). *Learning with corpora*. Houston: Athelstan.
- Argyris, C., Putman, R. & Mac Lain Smith, D. (1987). *Action science*. San Francisco: Jossey Bass Publishers.
- Bogaards, P. (1988). *Aptitude et affectivité dans l'apprentissage des langues étrangères*. Paris: Crédif-Didier.
- Boulton, A., Château, A., Pereiro, M. & Azzam-Hannachi, R. (2008b). Learning to learn languages with ICT – but how? *CALL-EJ*. <http://www.tell.is.ritsumei.ac.jp/callejonline/journal/9-2/boulton.html> (10.02.2011).
- Boulton, A. (2008a). Esprit de corpus: promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et corpus*, 3, 37-46.
- Boulton, A. (2008b). Looking for empirical evidence of DDL at lower levels. In: B. Lewandowska-Tomaszczyk (éd.), *Corpus linguistics, computer tools, and applications – state of the art* (pp. 581-598). Francfort: Peter Lang.
- Boulton, A. (2010). Consultation de corpus et styles d'apprentissage. *Cahiers de l'APLIUT*, 29, 1, 98-115. <http://apliut.revues.org/967> (24.02.2012).
- BYU-BNC. *British national corpus*. <http://corpus.byu.edu/bnc/x.asp> (10.02.10).
- Chevrier, J. & Charbonneau, B. (2000). Le savoir-apprendre expérientiel dans le contexte du modèle de David Kolb. *Revue des sciences de l'éducation*, 26, 2, 287-324. <http://id.erudit.org/iderudit/000124ar> (15.09.11).
- Cobb, T. (1999). Breadth and depth of vocabulary acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12, 345-360.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier.
- Foerster Von, H. (1960). On self-organizing systems and their environments. In: M. Yovits & S. Cameron (éds.), *Self-organizing systems*. Londres: Pergamon Press.

- Galisson, R. (1991). *De la langue à la culture par les mots*. Paris: Clé international.
- Grand Robert & Collins (2008). *Dictionnaire anglais-français*. Paris: Le Robert.
- Holec, H. (1979). *Autonomie et apprentissage des langues étrangères*. Strasbourg: Conseil de l'Europe. <http://www.lerif.net/A-GRAF/membres/Textes/CONSEIL%20DE%20LA%20COOPERATION%20CULTURELLE.pdf> (19.05.2013). [Repris dans: B. André (éd.) (1989). *Autonomie et enseignement / apprentissage des langues étrangères*. Paris: Didier/Hatier].
- Johns, T. (1986). Micro-concord: a language-learner's research tool. *System*, 14, 2, 151-162.
- Johns, T. ([1990] 1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. [Version remaniée: T. Johns & P. King (éds.), *Classroom concordancing*. *English Language Research Journal*, 4, 27-45.
- Kolb, D. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs: Prentice Hall.
- Landure, C. (2011). Data-driven learning: apprendre et enseigner à contre-courant. *Mélanges CRAPEL*, 32, 162-178. http://www.atilf.fr/IMG/pdf/163-178_landure.pdf. (15.05.2013)
- Lapaire, J. (2009). De la grammaire anglaise balado diffusée au portfolio grammatical électronique: vers de nouvelles pratiques adaptées à l'ère du numérique. *ALSIC*, 12. <http://alsic.revues.org/index1314.html> (08/09/2011).
- Le Prieult, H. (1996). *Grammaire progressive de l'anglais*. Paris: Belin.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL*, 19, 3, 269-286.
- Oxford advanced learner's dictionary* (1974). Oxford: Oxford University Press.
- Schmitt, N. & McCarthy, M. (1997). *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Schön, D. (1983). *The reflective practitioner: how professionals think in action*. New York: Basic books.
- Serre, F. (1992). La science en action: le rapport entre la science et la pratique professionnelle. In: R. Tessier & Y. Tellier (éds). *Changement planifié et développement des organisations: méthodes d'intervention, consultation et formation* (pp. 395-422). Québec: Presses de l'Université du Québec.
- Surcouf, C. (2010). Le dictionnaire bilingue peut-il s'intégrer profitablement dans une stratégie d'apprentissage d'une langue étrangère? *Cahiers de l'APLIUT*, 29, 1, 116-136.
- Test of English for international communication*. <http://www.toeic.eu/> (02.02.2010).
- Tournier, J. (2004). *Précis de lexicologie anglaise*. Paris: Ellipses édition.
- Wittorski, R. (2001). Contribution de l'apprentissage expérientiel et de la science action à la pratique professionnelle. In: M.-P. Mackiewicz (éd.), *Praticien et chercheur: parcours dans le champ social* (pp. 107-119). Paris: L'Harmattan.

Annexe 1

SURNAME

FIRST NAME

CLASS

E.

IUT EPINAL

1. En vous aidant du dictionnaire bilingue

- Rédiger une phrase en anglais en utilisant «**information**» (autre que les exemples figurant dans le dictionnaire)

a) Do you need some information?

- Traduire cette phrase en français

b) Avez-vous besoin d'information?

- Rédiger une phrase en anglais en utilisant «**inquiry**» (autre que les exemples figurant dans le dictionnaire)

c) We should ask inquiries.

- Traduire cette phrase en français

d) On devrait demander des renseignements.

2. En vous aidant du dictionnaire unilingue

- Rédiger une phrase en anglais en utilisant «**advice**» (autre que les exemples figurant dans le dictionnaire)

e) It's just an advice, but you should work.

- Traduire cette phrase en français

f) C'est juste un conseil, mais tu devrais travailler.

- Rédiger une phrase en anglais en utilisant «**suggestion**» (autre que les exemples figurant dans le dictionnaire)

g) This suggestion should help you.

- Traduire cette phrase en français

h) Cette suggestion devrait t'aider.

Annexe 2

SURNAME FIRST NAME CLASS E . IUT EPINAL

3. En vous aidant des phrases provenant du corpus BNC

- Rédiger une phrase en anglais en utilisant «**luggage**» (autre que les exemples présentés)

i) I will take your luggage.

- Traduire cette phrase en français

j) Je vais prendre vos bagages.

- Rédiger une phrase en anglais en utilisant «**suitcase**» (autre que les exemples présentés)

k) Is your suitcase full?

- Traduire cette phrase en français

l) La valise est-elle pleine?

Annexe 3

SURNAME

FIRST NAME

CLASS E.

IUT EPINAL

Advice or suggestion or both

1. The information and Advice is endless: in the case of fire, "stop, drop and roll.
2. Doctors may, in certain circumstances, give Advice about contraception to under 16-year-old girls without parental consent.
3. In the preceding chapters there is a lot of suggestion about how to apply systems ideas.
4. "Be yourself" is, interestingly enough, a piece of advice at home in secularized Western societies as much as in traditional religion.
5. These suggestion are mainly concerned with ways of making the choices implicit in the B & B.
6. There are both that the differences within the group labelled Asian are large.
7. In the course of its report it made several suggestion.
8. He then made one or two to advice for limiting her liability.

Information or inquiry or both

9. For a true design data base to be established more information needs to be available than just these associations.
10. You can retain in your computer a lot of information about what books are in the library.
11. If a piece of information arrives during trading in Japan, it will be reflected in the closing.
12. The job of the archive has been to preserve these inquiry storage media.
13. GP advertised a snip service in his local paper. And he got about 30 inquiry a day.
14. Two information with such machinery available to them, namely, that of the Law Commission.
15. Numerous both were made of the Financial Secretary to elucidate the resulting effect of the Bill.
16. In answer to these inquiry the Financial Secretary gave similar answers.

Luggage or suitcase or both

17. Small babies always need a lot of both: nappies, jars, bottles, wipes.
18. The friends joked about the amount of luggage Julie wanted to take. "She had enough clothes for a month.
19. Jafaar's two pieces of suitcase had been identified by the Lockerbie investigators.
20. They had "observed an individual approach Coding Station 206 with a single piece of both, place the luggage in a luggage container.
21. Clearly things are livelier when the luggage are packed off on Friday morning.
22. They haven't started packing their luggage yet, but they fear it may be inevitable.
23. Her life story is contained here ... eight suitcase of personal memorabilia sold today at Sotheby's.
24. Clothing like this used to be associated with dodgy del boy types who work out of suitcase at street markets.

Corpus et enseignants de français L2 en Flandre: une histoire réussie

Juliette DELAHAIE

Université Paris Diderot – Paris 7 et LLF (UMR 7111), UFR de linguistique,
bât. Olympe de Gouges, rue Albert Einstein, 75013 Paris, France
juliette.delahaie@linguist.jussieu.fr

Danièle FLAMENT-BOISTRANCOURT

Université Paris Ouest Nanterre La Défense et Modyco (UMR 7114)
200 avenue de la République, 92001 Nanterre, France
daniele.boistrancourt@kuleuven.be

Les comédiens font impression sur le public non lorsqu'ils sont furieux, mais lorsqu'ils jouent bien la fureur.

(Diderot, *Paradoxe sur le comédien*, 1773)

This paper presents the results of data-based study carried out in the field of teacher training for second language (L2) French. The project, initiated in 1994, involves the use of spoken French data gathered from French and Flemish speakers, known as the LANCOM corpus. LANCOM was used to train Flemish teachers of French in the late nineties and to define the attainment targets of the Flemish curriculum for French. In 2008 Delahaie carried out research on how three Flemish teachers of French taught a communicative spoken French course to their classes. She shows that only the teacher who benefited from the use of LANCOM in her training succeeds in making her pupils acquire authentic spoken French, even though, due to lack of practice, she is a less fluent speaker of French than her two colleagues. This study addresses the issue of how corpora can be used to meet teachers' needs when they have to deal with the thorny issue of spoken French ability.

Keywords: Corpus, learner corpora, teacher discourse, spoken French, training, spoken language grammar

1. Introduction

Notre propos sera ici de traiter de l'utilisation des corpus dans une perspective de formation d'enseignants, et par rapport à une compétence reconnue comme étant toujours difficile à faire acquérir à des non-natifs du français: l'expression orale. Dans une approche modulaire de la production orale (par ex. voir Roulet 1999), l'enseignant doit mettre l'accent sur au moins trois contraintes auxquelles est soumise la construction du discours, à savoir la contrainte linguistique (syntaxe, lexique, sémantique et phonologie), les contraintes liées à la situation de production (contexte social, univers de référence, etc.), et enfin les contraintes textuelles liées au plan d'organisation du discours. Les trois composantes de la "compétence à communiquer langagièrement" du Cadre Européen Commun de Référence pour les Langues ou CECRL (Conseil de l'Europe 2001) – linguistique,

sociolinguistique, pragmatique – reprennent d'ailleurs ces différents paramètres comme autant d'habiletés à faire acquérir. Or le problème est bien connu: ce que produit l'apprenant en situation scolaire reste toujours dans la plupart des cas de l'écrit oralisé (voir Flament-Boistrancourt 2001a; Bouchard 2005; Andersen 2006), l'accent étant mis sur la composante linguistique du discours. De plus, tout ce que l'on sait aujourd'hui en linguistique sur le français parlé et les conversations, quoique maintenant largement vulgarisé (voir entre autres Blanche-Benveniste 1997 ou Kerbrat-Orecchioni 1996), n'a encore que peu pénétré dans la réalité des dialogues et des exercices proposés dans les classes et les manuels de Français Langue Etrangère (FLE). Pourtant, les initiatives existent¹, sans pour autant constituer une pratique courante dans le domaine de l'enseignement des langues. On citera quelques manuels² importants dans le domaine, et notamment en anglais langue étrangère, le manuel construit à partir du *Cambridge International Corpus of North American English*, intitulé *Touchstone* (voir McCarthy 2004 pour une présentation). En FLE, la première initiative centrée autour de l'enseignement des compétences orales à partir de corpus est celle du Français Fondamental (1964), mais à part *De vive voix* (1972), parmi les manuels les plus récents qui se sont appuyés sur ses tables de fréquence, on citera la série *Parcours* (1997) destinée à l'enseignement du FLE en Belgique néerlandophone, ainsi que des manuels danois (par ex. Andersen & Fristrup 2005³). Le corpus d'Orléans constitue en quelque sorte la deuxième vague de création de corpus à des fins didactiques, mais là aussi, les exploitations se sont réalisées essentiellement à l'étranger: *Les Orléanais ont la parole* (1976) est un manuel rédigé par des enseignants anglais et qui exploite les données du corpus d'Orléans (voir Bergounioux *et al.* 1992). Pour la période récente, on citera le projet PFC (Phonologie du Français Contemporain) qui propose des exploitations didactiques d'un corpus à visée de recherche en phonétique et phonologie (voir Detey *et al.* 2010). Ces expériences réussies ne constituent qu'une partie, la plus visible, de ce que l'on peut faire avec des corpus de français parlé en matière d'enseignement des langues. De plus, les manuels à base de corpus n'ont de réel impact sur l'apprentissage que si les enseignants eux-mêmes sont bien formés pour les utiliser. En parallèle avec des essais prometteurs d'intégration de corpus dans les activités de la classe (voir Tyne 2012), nous voudrions ainsi montrer qu'une sensibilisation des enseignants au français parlé à partir de corpus oraux pourrait se révéler tout

¹ Voir Chambers (2009) pour une revue récente des projets de recherche autour de la constitution de corpus pour l'enseignement du FLE. Voir aussi Chambers (ce volume) et Kerr (ce volume).

² Une liste des manuels cités est donnée à la fin de la bibliographie.

³ La *Fransk Grammatik* (2005) est un manuel de grammaire à destination d'élèves danois débutants en français; il utilise les tables de fréquence du Français Fondamental pour le choix des items présentés et intègre de façon très réussie nombre d'acquis de la linguistique sur le français parlé et les conversations.

à fait pertinente. Pourtant, dans bien des cas, ce n'est pas la compétence langagière de l'enseignant qui est en cause car, qu'il soit natif ou non-natif, il possède généralement fort bien ce français parlé conversationnel, mais il ne le transmet pas à ses élèves lorsqu'il fait de l'oral (voir Flament-Boistrancourt 2001a).

Pour l'enseignement du FLE dans les établissements du secondaire de Belgique néerlandophone, nous avons tenté de résoudre le problème à l'occasion en 1994 du passage à l'approche communicative des programmes d'enseignement des langues, dans la Communauté flamande de Belgique. Ces nouveaux programmes impliquaient, pour les universités belges chargées de la formation initiale (l'Agrégation⁴) et continue des professeurs de langues, l'élaboration de nouveaux contenus d'enseignement. A l'Université néerlandophone de Louvain, qui a toujours eu une tradition de recherche dans le domaine des corpus de français parlé⁵, l'exigence en la matière a été forte: initier non seulement le futur professeur à une méthode communicative d'enseignement des langues, valable pour n'importe quelle langue, mais faire aussi en sorte que le français enseigné par le professeur soit véritablement celui qui est parlé par le francophone natif dans ses échanges quotidiens. Pour atteindre cet objectif, plutôt que d'exposer uniquement les professeurs néerlandophones de français en formation initiale et continue à la littérature linguistique sur le sujet, il a été décidé de procéder par contraste et de façon '*bottom-up*' ou ascendante. Les formateurs ont donc commencé par réaliser un corpus différentiel d'interactions verbales constitué des scènes typiques de celles que l'on trouve dans les manuels, jouées par des apprenants néerlandophones de niveau intermédiaire d'une part et par des francophones de France d'autre part. Connu sous le nom de corpus LANCOM⁶, ce corpus a donné lieu à de nombreuses publications et il a servi tout au long de l'année 1994-95 à la formation au communicatif des futurs professeurs de français de l'Université néerlandophone de Louvain.

⁴ Le diplôme belge d'Agrégation est l'équivalent du Capes français. Il s'obtient après une licence de philologie romane (4 ans) et à la suite d'une année dite d'Agrégation faite de stages et de cours à l'université.

⁵ Au début des années 1980, l'Université de Louvain-Leuven reçoit en gestion les enregistrements de l'Etude Sociolinguistique sur Orléans et procède à partir de ces enregistrements à l'élaboration d'un corpus de plus d'un million de mots: ELILAP ('Etude Linguistique de la Langue Parlée' – voir Debrock et al. 2000).

⁶ Voir <http://bach.arts.kuleuven.ac.be/elicop> (15.05.2013). Les concepteurs de LANCOM 1 ont été au nombre de trois: Mark Debrock (directeur au début des années 1990 du département de Linguistique Appliquée de l'Université de Louvain-Leuven), Raymond Gevaert (formateur au sein de l'Agrégation de l'Université de Louvain-Leuven et vice-président de la FIPF) et Danièle Flament-Boistrancourt. LANCOM 1, tel qu'il a été exploité au cours de l'année 1994-95, se composait de 39h d'enregistrements, dont 18 heures transcrites représentant un total de 160.593 mots.

Douze ans plus tard, à l'occasion de son travail de thèse, Delahaie (2008) complète LANCOM (désormais LANCOM 1) en lui adjoignant d'autres enregistrements différentiels sur le thème du tourisme (LANCOM 2). Ceux-ci sont réalisés d'une part dans une agence de voyage de la région lilloise et d'autre part dans les classes de trois professeurs de français de Belgique néerlandophone, dont un, Mme G, a bénéficié de la formation par LANCOM 1 en 1994-1995. Delahaie (2008) analyse le fonctionnement de ces trois enseignants dans leurs classes dans le domaine de l'enseignement de l'oral, à partir des jeux de rôle qu'ils mettent en place avec leurs élèves et de leur évaluation. Elle montre que Mme G est celle des trois professeurs qui réussit le mieux à faire produire à ses élèves un français parlé proche de celui du natif francophone.

Nous présenterons donc dans un premier temps les résultats obtenus dans LANCOM 1 et la manière dont ceux-ci ont été exploités avec les futurs professeurs de français en année d'Agrégation à l'Université de Louvain en 1994-95. Nous analyserons ensuite de façon détaillée le comportement des élèves et des enseignants de LANCOM 2 dans des activités de jeux de rôles et leur reprise, en nous concentrant notamment sur deux points précis qui s'étaient révélés être des zones importantes de contraste entre les francophones et les apprenants néerlandophones de français de LANCOM 1: l'utilisation du conditionnel dans les requêtes et un certain nombre de structures syntaxiques bien connues comme étant caractéristiques du français parlé (organisation thème-rhème, présence d'impératifs discursifs et de marqueurs de structuration de la conversation, etc.). Nous essaierons en conclusion de voir si l'on peut établir une corrélation entre la formation universitaire de Mme G et sa manière d'enseigner l'oral.

2. La formation au communicatif par LANCOM 1 en 1994-95

2.1 Les grands contrastes du corpus

Les apprenants néerlandophones de LANCOM 1 sont des adolescents en dernière année de secondaire dans des écoles de Gand, en moyenne d'un niveau B1 du CECRL en français. Le tableau 1 (voir ci-dessous) résume de façon synthétique les grandes différences trouvées entre la manière de dire de ces apprenants et celle des Français enregistrés. Le résultat ne surprend pas: le non-natif produit le plus souvent un oral qui déploie une syntaxe caractéristique de l'écrit, ce que Blanche-Benveniste (1997: 9) appelle "la langue du dimanche". Cette langue, qui est mobilisée par le francophone seulement dans des situations bien particulières, n'est pas le parler de tous les jours, celui des interactions de la vie quotidienne. L'utilisation de cette langue du dimanche par les apprenants néerlandophones était favorisée par un grand nombre d'interférences du néerlandais dans le domaine de la morphosyntaxe. Ce que produisaient nos apprenants en français, c'est en

effet ce qu'ils auraient fait en néerlandais, on le voit bien sur trois points précis:

- la façon de poser des questions, où ce qui domine chez le non-natif, c'est la structure par inversion du néerlandais;
- la formulation des requêtes non pas au conditionnel mais au présent, comme cela serait le cas en néerlandais;
- une organisation thème / rhème du discours avec des moyens syntaxiques propres au français et que nos apprenants n'utilisent donc pas: détachement en tête de phrase du syntagme nominal (SN) thématique et enchaînement sur le rhème, le plus souvent par l'intermédiaire d'un *c'est* dit "introduceur rhématique" (voir Flament-Boistrancourt 2008).

Ce sont ces trois points que nous avons décidé de travailler au cours de l'année 1994-95 avec les futurs professeurs de français de l'Agrégation de l'Université de Louvain-Leuven.

2.2 Exploitation formative des contrastes de LANCOM 1

LANCOM 1 sera exploité de trois manières différentes avec les futurs professeurs en formation:

1. présentation au futur professeur de français en formation des contrastes que manifestent dans le corpus les productions des néerlandophones et des francophones;
2. examen critique de la façon dont les trois points retenus ci-dessus sont alors traités par les manuels scolaires du marché belge;
3. élaboration de modèles phrastiques ou conversationnels à imiter par les élèves.

L'acte de question ayant été traité longuement dans nombre de publications antérieures (voir Flament-Boistrancourt *et al.* 1996, 1997; Flament-Boistrancourt & Cornette 1999; Flament-Boistrancourt 2001b, 2004), nous nous attarderons ici uniquement sur le cas du conditionnel et du traitement de l'opposition thème / rhème, que nous étudierons ensuite dans LANCOM 2.

	Néerlandophones	Francophones
Genre	2 monologues	1 vraie interaction: l'interlocuteur reprend les mots de l'autre, complète ses phrases, lui renvoie toutes sortes de signaux (phatiques, marques d'assentiment). Une co-construction effectuée avec une patiente recherche de consensus: interrogations méta-communicatives en <i>hein</i> .
Registres	familier (<i>boulot, tacot; tu vs. vous</i>)	soutenu
Syntactico-discursif	de l' écrit oralisé	de l' oral a) MSC ⁷ (<i>bon, alors, donc</i>) b) thématisations (<i>bon alors, pour les tarifs</i>) c) impératifs discursifs (<i>disons, allez, écoutez</i>)
Conversation	questions / réponses ouvertures, clôtures de la L1	des échanges (3 interventions)
Actes de langage	jussifs ('blunt') a) des consignes ⁸ en <i>devoir</i> (<i>tu dois</i> + infinitif) b) requêtes au présent (transfert de la L1)	en <i>falloir</i> (<i>il faudrait que...</i>) requêtes au conditionnel
Requêtes (baby-sitting) ⁹	SV (interrogations par intonation): 34% V-CL (ordre verbe + clitique): 41% ESV (interrogations en <i>est-ce que</i>): 24% interrogations dites intermédiaires ¹⁰ : 4%	81% 3% 15% 29%

Tableau 1. Le français parlé par les apprenants néerlandophones et par les locuteurs francophones: présentation différentielle

2.2.1 Le cas du conditionnel

Le conditionnel de politesse, rarement utilisé par les néerlandophones, était clairement ce qui caractérisait la manière de dire des francophones dans toutes les scènes d'embauche, d'agence de voyage ou de baby-sitting de LANCOM 1. L'extrait suivant le montre bien:

⁷ 'Marqueur de Structuration de la Conversation'.

⁸ Ces 'consignes' renvoient aux scènes d'embauche de LANCOM 1 où l'employeur décrit au candidat les tâches qu'il aura à accomplir.

⁹ Le 'baby-sitting' est une partie de LANCOM 1 étudiée entre autres dans Flament-Boistrancourt & Cornette (1999).

¹⁰ On en trouve la liste complète dans Flament-Boistrancourt & Cornette (1999): des interrogations totales dans la forme, mais partielles sur le fond, des assertions-demandes de confirmation, etc.

- (1) A- allô bonjour madame! je vous appelle pour l'annonce que vous avez: placée dans le journal pour une garde d'enfant = euh: **je voudrais** savoir si la place est toujours libre
 B- ah oui oui! je suis ravie que vous me téléphoniez étant donné que c'est assez urgent **je souhaiterais** quelqu'un = de disponible pour euh: à partir de ce mercredi de cette semaine
 (LANCOM 1, jeu de rôles du baby-sitting, côté francophone)¹¹

Les apprenants néerlandophones de LANCOM 1, en revanche, n'employaient pas ce type de conditionnel:

- (2) A- allô ici Nancy Vanheule c'est concernant de l'article qui était dans le journal = **je voulais** = demander quelques informations précises = par exemple com de combien d'enfants avez-vous de quel âge ils sont = combien d'heures je **dois** je **dois** travailler de quelle à quelle heure? et etcetera
 B- euh: oui un moment je vais vous donner les informations de = euh: je **veux** donc un un baby-sit pour mes enfants = et je **veux** une personne qui **veut** faire du baby-sit = du baby-sit pendant les week-ends euh: [...]
 (LANCOM 1, jeu de rôles du baby-sitting, côté néerlandophone)

Chez les locuteurs francophones, ce conditionnel dans les requêtes et les descriptions de tâches à accomplir peut même être associé à un lexique très familier, comme en témoigne l'extrait authentique suivant enregistré dans une agence de voyage d'un quartier bourgeois de la banlieue lilloise. Cette association paradoxale est bien la preuve s'il en était besoin, que ces manières de dire modalisées sont très ritualisées, et associées à différents registres de langue:

- (3) A- bonjour
 B- je **voudrais** donc quelques renseignements pour euh: pour aller au Ténéri euh: ouais aux Canaries donc à: Ténérife = [...]
 A- avec un hôtel je suppose
 B- avec un hôtel le plus **merdique** possible le moins cher possible puisqu'en fait c'est pour des étudiants
 (LANCOM 1, agence de voyage, côté francophone)

Comme nous le montrons alors dans nos formations, ce rôle modal d'adoucisseur du conditionnel français n'est jamais bien mis en évidence dans les grammaires pédagogiques du marché belge flamand de l'édition scolaire de cette époque-là (voir par ex. la *Grammaire 2000* et la *Grammaire pratique*). Ces dernières privilégient des présentations de type syntaxique: l'emploi du conditionnel dans les subordonnées (le futur du passé), les principales liées à un *si* et les indépendantes. Lorsque, dans ce dernier cas, la demande polie est recensée, elle l'est dans une liste de nombreuses autres valeurs modales possibles (irréel, éventualité, indignation, étonnement, etc.), qui ne fait donc pas ressortir l'importance à un niveau pragmatique du conditionnel dans les interactions entre francophones.

¹¹ Voir l'annexe 1 pour les conventions de transcription.

En réaction, nous avons donc insisté dans nos formations sur l'importance du conditionnel dans les requêtes. A partir de contextes situationnels précis puisés dans LANCOM 1, comme par exemple les extraits (1) et (2), nous faisons émerger le caractère systématique de l'emploi du conditionnel, là où dans la même situation un élève néerlandophone employait le présent. L'objectif de ce travail sur corpus a d'abord été d'opérer chez les enseignants une conscientisation contrastive des moyens utilisés par chaque langue. Ensuite, nous leur proposons une pédagogisation à destination de l'élève, en commençant par utiliser des exercices qui leur étaient familiers: par exemple, des exercices de systématisation ayant pour consigne d'adoucir des "formes un peu dures". Il était alors demandé à l'élève de faire passer au conditionnel des requêtes au présent (*je veux* → *je voudrais*), sans pour autant conjuguer tout le paradigme comme dans la tradition du *Bescherelle* (voir *Bescherelle* 2012). Nous leur présentons ensuite des modèles discursifs oraux et écrits où le conditionnel se trouvait mis en vedette. Ce que nous avons travaillé en formation d'enseignants sera par la suite repris par Arens *et al.* (1999). Ainsi dans le *Parcours bleu* 3, on montre bien à travers les situations de communication choisies, que ce conditionnel de politesse est tout autant de mise dans des échanges amicaux entre jeunes que dans des échanges quotidiens entre inconnus, ou dans un contexte plus institutionnel (au restaurant, dans les magasins, à l'école, etc.). On trouve par exemple dans le *Parcours bleu* (1999: 40) le modèle textuel suivant:

- (4) Anne-Sophie
 Aymeri et moi, nous sommes passés à 16h. Tu n'étais pas là. Tu pourrais nous appeler ce soir avant 19h? On aimerait aller au cinéma avec toi. Tu ne voudrais pas demander à Pierre de venir aussi? Ce serait chouette.
 Bisous
 Charlotte

Ce modèle écrit est fabriqué, mais il donne une grande impression de naturel, d'autant plus que le conditionnel est associé à des interrogations par intonation, caractéristiques elles aussi de la manière de dire du francophone, même si l'ensemble n'est pas parfait: l'emploi de *nous* et de *ne* est une caractéristique bien connue du français non spontané.

2.2.2 Traitement de l'opposition thème / rhème

L'opposition thème / rhème donnait lieu dans LANCOM 1 à des réalisations diverses et faisait très curieusement intervenir des introducteurs thématiques ayant peu retenu l'attention des linguistes, comme par exemple *niveau* et *par rapport à* (voir Flament-Boistrancourt 2008; Delahaie & Flament-Boistrancourt à paraître):

- (5) - Niveau prix, ça donne quoi?
 - Sinon, par rapport à la rémunération, je ne sais pas si vous avez vu ce qu'on proposait [...]

Poursuivant notre objectif de 'conscientisation contrastive' en formation d'enseignants, nous avons montré ces fonctionnements à nos futurs professeurs et nous leur avons expliqué leur intérêt au plan communicatif dans une langue comme le français: placer dans une position tonique un mot important du message ou encore utiliser, selon l'expression de Mel'čuk (1993), des "phrasèmes" réutilisables à peu de frais dans de nombreux autres contextes: *niveau -horaires / location de voitures / climat-, ça donne quoi?*

Ces types d'introducteurs thématiques n'ayant cependant pas encore fait l'objet d'études abouties chez les linguistes, nous nous en sommes tenu au plan de l'exploitation pédagogique à de simples structures à SN détaché suivies de rhèmes introduits par *c'est*. Habitué à une linguistique de la phrase et à des manuels fleurant encore bon l'héritage de la méthode directe, nos publics ne connaissaient *c'est* que sous l'angle du présentatif ou de l'outil de mise en relief. Pour bousculer cette représentation, nous présentions habituellement l'extrait authentique ci-dessous:

- (6) - Oui, et est-ce que vous avez de bons résultats heu ... vous êtes heu ... en quelles classes?
 - Ecoutez, je suis professeur au lycée de Sartrouville, donc **c'est** en banlieue, **c'est** un lycée assez modeste, **c'est** pas un lycée de champions, un lycée exceptionnel, et **c'est** en classe de première, de première A et de première G. Les premières A, **c'est** des élèves un peu littéraires, les premières G, **c'est** des élèves plus techniques.

(Emission *Apostrophes*, septembre 1984, Bernard Pivot, Patrick Grainville)¹²

Dans ce passage *c'est*, morphologiquement invariable, ne met pas en valeur un constituant de phrase (ou alors il met tout en valeur), mais fonctionne comme une sorte de connecteur discursif qui permet au plan informationnel de faire progresser l'énoncé.

Nous avons proposé aux enseignants des modèles pédagogiques exploitant cette structure, pour, par exemple, dire ce que l'on aime, activité que l'on trouve souvent dans les manuels:

- (7) A- Ton passe-temps favori, pour toi, c'est quoi?
 B- Moi, mon passe-temps favori, c'est d'abord le tennis. Mais ensuite, ce que j'aime faire aussi, c'est du vélo.
 A- Et pour toi, c'est quoi?
 C- Oh, pour moi, une après-midi agréable, c'est sortir avec mes copains.
 D- Tout à fait! Ça, c'est chouette, c'est même super! [...] ¹³

Ces modèles sont simples et cognitivement économiques. Ils font intervenir peu de verbes conjugués, privilégiant au contraire les noms et les infinitifs, et ils placent les informations importantes dans des positions accentuées.

¹² Dans les extraits, nous soulignons les mots importants par le caractère gras.

¹³ Ce dialogue est dû à Raymond Gevaert.

Pour conclure, nous voudrions insister sur la manière dont nous avons utilisé le corpus LANCOM 1 en formation d'enseignants à l'Université de Louvain. Loin de nous en tenir à une simple exposition des résultats de notre corpus, illustrés d'exemples bien choisis renvoyant à la littérature linguistique sur le sujet, nous avons toujours eu le souci de faire réfléchir le futur enseignant sur ses représentations métalinguistiques (le conditionnel n'est pas forcément lié à une condition ou à un niveau de langue, *c'est* peut être tout autre chose qu'un présentatif, etc.). De plus, nous avons montré que les résultats provenant de l'étude de notre corpus étaient toujours réinvestissables au plan pédagogique dès les niveaux débutants. Il nous est ainsi apparu que la meilleure manière de faire de la production orale avec un apprenant, c'était non seulement de lui permettre un accès direct à des documents authentiques de français parlé, mais aussi de lui donner des modèles à reproduire: des dialogues fabriqués où le naturel est atteint non par la copie du naturel, mais par une stylisation de celui-ci et une imitation maîtrisée de ses ressorts les plus fondamentaux. Douze ans plus tard, la constitution d'un LANCOM 2 à l'occasion de la thèse de Delahaie (2008) permet un retour sur cette formation.

3. LANCOM 2 en 2006: étude qualitative de jeux de rôles en classe

3.1 Présentation des données

Du côté francophone, les données que nous utilisons sont constituées de sept interactions dans une agence de voyage de la périphérie aisée de Lille (Marcq-en-Baroeul), avec trois employées différentes (20.000 mots environ). C'est cependant surtout le versant néerlandophone qui nous intéressera ici, formé de jeux de rôles et de leur évaluation, enregistrés en 2006 dans un collège de Belgique néerlandophone (Diksmuide). Du côté des apprenants, il s'agissait de jouer une interaction dans une agence de voyage, activité qui avait été préparée en amont par la création d'un scénario et un travail grammatical sur les formes jugées appropriées par les enseignants. Chaque interaction était ensuite reprise 'à chaud' par les enseignants pour évaluer et corriger les productions des élèves. Ont été enregistrés et filmés les trois enseignants de français du collège, que nous nommerons Mme G, Mme C et Mr L, dans des classes de niveaux différents, du premier et du deuxième degré belge (des élèves de douze à seize ans)¹⁴. Parmi les trois enseignants, seule Mme G a bénéficié de la formation mise en place à partir du corpus LANCOM 1 en 1994-1995. L'annexe 2 permet de récapituler les caractéristiques des données utilisées dans ce nouveau corpus.

¹⁴

Le système éducatif belge comporte six années de primaire et six années de secondaire, ce dernier étant divisé en trois degrés de deux ans chacun.

Au total, les données sont constituées de 36.000 mots (environ 5 heures et 17 minutes). Comme nous avons un peu plus de données sur Mr L que sur Mme C et Mme G, notre étude sera donc qualitative; contrairement aux grands corpus, elle ne permet pas de généraliser la situation décrite à l'ensemble des enseignants de Belgique néerlandophone. En revanche, elle permet de voir comment fonctionne un monde social, ici comment fonctionne la situation d'enseignement du français parlé dans un contexte précis¹⁵. Tout comme pour les récits de vie ou les entretiens sociolinguistiques, notre objectif est de décrire en profondeur une situation d'enseignement, en prenant en compte des paramètres constitutifs de cette situation que sont les profils linguistiques et géographiques des enseignants enregistrés et leur formation professionnelle. L'objet de notre étude sera de fournir une interprétation des phénomènes observés, qui puisse éventuellement servir à la formation d'enseignants.

3.2 Présentation générale des évaluations de jeux de rôles à l'agence de voyage

Tous les élèves des classes enregistrées ont eu des cours de français depuis l'école primaire, à raison de quatre heures par semaine, puis trois ou quatre heures à partir de l'école secondaire, en fonction de la section choisie. Ces collégiens ont donc un niveau intermédiaire-avancé qui permet de leur enseigner les caractéristiques du français parlé, domaine souvent jugé difficile. C'est ce que tentent de faire les trois enseignants enregistrés, mais avec plus ou moins de bonheur, et nous présenterons ici quelques caractéristiques de ces reprises.

Dès la présentation de l'organisation globale de la reprise des jeux de rôles, la différence entre les trois enseignants est claire. Mme G est orientée vers le français parlé, et au moment de la phase d'évaluation à laquelle participe toute la classe, elle divise le tableau en trois colonnes:

- (8) G- O.K merci bien et on va retravailler en **trois colonnes** = = \$ donc on a le **vocabulai:re grammai:re et cultu:re** ou ce qu'on appelle eu:h tout ce qui est les gestes le non-verbal bon = je trouve que ce premier groupe a très bien parlé le contact visuel était très bon donc je vous remercie [...]

(13-14 ans¹⁶, Mme G)

En réalité, la colonne *culture* accueillera toutes les remarques qui ne sont ni de la grammaire, ni du vocabulaire: formules de politesse, attention aux échanges rituels, formules toutes faites, etc. Mr L en revanche propose des critères beaucoup plus axés sur la correspondance oral / écrit et sur la traduction: 'prononciation', 'fautes grammaticales', et 'le français ≠ le

¹⁵ Voir à ce propos Bertaux (1997) sur le statut et les fonctions des données empiriques, et notamment des études qualitatives à partir de récits de vie.

¹⁶ Age des apprenants.

néerlandais'. Mme C enfin valorise l'écrit à travers l'oral: avant leur prestation orale, les apprenants doivent donner leur dialogue écrit à l'enseignante qui s'en servira pour évaluer les jeux de rôles. Les approches de l'évaluation proposées par les enseignants sont donc sensiblement différentes, les contenus également, et l'on voit d'emblée que seule Mme G fera référence à autre chose qu'au grammatical proprement dit, ce qu'elle a appelé le culturel, et qui relève en fait de la compétence pragmatique et sociolinguistique.

Ainsi, deux tendances contraires se dessinent dans le contenu des reprises proposées (voir Delahaie 2009 pour une analyse plus détaillée de celles-ci) et c'est surtout au niveau de la syntaxe que les disparités sont les plus fortes. Mme C et Mr L ont en effet tendance à privilégier l'ordre des mots de la phrase, notamment lorsqu'il s'agit de reprendre une structure interrogative. Ainsi en (9) et (10), Mme C et Mr L relèvent une erreur dans l'énoncé de leurs élèves, mais ils ne remarquent pas l'incongruité de l'inversion sujet-verbe, structure interrogative peu utilisée en français parlé (voir *supra* LANCOM 1):

- (9) C- [...] eu:h au début vous avez dit à *quel hôtel avous pensé* c'est quoi ça à *quel hôtel avous pensé* = essayez de vous corriger = à *quel hôtel a = vous = pensé* = à *quel* <INT> C écrit au tableau </INT>
E- à l'hôtel
C- voilà à *quel hôtel* = si on a *vous* on aura évidemment = *avez-vous pensé* oui↑
et puis [...]

(14-15 ans, Mme C)

- (10) L- [...] une structure un peu compliquée eu:h je crois que (prénom) a dit quelque chose comme = eu:h c'est un peu compliqué hein il faut pas trop se compliquer la vie = on peut dire *quand = pensez-vous* (prénom)
E- *de arriver*
L- oui et c'est c'est encore moins compliqué il faut pas de *de* = donc j/e crois qu'on peut dire *quand comptez-vous arriver = au Congo* [...]

(15-16 ans, Mr L)

L'enseignant met en valeur une formulation de la question par inversion, typique d'une situation plutôt formelle, associée à une recherche de vocabulaire (*compter / penser*). Or dans ce contexte situationnel précis, celui d'une interaction à l'agence de voyage, un francophone aurait plus volontiers dit *vous voulez arriver quand au Congo?* ou *quand vous voulez arriver au Congo?* C'est ce que montre l'étude des questions dans les scènes d'agence de voyage du côté francophone (Delahaie 2008): dans les sept interactions à l'agence de voyage du corpus LANCOM 2, les questions construites avec un mot interrogatif placé avant ou après la suite sujet-verbe non inversée constituent plus de 60% des interrogations partielles. Ces résultats ne sont pas surprenants et ont déjà été mis en valeur à maintes reprises, par Gadet (1989: 139) ou Coveney (1996). Mr L est d'autant moins sensible à la structure de la question à l'oral, qu'elle ne s'adapte pas aux traductions littérales du néerlandais. Or c'est souvent par ce biais qu'il sollicite la production de ses élèves:

- (11) L- [...] voilà euh autre chose par exemple euh (prénom) c'est (prénom) qui a posé la question et puis on a répondu eu:h *wat is de totale prijs* = eu:h j/e crois que quelqu'un a dit *qu'est-ce que c'est le prix total* = oui↑
E- *quel est la prix totale*
L- *quel est le prix total* il faut pas se compliquer la vie hein↑ on dit simplement *quel est le prix total*
(15-16 ans, Mr L)

Mme G au contraire ne passe pas par la traduction littérale dans notre corpus et, loin d'en rester au cadre de la phrase, elle tente même parfois de sensibiliser à la co-construction de l'interaction. En tant que construction collective, l'interaction implique en effet une détermination réciproque et continue des comportements langagiers des partenaires en présence (voir Kerbrat-Orecchioni 1996). Cela passe notamment par l'emploi de signes d'engagement mutuel dans la conversation, comme les marqueurs d'accord dont Mme G souligne l'importance dans cet extrait:

- (12) G- [...] n'oubliez pas non plus les interjections au téléphone = il y a des groupes qui font ça très très bien quelles sont les interjections au téléphone? pas tout tout le temps *hm hm hm hm* qu'est-ce qu'on peut dire d'autre? = aucune idée? oui
E- *d'accord*
G- *d'accord* = oui? (prénom)?
E- *bien*
G- *bien* = autre chose?
E- *je vois*
G- *je vois bien entendu je comprends c'est parfait etcetera* hein donc il faut compléter il faut meubler un petit peu les silences [...]
(13-14 ans, Mme G)

Mme G a bien compris que le francophone de France n'aimait pas les silences, et surtout qu'il avait sans cesse besoin de l'approbation de son interlocuteur pour poursuivre son discours. En revanche (peut-être par influence du néerlandais?), elle se trompe sur le choix des mots. Un recensement exhaustif des marqueurs d'accord dans les interactions à l'agence de voyage de LANCOM 2 a en effet montré que les marqueurs les plus employés étaient *oui* (49,1% des marqueurs d'accord), *voilà* (13,8%), *d'accord* (8,9%), et *OK* (9,3%)¹⁷, mais pas *bien entendu*, *je vois* ou *je comprends* (voir Delahaie 2009).

Au-delà de ces tendances générales de reprise, nous nous intéresserons maintenant à la façon dont dans LANCOM 2, le conditionnel de politesse et la structuration thème-rhème sont traités.

3.3 Le conditionnel de politesse dans LANCOM 2

Dans les jeux de rôles de la classe de Mme G, il y a un grand nombre de conditionnels de politesse: nous en avons recensé 90 (sur 5.735 mots), essentiellement avec les auxiliaires modaux *pouvoir* et *vouloir*. Ces

¹⁷ Le pourcentage est calculé par rapport au nombre total de marqueurs d'accord relevés dans le corpus (644 occurrences de marqueurs d'accord pour 20.000 mots retranscrits)

conditionnels sont beaucoup moins fréquents dans les jeux de rôles des deux autres enseignants: 19 occurrences chez Mr L (sur 8.892 mots), 23 chez Mme C (sur 5.848 mots), soit 1,57% chez Mme G vs. 0,213% chez Mr L et 0,39% chez Mme C. Or ces différences ne sauraient être corrélées au niveau des élèves, car c'est justement Mme G qui a les apprenants les moins avancés en français (les 13-14 ans), et dont, pourtant, les requêtes en agence de voyage sont toujours au conditionnel:

- (13) E1- bonjour asseyez-vous
 E2- merci
 E3- merci
 E1- madame oui
 E2- nous sommes membres d'un mouvement de jeunesse Chiro et **nous voudrions** aller en colonie à Sparte
 E1- ah Sparte euh avez-vous trouvé quelque chose concret dans le dépliant?
 (13-14 ans, Mme G)

Chez Mme C et Mr L, à quelques exceptions près, les apprenants formulent leurs requêtes au présent, comme cela serait le cas en néerlandais, y compris chez les 15-16 ans d'un niveau plus avancé en français (B1⁺ selon le CECRL):

- (14) E1- bonjour
 E2- bonjour
 E3- bonjour
 E1- euh est-c/e que je peux vous je peux vous aider?
 E2- oui
 E3- bah oui euh nous sommes jeunes un jeune couple et
 E2- et nous sommes ensemble depuis un an et on **veut** célébrer ça
 E3- oui et moi **je donne** un cadeau pour célébrer ça
 E2- oui **on veut** aller on veut faire un euh excursion de de ville
 E1- et au quel pays avez-vous pensé? au quel pays
 (15-16 ans, M.L)

- (15) E1- bonjour je peux vous aider?
 E2- bonjour madame je **veux** réserver un hôtel eu:h un un gîte rural près du parc de Mercantour
 E1- quand **voudrez-vous** aller = et: avec combien de personnes?
 E2- c'est pour le mois de septembre et c'est pour deux personnes
 (14-15 ans, Mme C)

Une telle différence d'emploi chez les apprenants peut s'expliquer par le fait que Mr L et Mme C ne présentent jamais de requêtes au conditionnel à leurs élèves, contrairement à Mme G qui les pointe de façon explicite à l'apprenant, ici dans une discussion avec l'informateur E2 venu filmer la classe:

- (16) G- OK = OK très bien pour le premier groupe
 E2- oui bravo c'est bien hein le conditionnel hein
 G- j'ai très peu j'ai très peu de choses
 E2- on est très impressionnés
 G- **le conditionnel ils sont polis et tout**
 E2- ouais ils sont polis oh là là
 (13-14 ans, Mme G)

3.4 Traitement de l'opposition thème / rhème dans LANCOM 2

Nous avons vu ci-dessus que la formation à partir de LANCOM 1 avait mis l'accent sur les structures à SN détaché suivis d'un *c'est* introducteur rhématique. Dans les jeux de rôles de LANCOM 2, la structure avec un *c'est* rhématique n'est pas naturelle chez l'apprenant néerlandophone. Il en fait souvent l'économie pour se couler dans le moule syntaxique sujet + copule:

- (17) E1- et qu'est-ce que le prix exact?
E4- le prix exact **est** mille sept cents euros
(14-15 ans, Mme C)

A nouveau, les disparités entre les classes de nos trois enseignants sont fortes. Certes, tous les apprenants utilisent des *c'est* en très grand nombre (1,952% dans les jeux de rôles joués par les apprenants de Mme G, 2,72% chez Mr L et 2,99% chez Mme C), mais il s'agit souvent chez Mr L et Mme C de *c'est* dits présentatifs, sans reprise du SN:

- (18) E1- et quelle région?
E2- euh nous euh
E3- le Costa Brava peut-être?
E1- ah oui bonne choix c'est près de la côte et près de Barcelone et euh
E3- ah
E2- ah oui c'est bien
(14-15 ans, Mme C)

En revanche, chez Mme G, l'ensemble des classes utilise une structure en *c'est* avec reprise dans le contexte particulier du 'numéro de téléphone', ce que ne font jamais les apprenants des deux autres enseignants:

- (19) E1- et maintenant votre adresse et votre numéro de téléphone
E3- notre adresse est rue du moulin numéro six à Bruxelles =
E1- et le numéro?
E3- euh **c'est le** zéro quatre = soixante-dix-sept = cinquante = vingt-et-un = trente
E1- donc rue du moulin blanc six à Bruxelles et **le numéro c'est le:** c'est: zéro quatre soixante-dix-sept cinquante vingt-et-un trente d'accord?
(13-14 ans, Mme G)

Mme G insiste de façon explicite sur ces structures avec ses élèves, et elle le fait sans recourir à aucune métalangue, en les présentant tout simplement comme des séquences préfabriquées:

- (20) G- [...] petit détail (prénom) si tu donnes ton numéro de téléphone qu'est-ce qu'on dit? [...] *mijn telefoon is?*
E- *mon numéro de téléphone c'est*
G- *c'est le* j'avais pas entendu le ce hein oui tu l'as dit j'avais pas compris **vous notez encore une fois mon numéro de téléphone c'est le et pas est le hein il faut noter le c'est** [...]
(13-14 ans, Mme G)

Parmi les trois enseignants étudiés, seule Mme G réussissait vraiment à faire en sorte que les élèves produisent une morphosyntaxe caractéristique d'un français parlé qui ne serait pas la langue du dimanche. Pour rendre compte de cette disparité, nous nous intéresserons donc à la formation universitaire et à

l'origine géographique de ces trois enseignants, cette dernière jouant un rôle identitaire important au sein des différentes communautés linguistiques de Belgique (voir Francard: 1995)¹⁸.

4. Conclusion: corpus et conscientisation d'enseignant

Mr L et Mme G sont tous les deux des néerlandophones, originaires de la province belge de Flandre occidentale et ils ont appris le français comme une langue étrangère, même si la proximité avec la France et les séjours qu'ils y ont faits les ont amenés à développer une excellente maîtrise de cette langue. Mme C en revanche est une francophone de Bruxelles qui comprend et parle bien le néerlandais. Or c'est elle, la francophone, qui est la plus attachée à la transmission, même à l'oral, d'un français écrit normé ("la langue de Molière" *sic*). Contrairement à ses deux collègues néerlandophones, et pour les raisons particulières que nous détaillons ci-dessous, elle est hostile à tout changement de manuel dès qu'elle y relève des mots comme *truc* ou *machin*. Comme Mr L, Mme C n'a jamais reçu d'enseignement sur le français parlé et le fonctionnement des interactions verbales en français au cours de sa formation universitaire initiale, mais elle se montre une gardienne plus vigilante du temple de la norme écrite que son collègue néerlandophone. Cette réaction s'explique sans doute en raison d'une double insécurité linguistique: l'insécurité habituelle du Belge francophone par rapport au français de Paris (voir Blampain *et al.* 1997), mais aussi très certainement, pour une francophone installée en Flandre profonde, un sentiment d'insécurité face à la montée en puissance du néerlandais dans la partie néerlandophone du pays. En effet, les familles bourgeoises n'y parlent plus français à la maison, et la maîtrise du français n'est plus, comme il y a un siècle, nécessaire pour faire carrière ; très récemment, des parents résidant à Bruges se sont plaints auprès des directeurs d'école et dans la presse (voir le Journal namurois *Province* du 1^{er} septembre 2011 et *De Morgen* du 31 août 2011) que, dans les écoles de leurs enfants, il ne soit parlé que français pendant les cours de français!

Au plan universitaire, Mme C et Mr L ont reçu en leur temps la formation universitaire (licence de philologie romane) et professionnelle (l'année d'Agrégation belge), telle qu'elle est dispensée dans les universités belges avec un fort accent mis sur la littérature et la grammaire normative (*Le bon Usage* de Grevisse & Goosse 1993). Les stages de formation continue dont ils peuvent bénéficier au cours de leur vie professionnelle (notamment avec le *Cavilam* et *Formacom*), portent toujours sur de nouvelles approches en didactique des langues (CECRL et perspective actionnelle), mais l'accent est

¹⁸ L'enregistrement des évaluations et jeux de rôles s'est accompagné de nombreux entretiens avec les enseignants, aussi bien à propos de leur pratique enseignante qu'à propos de leur formation universitaire, en avril et octobre 2006.

mis sur la pédagogie, et il n'y a jamais de réflexion sur la langue comme matériau linguistique. Mme G a également eu ces formations, mais elle a bénéficié en plus au cours de son année d'Agrégation d'une initiation au communicatif par le biais du corpus LANCOM 1, et l'on serait ainsi tenté de penser que son enseignement pertinent de l'oral n'est pas uniquement le fruit du hasard. Dans tous les cas, on pourra établir une corrélation entre sa formation avec LANCOM 1 et ce qu'elle enseigne, même si cette hypothèse forte reste à vérifier empiriquement (il faudrait pour cela renouveler l'expérience ou retrouver d'autres enseignants formés par LANCOM 1).

Ce que l'on peut néanmoins constater, c'est que Mme G a retenu les leçons de LANCOM 1, et l'on peut avancer plusieurs raisons à cela. Elle les a retenues sans doute parce que nous procédions constamment par contraste, en pointant les différences à l'aide des données de notre corpus entre la manière de dire du natif et celle du non-natif, les fonctionnements décrits dans les grammaires pédagogiques et ceux réellement à l'œuvre dans le parler du francophone. Si Mme G a mis LANCOM 1 en pratique, c'est sans doute aussi parce que nous ne nous contentions pas de bousculer les représentations, mais que nous donnions des outils (les modèles à imiter). Enfin a peut-être également joué un rôle le fait que Mme G a fait partie des sept étudiants qui, dans le cadre de leur mémoire de fin d'études, ont transcrit LANCOM 1 et ont analysé quantitativement (établissement de statistiques) et qualitativement les contrastes qu'ils observaient dans leur partie. Dans le cas de Mme G, la formation a donc été participante et c'est sans doute ce qui a facilité les prises de conscience (voir Tyne 2009 sur l'utilité de la transcription comme tâche en L2). La méthode dite de la gouvernante a longtemps imposé l'idée que l'on ne pouvait bien apprendre une langue qu'au contact de natifs, mais ce que montre notre étude, c'est qu'un professeur non-natif bien formé peut être meilleur pour faire de l'oral qu'un natif non formé.

Pour conclure, un travail de prise de conscience de la langue parlée par l'étude sur corpus, à la fois à travers le contact direct avec les données, mais aussi à travers une réflexion sur la pédagogisation de celles-ci, telle est selon nous la direction que doit prendre la formation d'enseignant pour améliorer l'enseignement du français parlé. L'expérience collective de LANCOM 1 et l'exemple de Mme G tendent à montrer que le recours à des corpus dans une formation d'enseignants n'est productif pour les compétences orales que s'il se fait par le biais d'une pragmatique contrastive, de nature à la fois réflexive et participante.

BIBLIOGRAPHIE

- Andersen, H. L. (2006). Authenticity and textbook dialogues. In: H. L. Andersen, K. Lund, & K. Risager, (éds.), *Culture in language learning* (pp. 87-104). Aarhus: University Press.

- Bergounioux, G., Baraduc, J. & Dumont, C. (1992). L'étude Socio-Linguistique sur Orléans (1966-1991), 25 ans d'histoire d'un corpus. *Langue française*, 93, 74-93.
- Bertaux, D. (1997). *Les récits de vie*. Paris: Nathan.
- Bescherelle (2012). *La conjugaison pour tous*. Paris: Hatier.
- Binon, J. & Desmet, P. (2006). La reconfiguration et formacom. Une tentative de renouveler la formation didactique initiale et continue des enseignants de français langue étrangère ou seconde (FLES). *Communication présentée au Congrès européen de la FIPF* (Vienne, 2-5 nov. 2006).
- Blampain, D., Klinkenberg, J.-M., Wilmet, M. & Goosse, A. (1997). *Le français en Belgique*. Louvain-la-Neuve: Duculot.
- Blanche-Benveniste C. (1997). *Approches de la langue parlée en français*. Paris: Ophrys.
- Bouchard, R. (2005). Le cours, un événement oralographique structuré: étude des inter-actions pédagogiques en classe de langue et au-delà ... *Le Français dans le monde: recherches et applications*, juillet 2005, 64-74.
- Chambers, A. (2009). Les corpus oraux en français langue étrangère: authenticité et pédagogie. *Mélanges CRAPEL*, 31, 15-33. http://www.atilf.fr/IMG/pdf/melanges/01_Chambers.pdf (15.05.2013).
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier.
- Coveney, A. (1996). *Variability in spoken French. A sociolinguistic study of interrogation and negation*. Exeter: Elm Bank Publication.
- Debrock, M., Mertens, P., Truyen, F. & Brosens, V. (2000). *ELICOP (Etude Linguistique de la Communication Parlée): constitution et exploitation d'un corpus de français parlé automatisé. Preprints van het Departement Linguistiek*, 172. Université de Louvain – K.U. Leuven.
- Delahaie, J. (2008). *Français parlés et français enseignés. Analyses linguistiques et didactiques de discours de natifs, de non-natifs et d'enseignants*. Thèse de doctorat. Université Paris Ouest Nanterre La Défense.
- Delahaie, J. (2009). Oui, voilà ou d'accord? Enseigner les marqueurs d'accord en classe de FLE. *Synergies pays scandinaves*, 4, 17-34.
- Delahaie, J. & Flament-Boistrancourt, D. (à paraître). Questions de variation: autour de quelques locutions méconnues de l'oral, niveau, par rapport à et en termes de. *Corpora and Language in Use*. Louvain: Presses Universitaires de Louvain.
- Detey, S., Durand, J., Laks, B. & Lyche S. (éds.) (2010). *Les variétés du français parlé dans l'espace francophone. Ressources pour l'enseignement*. Paris: Ophrys.
- Flament-Boistrancourt, D. & Debrock, M. (1996). Le corpus LANCOM: bilan et perspectives. *ITL – Review of Applied Linguistics*, 111-112, 1-36.
- Flament-Boistrancourt, D. & Debrock, M. (1997). L'acte de question dans des scènes d'embauche à partir d'un corpus différentiel (francophones vs. apprenants néerlandophones de français). *Revue Parole*, 2, 93-120.
- Flament-Boistrancourt, D. & Cornette, G. (1999). Bon français ou vrai français? Une étude de l'acte de question menée à partir d'un extrait du corpus LANCOM: les scènes dites du baby-sitting. *Travaux de linguistique*, 38, 119-153.
- Flament-Boistrancourt, D. (2001a). Jeux de rôle et discours d'enseignants en Belgique néerlandophone: analyse d'un double bind ordinaire. *Langue française*, 131, 66-88.
- Flament-Boistrancourt, D. (2001b). Pragmatique et approche communicative: la contribution du corpus LANCOM. *Le français dans le monde, recherches et applications*, juillet 2001, 143-170.
- Flament-Boistrancourt, D. (2004). Quelle contribution un corpus natif/ non natif est-il susceptible d'apporter à la linguistique et à l'enseignement du français L2: le cas de LANCOM? *Pré-textes franco-danois*, 4, 45-69.
- Flament-Boistrancourt, D. (2008). Prix et salaires en français à la lumière du corpus LANCOM: des chiffres ou des thèmes? In: C. Parpette & M.-A. Morel (éds.), *L'oral en représentation(s): décrire, enseigner, évaluer* (pp. 91-108). Louvain: Editions Modulaires Européennes.

- Francard, M. (1995). Nef des fous ou radeau de la Méduse? Les conflits linguistiques en Belgique. *Linx*, 33, 31-46.
- Gadet, F. (1989). *Le français ordinaire*. Paris: Armand Colin.
- Grevisse, M. (1936 [éd. Goosse, A. 1993]). *Le bon usage. Grammaire française*. Paris, Louvain-la-Neuve: Duculot.
- Kerbrat-Orecchioni, C. (1996). *La conversation*. Paris: Seuil.
- McCarthy, M. (2004). *Touchstone: from corpus to coursebook*. Cambridge: Cambridge University Press. <http://www.cambridge.org/us/esl/Touchstone/teacher/images/pdf/CorpusBookletfinal.pdf>, (20.07.2011 via Internet Archive).
- Mel'čuk, I. (1993). La phraséologie et son rôle dans l'enseignement/apprentissage d'une langue étrangère. *Etudes de Linguistique Appliquée*, 92, 82-111.
- Roulet, E. (1999). Une approche modulaire de l'enseignement-apprentissage de la compétence discursive. *Actas do 4º Encontro Nacional do Ensino das Línguas Vivas no Ensino Superior em Portugal, Faculdade de letras da Universidade do Porto*, 19-27.
- Tyne, H. (2009). Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL*, 31, 91-111. http://www.atilf.fr/IMG/pdf/melanges/05_Tyne.pdf (16.05.2013).
- Tyne, H. (2012). Corpus work with ordinary teachers: data-driven activities. In: A. Boulton & J. Thomas (éds.), *Teaching and language corpora: input, process and product*. Brno: Masaryk University Press.

Manuels

- De vive voix*. Mauger, M.-T., Argaud, M. & Neveu, P. (1972). Paris: Didier-Crédif.
- Fransk Grammatik*. Andersen, H.-L. & Fristrup, D. (2005). Copenhagen: Alinea.
- Grammaire 2000*. Spiegeleer De, J. et al. (1995, 11^e éd.). Anvers: Uitgeverij Plantijn.
- Grammaire pratique, une grammaire française pour néerlandophones*. D'Haene, S. & De Rammelaere, P. (1983). Anvers: Standaard Educatieve Uitgeverij.
- Grammaire Trajet: Franse basisgrammatica voor Nederlandstaligen*. Raes, M.-A. & De Clercq, F. (1994, 3^e éd.). Kapellen: Uitgeverij Pelckmans.
- Les Orléanais ont la parole*. Biggs, P. & Dalwood, M. (1976). Londres: Longman.
- Parcours bleu: Frans 3^{de} jaar SO*. Arens, C. et al. (1999). Kapellen: Uitgeverij Pelckmans.
- Touchstone*. McCarthy, M., McCarten, J. & Sandiford, H. (2005). Cambridge: Cambridge University Press.

Annexe 1

Conventions de transcription utilisées pour les données de LANCOM 1 et 2

\$ brouhaha (voix) ou bruits (chaises, portes)
 () interprétation probable de ce qui a été dit
 / élément non prononcé
 : syllabe allongée
 = pause courte
 = = pause plus longue
 _____ soulignement: chevauchements de parole

Annexe 2

Jeux de rôles à l'agence de voyage et évaluation (LANCOM 2): présentation des données

	Mme G	Mr L		Mme C	
	13-14 ans ^e	15-16 ans	13-14 ans	14-15 ans	13-14 ans
nb de jeux de rôles	14	10	11	3	5
nb de mots	5.735	4.901	3.983	3741	2.107
durée totale	43'11"	34'02"	32'32"	25'43"	15'57"
nb d'évaluations	15	5	10	0	5
nb de mots	5.865 mots	1.460	6.706	0	2.434
durée des évaluations	53'42"	14'52"	76'25"	0	20'37"
Total mots retranscrits	11.600	6.361	10.689	3.741	4.541
Total durée	96'52"	48'54"	108'57"	25'43"	36'34"

Corpus linguistics and data-driven learning: a critical overview

Alex BOULTON

Université de Lorraine et Crapel, ATILF (UMR 7118)
44 avenue de la libération, 54000 Nancy, France
alex.boulton@univ-lorraine.fr

Henry TYNE

Université de Perpignan Via Domitia et VECT (EA 2983)
52 avenue Paul Alduy, 66000 Perpignan, France
henry.tyne@univ-perp.fr

L'utilisation de corpus s'est avérée intéressante dans de nombreux domaines depuis plus de vingt ans, notamment en didactique des langues. Un état de l'art des études publiées dans ce domaine nous permettra de constater que si les résultats sont en général assez positifs (l'apprentissage a lieu), l'exploitation de corpus reste une activité marginale. Si, en général, de telles recherches en didactique des langues ne donnent pas toujours des résultats fulgurants, on pourrait faire l'hypothèse que les approches sont plus ou moins adaptées à certains types d'apprenants, ou à certains types d'activité, et qu'elles contribuent au bagage général des outils et des techniques disponibles. Qui plus est, avec une approche sur corpus, on pourrait de même croire que les étudiants deviennent tout simplement des apprenants plus conscients de leur apprentissage grâce au travail effectué. Nous suggérons que l'apport des corpus pour l'apprentissage des langues offre ainsi de nombreuses possibilités.

Mots-clés: Corpus, apprentissage sur corpus, pratiques ordinaires d'enseignement-apprentissage, processus, TIC

1. Introduction

Corpus linguistics is not simply a recondite field of research within linguistics: it affords practical methodologies and tools to further the study of all aspects of language use (e.g. Meyer 2002: 11-28; McEnery *et al.* 2006: 8), not least in areas related to teaching and learning foreign or second languages (L2). Here, the impact of corpora has been termed "revolutionary" in the truly Kuhnian sense of the term (e.g. McCarthy 2008: 564). Corpus studies have helped to show how language is really used in authentic communicative situations (Gilmore 2007), and have been influential in informing dictionaries, grammar books, usage manuals, syllabuses, evaluation methods, and course materials. It is also possible for teachers and learners to make direct use of corpora themselves in what Johns ([1990] 1991) has called "data-driven learning" (DDL). DDL was considered "radical" (Johns 1988: 20) then; more than two decades on, it still may be for those who believe it gives learners total control of what to look for and how, with one thing leading to another in an unplanned manner (cf. Bernardini 2000).

However, DDL has come to mean different things to different people and is not an all-or-nothing affair: learners can do many things with corpora, and elements of DDL can be found along a continuum between highly-controlled, closed exercises and the unplanned or 'serendipitous' approach mentioned above (cf. the "soft" and "hard" versions described by Gabrielatos 2005: 17). Indeed, it can cast the learner as "researcher" (Johns 1988: 14) involved at all stages, even creating their own corpora giving rise to pedagogical focus on the dynamic aspects of corpus creation and fieldwork, not just on end-state consultation (cf. Koester 2010). In this sense, the potential for pedagogical uses of corpora is "limited only by the imagination of the user", whoever that may be (Breyer 2006: 162). Given such a range of uses and interpretations, we opt for a broad definition of DDL here, such as that proposed by Gilquin and Granger (2010: 359): "data-driven learning (DDL) consists in using the tools and techniques of corpus linguistics for pedagogical purposes" – whatever the corpus linguistic tools and techniques, and whatever the pedagogical purposes.

Although DDL has generated considerable research interest over the years, it has yet to extend very far beyond the research community. Römer (2009: 92) has highlighted what she sees as the need to undertake a "corpus mission" to increase the visibility of DDL. Obviously, any such initiative should attempt to target both the learners and teachers involved. However, at a more practical level, it would be difficult to train more than a tiny fraction of them directly or produce more than a handful of resources within a reasonable time-span. It may thus be more effective to target publishers and decision-makers further upstream (for teacher training, syllabuses and curricula, examinations, etc.) as recommended by McCarthy (2008). It may also be possible to connect DDL better with existing teaching and learning activities (Tyne 2012), thus promoting it as 'ordinary' practice rather than focusing only on the most radical procedures which may not be suited to many 'ordinary' teaching / learning environments (Mukherjee 2006).

Section 2 of this paper briefly summarises the general arguments in favour of DDL before moving on to the empirical evidence. Section 3 looks at reaching teachers and learners through corpora, tools and materials incorporating aspects of DDL. In section 4, an attempt is made to show that DDL doesn't have to be presented as "revolutionary" (Johns [1990] 1991: 27), and that overstating this may in fact serve to hinder the spread of such activities. This section focuses on a particular difference in underlying philosophy between 'end-state' corpus use and more inclusive, dynamic approaches to DDL. To this end, we discuss the merits of cases where learners are involved in what we will loosely term "corpus work" (Tyne 2012), including many learning opportunities which are quite compatible with a number of existing activities and exercises that form the common stock of practices already used by

language teachers in various educational contexts. Finally, in section 5 we look at arguments in favour of a broader interpretation of DDL, highlighting how data can drive learning through various activities, whether or not it happens to derive from a 'corpus' in the modern sense of large collection of authentic text in electronic format.

2. Arguments and evidence for DDL

Johansson (2009: 42) makes the useful point that corpora "should not be used in language teaching just because we now have this wonderful tool and would like to apply it in language teaching as well". This type of argument has been levelled at pedagogical applications of information and communications technology (ICT) in general as Salaberry (2001) reminds us; it certainly seems axiomatic that "practical developments should not be technology-driven, but rather theory- and pedagogy-driven" (Chambers *et al.* 2004: 1). In other words, what are needed are sound arguments in favour of DDL, supported by some solid empirical evidence of its effects.

A considerable number of papers (including some in this volume) have outlined the theoretical and pedagogical benefits of DDL, and hardly need repeating here. Amongst other things, DDL is alleged to enhance cognitive and metacognitive skills, increase sensitivity to authentic language use, provide an interactive approach to constructivist discovery learning, foster motivation especially through individualisation, promote reusable and transferable skills, favour autonomy for life-long learning, and correspond largely to current theories of second language acquisition (specifically, see Hoey 2005 on "lexical priming"; Taylor 2012 on the "mental corpus"). Although doubts have been raised about some of these issues – the advantages, for example, of authentic documents (Widdowson 2000), constructivism (Matthews 2003), learner-centred, discovery-learning approaches (Kirschner *et al.* 2006), task-based instruction and noticing (Swan 2005) – many are nevertheless basic tenets of contemporary educational systems around the world, and will therefore not be further debated here; see Boulton (2009c) for a discussion. For the time being, if we simply assume the positive arguments hold, this begs the question of why DDL has not attracted a wider following, and specifically whether it translates well into actual practice.

In a review of language teaching methodologies, Nunan (2007) suggests that the zeitgeist of the 1970s and 1980s allowed the emergence of a number of methods (from Suggestopedia to the Silent Way) that were "data free", drawing sustenance from rhetoric rather than empirical support" (Nunan 2007: 9). Yet he goes on to say that today, decision-makers typically require "some kind of evidence" before embracing innovation. In the case of DDL, it is commonplace to lament the lack of empirical evaluation of its actual use. However, a recent survey by Boulton (2010c) finds a number of studies which

do provide empirical assessment of learning outcomes from DDL approaches¹. Without exception, the research papers reviewed by this author report that some learning can and does take place. Furthermore, DDL is not only *effective* in this sense, but it is also consistently found to be as *efficient* as traditional teaching practices, and generally more so. Although the evidence from many individual papers is less conclusive than might perhaps be hoped, the number of studies pointing in broadly the same direction justifies cautious acceptance of the idea that corpus consultation is useful in language learning. This lead is currently being pursued in a meta-analysis conducted by Cobb and Boulton (in preparation), which finds a large effect size over 21 separate studies where sufficient data was available.

Many of these studies are concerned to show that DDL is *possible*, with an emphasis on quantifiable data which are easier to gather from very specific language items and comparatively short-term outcomes in experimental conditions. However, as we have seen above, the main advantages generally attributed to DDL reflect long-term underlying changes in behaviour, language awareness, sensitivity, autonomy, etc., which are inevitably more difficult to evaluate given the complex nature of language learning in general. Partly for this reason, no doubt, many of the studies that do look at learners' behaviour rely on qualitative instruments such as interviews from which it is often impossible to draw any quantitative data at all (e.g. Sun 2003), or at best raw percentages concerning learners' attitudes towards DDL averaged from Likert-scale questionnaires (e.g. Varley 2009). This is not to question the validity of qualitative research, nor the valuable insights it can afford into the processes involved in DDL. However, Richards (2009) reports that qualitative studies account for less than a quarter of those published in applied linguistics journals, with the field as a whole showing a marked preference for quantitative studies supported by statistical analysis.

Occasional studies do allude to increased language awareness and sophistication in dealing with authentic data or longer-term changes in learners' behaviour, especially in using corpora as a reference resource. For writing, Yoon (2008) reports that students claimed increased confidence and autonomy as well as language awareness resulting from their corpus work, while a third of the learners in Gaskell and Cobb's study (2004) became persistent independent users and claimed they would continue in the future. O'Sullivan and Chambers (2006) similarly found many of their students using corpora for translation purposes also intended to continue with their use. However, following Landure (this volume), we would add that there is clearly a need for learners to have some idea of the benefits of taking up new techniques and to have access to training in new methods (though see

¹ An updated supplement now extends to 116 empirical evaluations of corpus use in learning and teaching; see <http://bit.ly/STZegS> ('autres utilisateurs' > 'connexion anonyme', 16.05.2013).

Frankenberg-Garcia 2005 – following paragraph). Of course, there is also a difference between such students' declarations or stated intentions and what actually happens, with little work being carried out to examine the question (cf. Chambers 2007). Charles (2012a) does, however, find that, by the end of her course, over half her students are using a corpus at least once a week for their own purposes; and provisional results from a follow-up study suggest that a year later, half of students are still using a corpus at least once a week (2012b).

Very few studies evaluate long-term effects (such as Delahaie & Flament-Boistrancourt this volume), and those that do often report equivocal results. For example, Hafner and Candlin (2007) find learners rapidly resorting to earlier learning practices in preference to concordancing, while Frankenberg-Garcia (2005) shows that training does not reduce a marked preference for specialist-mediated resources such as dictionaries over materials involving greater autonomy on the part of the learner – specifically corpora. In a rare example of wider benefits, Allan (2006) and Johns *et al.* (2008) independently report that their learners, after initiation to DDL, scored better on items not specifically covered in their course. But these potentially major findings are incidental to the original experimental aims, and are in need of replication in controlled studies. To date, the only attempt along these lines appears to be Boulton (2011c), who finds some evidence that learners improved their noticing skills following corpus work over a medium-term period.

One issue arising from quantitative studies in applied linguistics is that they tend to treat all learners as equal, concealing the diversity of the individuals involved. To put it another way, if average outcomes are not particularly high, it may still be that some learners derive great benefit from DDL (and others little, if any). Surprisingly, little work has been conducted to see which learners DDL may be most suited for. It is generally assumed that it is only appropriate for small groups of advanced, sophisticated, motivated, autonomous students (Hughes 2010: 407), and these are the populations generally investigated. However, this is likely to be in part at least for logistical reasons, with researchers in university environments having easier access to such learners (see Boulton 2009a). Indeed, the few studies that do test DDL with lower levels of language proficiency seem to be as promising as any – Yoon and Hirvela (2004) even obtain better results with intermediate than advanced learners. Varley (2009) and Boulton (2010a) therefore suggest that DDL may be a useful alternative for learners who have not been successful with the teaching they have been used to in the past.

It has been claimed on a number of occasions that inductive DDL work may be most suited to learners with inductive preferences and learning styles (see discussion in Flowerdew 2009). This much would seem uncontroversial, yet here too empirical studies are rare and often incomplete. For example, Lee

and Liou (2003) and Chan and Liou (2005) both report DDL working best for learners with an inductive preference, but provide no details of the instrument used to test this. Turnbull and Burston (1998) report a case study where a field-independent learner with integrative motivation was considerably more receptive to corpus use than a field-dependent one with only instrumental motivation; but again, no indication is provided of how these styles were assessed. Finally, in a pair of experiments, Boulton (2009b, 2010b) uses a standard psychometric instrument to compare learning styles against learners' receptiveness to DDL and their learning outcomes: while some correlations are found, they tend to be fairly weak overall. An optimistic conclusion might be that DDL is in fact suitable for all types of learner, but this would not explain the variation in how different learners react to DDL. Alternative explanations might be sought elsewhere, for example in past experience (cf. Yoon 2008), aptitude or motivation (Chambers 2005).

To conclude: there are many appealing arguments in favour of DDL, and more empirical evidence than is sometimes supposed. The results as a whole are about as encouraging as might reasonably be hoped in any area of language teaching and learning, though the research is inevitably in need of extension and refinement, particularly large scale, quantitative or mixed-method studies (Chambers 2007) to address the specific long-term advantages that DDL is said to hold. In other words, in addition to looking at specific learning outcomes, we also need to focus on the *processes* involved in a DDL approach (O'Sullivan 2007), and allow for the dynamic nature of individual differences (Boulton 2012). Nonetheless, given the arguments and evidence currently available, the obvious question is how to introduce teachers to corpus work (cf. McCarthy 2008).

3. Reaching teachers and learners: resources and materials for DDL

Innovation in curricula, examinations, teacher-training programmes, and so on is generally decided by institutions, local education authorities, or even at a national level, based on advice by experts who are familiar with current research. While such changes may be imposed from above, they are unlikely to be successful unless accepted by the people directly concerned – learners and teachers (see Delahaie & Flament-Boistrancourt this volume). As long as DDL remains the preserve of the research establishment, teachers are unlikely even to be aware of it: a survey by Borg (2009) of 500 language teachers around the world shows that research-engagement (including reading academic publications) is comparatively rare among practising teachers. Specifically regarding corpus use in teaching, Tribble has been conducting regular on-line surveys since 2000, the most recent of which (2012) attracted nearly 600 responses, mostly from the university sector. Three quarters of respondents use corpora in their teaching in one way or

another, but only a tenth explicitly in data-driven learning activities. Nearly half (42%) claimed access to corpora had changed their teaching "a lot", and only 13% "not much / at all"; reasons for under-exploitation included lack of time or awareness of corpus potential, as well as the need for more relevant corpora and available resources.

These surveys underline the crucial role played by the teacher in deciding classroom activities; as McCarthy (2008: 565) puts it, teachers need to become "central stakeholders" in promoting change. There have been attempts to introduce corpora to both pre-service courses (e.g. Breyer 2009) as well as in-service training (e.g. Mukherjee 2004). However, corpus literacy as reported in these papers is limited to individual institutions on an experimental basis; elsewhere, it is understandable that teachers (and their teacher-trainers) are generally reluctant to spend time on areas that do not directly prepare for the qualification (Breyer 2009), however useful they may be later on. Teachers are also wary of regular "revolutions" which have difficulty living up to promises (Littlejohn 1998: 190), and are thus less likely to be convinced by theoretical research than by access to simple tools and materials that can be used with minimum preparation to make their jobs easier. This section therefore considers some of the issues arising from the resources currently available.

Although large, principled corpora are generally designed with linguistic research in mind, they can also be used in language learning (e.g. Aston 1997). To take a single example, the popular BYU site for free access to on-line corpora is provided by Mark Davies, notably including his interface to the British National Corpus (BNC)² as well as his own corpora for English, Portuguese and Spanish. Davies (personal communication) logs the numbers of different types of users for these corpora; significantly, 42% are from users who claim their main interest to be as language learners, a further 16% as teachers (with the rest divided between researchers, translators and other users). While such resources come across as being rich and full of potential in the eyes of researchers, for the average language teacher in need of easy, reliable methods and tools, they may appear to be quite daunting or, at best, unwieldy or not fitting particular language uses (see comments by teacher A in Tyne 2012 on the use of the *Corpus del español* in communicative language teaching).

Smaller corpora may be more "pedagogically relevant" (Braun 2005), more accessible and easier for learners to use, especially when tailored to the needs of a specific target population (cf. Ghadessy *et al.* 2001; Koester 2010). Moreover, as Thompson (2001) points out, a small corpus also allows manual searching without a concordancer, giving learners access to the individual text

² A selection of websites and corpus resources / tools is given at the end of the bibliography.

as well as the corpus, which may be a more satisfactory way into corpora for many learners (Flowerdew 2009). While this may not be considered typical corpus linguistics in the sense that it does not involve a *large* or even *electronic* collection of texts, it can still make use of a principled collection of texts, and allows learners to engage more readily with the data in a familiar linear presentation in addition to concordancing (see section 4). Elsewhere, Aston (2002) suggests increased involvement by getting students to collect data for the corpus, and Seidlhofer (2000) by using a corpus of language produced by the learners themselves. Mparutsa *et al.* (1991) were among the first to suggest corpora compiled of student textbooks to increase immediate relevance, though of course this raises the question of whether textbooks reflect "authentic" language (e.g. Meunier & Gouverneur 2009). However, the *processes* of corpus consultation will still be the same, and most of the benefits hold. There may then be an argument for grading authentic texts or query output from a corpus, or even using simplified readers as proposed by Allan (2009); the idea is certainly appealing, as graded readers are the most consulted corpus on the Compleat Lexical Tutor (cf. Cobb forthcoming). Multi-modal corpora might also increase pedagogical relevance, with sound and / or video aligned with the transcriptions (e.g. Braun 2007), though technological considerations may prove an issue.

While large corpora can be used for pedagogical purposes, many are not generally available for public use: locating an appropriate corpus was one of the most difficult aspects for Breyer's (2009: 165) teacher trainees. This is, of course, even more the case for languages other than English. However, as Kerr (this volume) and Chambers (this volume) point out, there are a number of corpora available for French as a foreign language (see also Kamber 2011 for a presentation of the Wortschatz corpus for DDL in French). The corpora in question are, though, generally based on Internet sources, which inevitably constrains their scope and relevance – indeed, this would seem to be a general feature for corpus linguistics today (Tognini-Bonelli 2010). While existing corpora may adequately serve a number of purposes in general language description, there is a distinct lack of genres pertaining to the types of communicative situations generally targeted in modern foreign language learning today (cf. the SACODEYL and Backbone projects which are an exception to this observation – see Kerr this volume; Chambers this volume). Frequently, then, a teacher's only access to a truly pedagogically relevant corpus might be via a DIY-approach: a sort of 'if you ain't got it, build it' philosophy.

A number of tools are available to assist with corpus collection (e.g. SketchEngine; BootCat) and subsequent consultation (e.g. WordSmith Tools; AntConc), or in using the web as a type of corpus in itself (e.g. WebCorp; Web Concordancer). But again, most such tools are originally designed with

linguistic research in mind, and not all teachers will find it easy to exploit them for pedagogical aims. This is highlighted in Kosem's (2008) review of a number of common tools for corpus analysis, which finds the majority are not sufficiently user-friendly or adapted for use by teachers or learners. He strongly advocates involving teachers in designing new tools from scratch, a position supported by McCarthy (2008: 565), who promotes the "teacher as lobbyist, lobbying academics and publishers and telling those academics and publishers what it is they want and need of this new technology".

A number of tools have been built by researchers with a strong teaching interest. Examples include Cobb's Compleat Lexical Tutor (see Cobb forthcoming on tools for French), Greaves' Virtual Language Centre, and Kaszubski's IFA Concordancer. These initiatives and many others are extremely valuable, but generally stem from the work of individuals or small groups of researchers with limited resources initially addressing a local need. Truly user-friendly tools designed from the start with broad pedagogical aims in mind will require substantial investment to produce.

Even the best of tools and corpora are likely to encounter a certain number of difficulties, and there are certainly sceptics to their use (e.g. Dellar 2003). Amongst other things, it might be claimed that the 'authentic' language in corpora is not relevant to learners, and does not convey 'reality' (Widdowson 2000); that the quantities of results (especially in the form of truncated concordance lines) lead to drowning in data or a blinkered approach meaning learners do not look beyond the concordances (Kamber 2011); that the often 'messy' nature of the data due to tagging errors makes it necessary to check or even edit concordance outputs before use with learners (see Kerr this volume); that the approach is time-consuming, mechanical and tedious, requiring considerable motivation; that learners are unable to take sufficient responsibility for their own learning, and liable to formulate inappropriate queries and come to wrong conclusions; that the learners (and teachers) require advanced ICT skills and corpus training; that adequately equipped computer rooms are lacking; and so on (see Boulton 2009c for a fuller discussion). Crucially, the majority of these issues relate to practical concerns and logistical barriers to implementing DDL, which suggests it *should* be possible to overcome them. Indeed, we have seen that many empirical studies are at pains to do just this by showing that DDL is possible even in comparatively unpromising circumstances: for example, examining uses of DDL with a minimum of technology (e.g. Boulton 2010a; Tyne 2012), with learners at comparatively low levels of language proficiency, motivation and sophistication (e.g. Hadley 2002), and where the teacher retains substantial control over the various decisions involved (e.g. Koosha & Jafarpour 2006; see also Kerr this volume).

In other words, studies such as these are attempting to show that DDL can be of immediate benefit to far more learners in a far wider range of contexts than is usually supposed. For this to happen, DDL needs to be introduced not only as efficient, but as relevant, practical, ordinary practice that does not dramatically overthrow everything teachers and learners know and are comfortable with (Boulton 2009c). Thomas (2008: 468) is clear on this: "no amount of favourable empirical evidence is going to convince teachers or publishers to involve themselves in the use of corpora. Its proselytization depends on having the tools and techniques used by more teachers".

Help is therefore needed to demonstrate clearly how corpora can be used in language teaching and learning. One of the few practical guides to using corpora in language teaching is Tribble and Jones' (1997) *Concordances in the Classroom*; while full of useful ideas, it remains divorced from any specific corpus or interface, and still requires substantial input from the teacher. The same can be said for a number of on-line courses and websites that provide 'how-to' introductions to using corpora (e.g. Lamy & Klarskov Mortensen 2007). What are needed are exercises, activities and examples related to a specific, available corpus and ready for immediate use (O'Keeffe *et al.* 2007: 247). This is clearly expressed as a main concern by Breyer's (2009) student teachers, who also felt that closed tasks with known outcomes would make life easier, at least at the start.

The simplest way of introducing innovation in general is, in fact, likely to be through paper-based materials (cf. Breyer 2009: 156); indeed, much of Johns' own work (e.g. [1990] 1991) was with printed concordances. Coursebooks can thus be considered a "powerful device" (Littlejohn 1998: 190) in spreading new ideas, and it is reassuring that increasing numbers of materials (for English at any rate) do take account of corpus data (Boulton 2010d). However, in the overwhelming majority of cases the corpus input is all but invisible – as one author of a leading corpus-based coursebook puts it, "teachers and learners should expect that, in most ways, corpus informed materials will look like traditionally prepared materials" (McCarthy 2004: 15). This is understandable, but there is no reason to stop there: Higgins and Johns (1984: 93) suggested nearly 30 years ago that concordances and visibly DDL-like activities could be integrated directly into published materials. This has only occasionally been taken up in coursebooks (e.g. Mohamed & Acklam 1995), slightly more frequently in vocabulary-related materials (e.g. Burdine & Barlow 2008). The only book given entirely to a DDL approach is Thurstun and Candlin's (1997) *Exploring Academic English*, though perhaps the most radical approach is adopted in the Cobuild *Concordance Samplers* series (e.g. Goodale 1995), which provides large numbers of concordances along with ideas for exploitation. Boulton (2010d) discusses these and other paper-based DDL materials, but concludes that very little is currently available. Support for this

can be found in the fact that most researchers create their own materials. Though they may on occasion make these available on line for immediate print-out³, it is clearly not possible to rely on the goodwill of individual enthusiasts to supply materials for a sufficient range of needs and language points, hence the need for involvement from publishers. The DDL materials that are available are typically product-oriented, featuring closed activities with known outcomes (Breyer 2009: 166), with teachers and learners as passive consumers (cf. McCarthy 2008). Their attraction is that they are immediately usable for instant benefits, and may represent a first step towards more independent, autonomous corpus-based work.

DDL and corpus consultation are certainly no panacea. However, they may be representative of "something more than a distinctive technique, in fact of a methodology" (Johns & King 1991: iii-iv). Obviously, while few would wish DDL to be used to the exclusion of all other activity, once introduced, it may be integrated to the regular aspects of classroom practice, and may even have more extensive knock-on effects on classroom activity. The following section outlines some of our studies of corpus-related language learning activities in a variety of situations. The aim here is to show that the learner-as-researcher approach (Johns 1988) inherent in DDL may rather be viewed as an extension of ordinary practice insofar as it builds on many aspects of existing language teaching and learning methodology. We go on to discuss these uses, arguing that, although in some ways they appear to be quite different from conventional DDL activities, they can be conceived within a general 'language-data-for-learning' framework, focusing on certain methodological aspects underpinning corpus linguistics.

4. Processes in corpus work: involving learners in corpus creation

Section 2 presented some evidence that the *type* of enquiry could be a positive factor in learning. In Johns' original formulation of DDL, the enquiry itself comes across as being all-important as learners engage directly with corpus data, "the underlying assumption being that effective language learning is a form of research" (Johns [1990] 1991: 30). As O'Sullivan (2007) observes, interest in the benefits of enquiry are the result of a shift towards process-oriented methods, couched in cognitive constructivist theory. In this section we look at some case studies that relate to the dynamic processes involved in learning with corpora, concentrating on corpus creation in various forms. These studies illustrate an alternative means to applying problem-solving or task-based methodologies to learning with corpora, concentrating on the 'homemade' approach which, as Aston (2002) observes, offers many

³ Among the best known are Tim Johns' kibbitzers, notes based on 77 real examples of problems encountered in students' writing and explored using a concordancer; 34 are currently available on Mike Scott's webpage at <http://www.lexically.net/TimJohns/index.html> (16.05.2013).

advantages for learning, favouring ownership, design control, critical awareness, and so on.

A study by Tyne (2012) looked at corpus use with high school teachers of Spanish in France in daily classroom activities using an emic approach, i.e. looking at what teachers (who are not researchers) do from the inside. In one case, 35 learners (two classes of 17 and 18) were asked to work on short news items of general interest in relation to on-going topic work. The Spanish teacher began by getting students to look at seven 'typical' short texts in French in order to pick out some of the structural characteristics. As Flowerdew (2009) argues, in the light of studies by Charles (2007) and Weber (2001), consciousness-raising activities at the level of text (e.g. acquainting students with individual essays in the case of Weber's 2001 "concordance- and genre-informed approach"), can be a useful starting point in the classroom rather than a bottom-up, concordance-based approach. In the Spanish classroom, identification of typical structures and language forms took place in individual texts and these were immediately collated and spontaneously 'verticalised' in a kind of "blackboard concordancing" (Johns 1993): groups of learners worked on different pages of text and produced "hand concordances" (Willis 1998) of key items for whole group work. The students went on to compare their results in French against six articles in L2 Spanish, repeating the same processes (e.g. picking up on use of the *pretérito indefinido* in Spanish vs. the *passé composé* in French). Students referred to the resulting DIY concordance lines in order to notice recurrent lexico-grammatical features, and set about producing a pastiche in a similar genre. Such an exercise exploits meaningful activities (i.e. text creation) to highlight the typical features of the genre in terms of discourse, lexis, tenses and so on. The small corpus enables work on the data at the level of text (see section 3), as well as making it easier for the learners to manipulate and "authenticate" the corpus (cf. Braun 2005), and the processes (i.e. gathering texts, bringing them together, noticing forms together, etc.) emphasise key features in preparation for any future corpus work.

Another paper by Tyne (2009a) highlights the benefits of having learners, admittedly advanced, collect and transcribe data. This study shows how learners become acquainted with sociolinguistic features in L2 French (variable *ne* deletion, liaison, lexis, etc.) through intense and 'physical' contact with the data: students were able to discover features they did not (consciously) know before starting the investigation. In the words of the learners themselves, transcribing authentic speech was considered "a challenge", "quite time-consuming and frustrating at times", but rewarding for most who found it "definitely very satisfying" and "a real learning experience". Transcription was found to encourage noticing in a way that simply hearing language does not, allowing students to "really engage with the language" as

they "get to hear French, how it is really spoken". As Koester (2010) observes, involving students in fieldwork, transcription and corpus compilation allows a greater focus on the qualitative, human dimension of language data. Moreover, in other studies on transcription as a learner exercise (often carried out in the form of self-transcription), we find evidence of enhanced noticing (Lynch 2001; Stillwell *et al.* 2010; Mennim 2012). In Tyne (2009a), sociolinguistic variation (cf. Conrad 2004) was made apparent as the learners encountered it first-hand. Nearly all students felt their French had improved as a result of having done transcription: they encountered "new vocabulary and phrases", and "discover[ed] how different words / registers etc. are used in real language, something which you cannot always get from textbook study".

While the development of spoken skills has come to the fore in communicative language teaching, spoken corpora have (chiefly because of difficulties in accessing and transcribing data; see Meyer 2002: 56) yet to make an impact in terms of pedagogical applications (Mauranen 2004). Oral communication typically occupies an important place in most syllabuses but paradoxically is often absent from corpus-informed learning materials or methodologies (although it is worth pointing out that the rare examples of corpus-informed materials for French are all oral-based – see Delahaie & Flament-Boistrancourt this volume). In a secondary school setting, Tyne (2012) looks at how a Spanish teacher chose to build on existing listening comprehension exercises in class, introducing transcription of material recorded by the students themselves in the form of interviews with the Spanish language assistant in order to build up a corpus of spoken Spanish. The study finds that not only is transcription a simple means of making sure that students have actually listened in detail to given extracts as homework, it also serves to draw attention to detail by bringing the learners into the data – they cannot just superficially adopt an 'I've done that' approach to listening. Importantly, the fact that each student is asked to transcribe several minutes of speech is a means of creating a collaborative local corpus – important since it includes 'meaningful' data (i.e. data that the learners can relate to and feel for), and can serve to target better specific objectives (Aston 2002; Koester 2010). All 25 students in this study felt that transcription was a fruitful way to improve their Spanish in one way or another, commenting in particular on the attention to detail (cf. Lynch 2001) and discovery of 'new' things (cf. Tyne 2009a) as well as new ways of working.

5. Getting into DDL: broadening horizons

In his discussion of the issues arising in applying corpora to language teaching, Sinclair (2004: 288) claims that "both teacher and student can make use of a corpus right away, with only a modest few hours of orientation". He adds that "you will need a computer of normal performance, a corpus and some query software" (*ibid.*). If it really is so straightforward, one might expect corpora to have become commonplace in language teaching and learning, though this is clearly not the case as we have seen. In fact, while computers have become fairly widespread in education in many parts of the world (though by no means available in all language classrooms), corpus investigation certainly has not. Mauranen (2004) reminds us just how difficult it is to introduce "corpus thinking" into teaching. Few language teachers are corpus linguists, but they are in many cases involved in activities which require data gathering and compilation: recording, listening, transcribing, assembling collections of texts, etc. In the first two high school studies documented in the previous section, it is important to note that the individuals concerned (see Tyne 2012) are 'ordinary' teachers rather than corpus linguists. Both of them, having undertaken postgraduate training in applied linguistics (involving a mild amount of corpus work amongst other things), set about using those aspects of corpus work they felt were compatible with their existing tasks (including continued collaboration with more 'conventional' colleagues, more or less allergic to 'new' techniques). These teachers have quite a liberal view of the uses of corpora for language learning and they see student involvement in the processes of corpus work and in the manipulation of data to be a key issue, rather than in the study of existing data. In addition to technical and logistical barriers, teachers and learners may, in general, have difficulty in assuming new roles, in adapting immediately to unaccustomed freedom (cf. Kerr this volume): moving from the 'teacher as knower' with the learner waiting to be taught in a knowledge transmission model to the role of helper, encouraging learners to become more active and take on more responsibility, etc. (Johns [1990] 1991). Or quite simply, the general teaching / learning atmosphere (particularly in the school setting) may not lend itself to less direct techniques.

Returning to the question of whether such work leads to language learning as such, the findings of the studies in the previous section serve to show that it really does depend on what we call language learning and how we measure it. The cases presented here highlight how individual learners work and the processes involved, thus allowing an additional and more individual type of analysis to purely quantitative studies (see section 2 above). For example, for ordinary language teachers, the students' marks are standard measures of achievement, together with overall contentedness, smiling faces, etc. – did they complete the exercise and did they do well? In Boulton (2010c), learning outcomes were specifically aligned with a demonstration of the benefits of

noticing through investigative work. Elsewhere, for Tyne (2009a), achievement was construed in terms of sociolinguistic outcomes: some students didn't know about variable *ne* deletion in L2 French before taking this course – they did afterwards; some students didn't realise that people do not speak using sentences and that even native speakers hesitate, etc. – they did afterwards; and so on. Had more conventional teaching methods worked, the advanced fourth-year students might reasonably be expected to have already mastered such features effectively (see Tyne 2009b). Of course, simply 'being taught' might have enabled them to get straight to the point, but a constructivist approach allowed greater immediacy, personalisation, involvement, and any number of incidental benefits which are difficult to assess in a traditional research paradigm (cf. Boulton 2012).

The notion of discovery in learning is an obvious benefit put forward through corpus work and DDL. For example, in a university setting, Boulton (2011a) reports how 30 English MA students in France produced projects over which they had almost total control. In this distance degree programme, the course introduced the concept of corpus linguistics and outlined the requirements, providing on-line help throughout. However, the students were left largely to their own devices to 'discover' how to use corpora for their own purposes, identify an area for research, choose or compile a corpus, locate and master appropriate software for application to their stated objectives, and so on. Only three students relied solely on existing corpora, the others building their own from scratch, generally from material found on the internet, and most often comparing the results against a large on-line corpus. The students' own research questions were often not explicitly linguistic in nature but reflected their other interests, whether academic (literary, cultural, social, historical, etc.) or personal (musical, ICT, religious, professional, etc.), thus underlining the interdisciplinary applications of corpus linguistic methodologies. Given the variation in the purposes the individual students found for corpora, the emphasis of the course was as much on the process as the product – the students were told they would be evaluated not just on the finished paper, but equally on their initiative and creativity, their strategies in dealing with problems and their reasoning from start to finish. The underlying processes and task-based nature of the assignment involved considerable exposure to and thinking about a wide range of language-related issues (morphology, collocations, chunking, frequency, genre, discourse, etc.), not to mention more general implications for ICT skills, report-writing, critical thinking, problem-solving, and so on.

The type of approach we have detailed in section 4 certainly does go some way beyond traditional acceptations of DDL, but is nevertheless firmly grounded in the basic methods and understandings of corpus linguistics applied to language teaching and learning in accordance with Gilquin and

Granger's (2010: 359) definition given above ("using the tools and techniques of corpus linguistics for pedagogical purposes"). This paper has attempted to show that it constitutes a legitimate area of enquiry in teaching and learning with corpora – understanding what a corpus is, how it is created, and how the very act of creating, even at a very small, classroom exercise level, can in itself be useful for language learning (Maia 1997; Aston 2002). What is new, of course, is that the target teaching-learning activities themselves may not directly relate to the end-state corpus itself but rather to methodological aspects involved in its creation.

6. Conclusion

In this paper, we have outlined some of the ways corpora can be used directly in language learning. The basic idea dates back at least 25 years but, despite a number of powerful arguments, backed up by an increasing body of empirical research, data-driven learning has still to penetrate mainstream classroom practice and to build on existing activities. One problem identified here is the scarcity of appropriate resources, suggesting a lack of communication between researchers, teachers, and decision-makers. Another is the perception that the 'hard-core' approach to autonomous, hands-on concordancing is difficult to reconcile with the reality of many language classrooms with 'ordinary' teachers and learners.

Although difficult to quantify, there is good reason to believe that the dynamic aspects of corpus work can lead to an enormous number of beneficial transfer effects for a tremendous variety of issues connected with language and beyond. By placing the focus on these, a number of objections are no longer viewed as 'problems' for the learner: a small *ad hoc* 'corpus' has the advantage of reinforcing the human dimension of the data and thus bridging an important gap between learner and corpus, accepting that individual activities may focus on a 'peripheral' aspect in corpus linguistics, more connected to methodological concerns than formal descriptive ones. While these are perhaps less dependent on expensive, inaccessible, pedagogically inappropriate tools, they also build on widespread teaching practices already in place – the use of authentic documents, discovery learning, constructivism, autonomisation, ICT, and so on (cf. Johns 1993: 8) – thus facilitating adoption by ordinary teachers and integration into ordinary classroom activity (cf. Boulton & Tyne forthcoming).

REFERENCES

- Allan, R. (2006). Data-driven learning and vocabulary: investigating the use of concordances with advanced learners of English. *Centre for Language and Communication Studies, Occasional Paper, 66*. Dublin: Trinity College Dublin.
- Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal, 63*, 23-32.

- Aston, G. (1997). Small and large corpora in language learning. In: B. Lewandowska-Tomaszczyk & J. Melia (eds.), *Practical applications in language corpora* (pp. 51-62). Lodz: Lodz University Press.
- Aston, G. (2002). The learner as corpus designer. In: B. Kettemann & G. Marko (eds.), *Teaching and learning by doing corpus analysis* (pp. 9-25). Amsterdam: Rodopi.
- Bernardini, S. (2000). Systematising serendipity: proposals for concordancing large corpora with language learners. In: L. Burnard & T. McEnery (eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 225-234). Frankfurt: Peter Lang.
- Borg, S. (2009). English language teachers' conceptions of research. *Applied Linguistics*, 30, 3, 358-388.
- Boulton, A. (2009a). Testing the limits of data-driven learning: language proficiency and training. *ReCALL*, 21, 1, 37-51.
- Boulton, A. (2009b). Corpora for all? Learning styles and data-driven learning. *Proceedings of 5th Corpus Linguistics Conference. University of Liverpool, 20-23 July*. <http://ucrel.lancs.ac.uk/publications/cl2009/> (27.05.2013).
- Boulton, A. (2009c). Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35, 1, 81-106.
- Boulton, A. (2010a). Data-driven learning: taking the computer out of the equation. *Language Learning*, 60, 3, 534-572.
- Boulton, A. (2010b). Consultation de corpus et styles d'apprentissage. *Cahiers de l'APLIUT*, 29, 1, 98-115.
- Boulton, A. (2010c). Learning outcomes from corpus consultation. In: M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (eds.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching* (pp. 129-144). London: Equinox.
- Boulton, A. (2010d). Data-driven learning: on paper, in practice. In: T. Harris & M. Moreno Jaén (eds.), *Corpus linguistics in language teaching* (pp. 17-52). Bern: Peter Lang.
- Boulton, A. (2011a). Bringing corpora to the masses: free and easy tools for language learning. In: N. Kübler (ed.), *Corpora, language, teaching, and resources: from theory to practice* (pp. 69-96). Bern: Peter Lang.
- Boulton, A. (2011b). Data-driven learning: the perpetual enigma. In: S. Goźdz-Roszkowski (ed.), *Explorations across languages and corpora* (pp. 563-580.). Frankfurt: Peter Lang.
- Boulton, A. (2011c). Language awareness and medium-term benefits of corpus consultation. In: A. Gimeno Sanz (ed.), *New trends in computer-assisted language learning: working together* (pp. 39-46). Madrid: Macmillan ELT.
- Boulton, A. (2012). Computer corpora in language learning: DST approaches to research. *Mélanges CRAPEL*, 33, 79-91.
- Boulton, A. & Tyne, H. (forthcoming). Corpus-based study of language and teacher education. In: M. Bigelow & J. Ennser-Kananen (eds.), *Handbook of educational linguistics*. New York: Routledge.
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17, 1, 47-64.
- Braun, S. (2007). Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL*, 19, 3, 307-328.
- Breyer, Y. (2006). My concordancer: tailor-made software for language learners and teachers. In: S. Braun, K. Kohn & J. Mukherjee (eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods* (pp. 157-176). Frankfurt: Peter Lang.

- Breyer, Y. (2009). Learning and teaching with corpora: reflections by student teachers. *Computer Assisted Language Learning*, 22, 2, 153-172.
- Burdine, S. & Barlow, M. (2008). *Business phrasal verbs and collocations*. Houston: Athelstan.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9, 2, 111-125.
- Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In: E. Hidalgo, L. Quereda & V. Santana (eds.), *Corpora in the foreign language classroom* (pp. 3-16). Amsterdam: Rodopi.
- Chambers, A., Conacher, J. & Littlemore, J. (eds.) (2004). *ICT and language learning: integrating pedagogy and practice*. Birmingham: University of Birmingham Press.
- Chan, P-T. & Liou, H-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18, 3, 231-251.
- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6, 4, 289-302.
- Charles, M. (2012a). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31, 93-102.
- Charles, M. (2012b). Student corpus use: Giving up or keeping on? *10th Teaching and Language Corpora (TaLC) International Conference*. Uniwersytet Warszawski, 11-14 July.
- Cobb, T. (forthcoming). A resource wish-list for data-driven learning in French. In: H. Tyne, V. André, C. Benzitoun, A. Boulton & Y. Greub (eds.), *Ecological and data-driven perspectives in French language studies*. Newcastle: Cambridge Scholars Publishing.
- Cobb, T. & Boulton, A. (in preparation). Classroom applications of corpus analysis. In: D. Biber & R. Reppen (eds.), *Cambridge handbook of corpus linguistics*. Cambridge: Cambridge University Press.
- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In: J. Sinclair (ed.), *How to use corpora in language teaching* (pp. 67-85). Amsterdam: John Benjamins.
- Dellar, H. (2003). What have corpora ever done for us? *Developing Teachers*. http://www.developingteachers.com/articles_tchtraining/corporapf_hugh.htm (18.12.2012).
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: a critical evaluation. *International Journal of Corpus Linguistics*, 14, 3, 393-417.
- Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18, 335-355.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal*, 8, 4, 1-35. <http://tesl-ej.org/ej32/a1.html> (18.12.2012).
- Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32, 3, 301-319.
- Ghadessy, M., Henry, A. & Roseberry, R. (eds.) (2001). *Small corpus studies and ELT*. Amsterdam: John Benjamins.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40, 2, 97-118.
- Gilquin, G. & Granger, C. (2010). How can data-driven learning be used in language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 359-370). London: Routledge.
- Goodale, M. (1995). *Concordance samplers 2: phrasal verbs*. London: HarperCollins.
- Hadley, G. (2002). Sensing the winds of change: an introduction to data-driven learning. *RELC Journal*, 33, 2, 99-124.

- Hafner, C. & Candlin, C. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6, 4, 303-318.
- Higgins, J. & Johns, T. (1984). *Computers in language learning*. London: Collins.
- Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London: Routledge.
- Hughes, R. (2010). What a corpus tells us about grammar teaching materials. In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 401-412). London: Routledge.
- Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In: K. Aijmer (ed.), *Corpora and language teaching* (pp. 33-44). Amsterdam: John Benjamins.
- Johns, T. (1988). Whence and whither classroom concordancing? In: T. Bongaerts, P. de Haan, S. Lobbe & H. Wekker (eds.), *Computer applications in language learning* (pp. 9-27). Dordrecht: Foris.
- Johns, T. ([1990] 1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. [Revised version: T. Johns & P. King (eds.), Classroom concordancing. *English Language Research Journal*, 4, 27-45].
- Johns, T. (1993) Data-driven learning: an update. *TELL&CALL*, 2, 4-10.
- Johns, T., Lee, H. & Wang, L. (2008). Integrating corpus-based CALL programs and teaching English through children's literature. *Computer Assisted Language Learning*, 21, 5, 483-506.
- Johns, T. & King, P. (eds.) (1991). Classroom concordancing. *English Language Research Journal*, 4.
- Kamber, A. (2011). Contexte et sens: utilisation d'un corpus écrit dans l'enseignement/apprentissage du FLE. *Travaux neuchâtelois de linguistique*, 55, 199-218.
- Kirschner, P., Sweller, J. & Clark, R. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 2, 75-86.
- Koester, A. (2010). Building small specialised corpora. In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 66-79). London: Routledge.
- Koosha, M. & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL adult learners. *Asian EFL Journal Quarterly*, 8, 4, 192-209. http://www.asian-efl-journal.com/December_2006_EBook.pdf (18.12.2012).
- Kosem, I. (2008). User-friendly corpus tools for language teaching and learning. In: A. Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 183-192). Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Lamy, M-N. & Klarskov Mortensen, J. (2007). Using concordance programs in the modern foreign languages classroom. Module 2.4. In: G. Davies (ed.), *Information and communications technology for language teachers (ICT4LT)*. Slough: Thames Valley University. http://www.ict4lt.org/en/en_mod2-4.htm (18.12.2012).
- Lee, C-Y. & Liou, H-C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning*, 27, 3, 35-56.
- Littlejohn, A. (1998). The analysis of language teaching materials: inside the Trojan Horse. In: B. Tomlinson (ed.), *Materials development in language teaching* (pp. 190-216). Cambridge: Cambridge University Press.
- Lynch, T. (2001). Seeing what they meant: transcribing as a route to noticing. *ELT Journal*, 55, 2, 124-132.
- Maia, B. (1997). Making corpora: a learning process. In: G. Aston, L. Gavioli & F. Zanettin (eds.), *Proceedings of corpus use and learning to translate*. Bertinoro: Università di Bologna. <http://www.sslmit.unibo.it/cultpaps> (04.06.08 via Internet Archive).
- Matthews, W. (2003). Constructivism in the classroom: epistemology, history, and empirical evidence. *Teacher Education Quarterly*, 30, 3, 51-64.

- Mauranen, A. (2004). Speech corpora in the classroom. In: G. Aston, S. Bernardini & D. Stewart (eds.), *Corpora and language learners* (pp. 195-211). Amsterdam: John Benjamins.
- McCarthy, M. (2004). *Touchstone: from corpus to coursebook*. Cambridge: Cambridge University Press.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41, 4, 563-574.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. London: Routledge.
- Mennim, P. (2012). Learner negotiation of L2 form in transcription exercises. *ELT Journal*, 66, 1, 52-61.
- Meunier, F. & Gouverneur, C. (2009). New types of corpora for new educational challenges: collecting, annotating and exploiting a corpus of textbook material. In: K. Aijmer (ed.), *Corpora and language teaching* (pp. 179-201). Amsterdam: John Benjamins.
- Meyer, C. (2002). *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.
- Mohamed, S. & Acklam, R. (1995). *The intermediate choice*. Harlow: Longman.
- Mparutsa, C., Love, A. & Morrison, A. (1991). Bringing concord to the ESP classroom. In: T. Johns & P. King (eds.), *Classroom concordancing. English Language Research Journal*, 4, 115-134.
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In: U. Connor & T. Upton (eds.), *Applied corpus linguistics: a multidimensional perspective* (pp. 239-250). Amsterdam: Rodopi.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: the state of the art – and beyond. In: S. Braun, K. Kohn & J. Mukherjee (eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods* (pp. 5-24). Frankfurt: Peter Lang.
- Nunan, D. (2007). The nineteen-eighties. In: H. Brown, E. Tarone, M. Swan, R. Ellis, L. Prodromou, U. Jung, A. Bruton, K. Johnson, D. Nunan, R. Oxford, C. Goh, A. Waters & S. Savignon, *Forty years of language teaching. Language Teaching*, 40, 1-15.
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL*, 19, 3, 269-286.
- O'Sullivan, Í. & Chambers, A. (2006). Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15, 1, 49-68.
- Richards, K. (2009). Trends in qualitative research in language teaching since 2000. *Language Teaching*, 42, 2, 147-180.
- Römer, U. (2009). Corpus research and practice: what help do teachers need and what can we offer? In: K. Aijmer (ed.), *Corpora and language teaching* (pp. 83-98). Amsterdam: John Benjamins.
- Salaberry, R. (2001). The use of technology for second language learning and teaching: a retrospective. *Modern Language Journal*, 85, 1, 39-56.
- Seidlhofer, B. (2000). Operationalizing intertextuality: using learner corpora for learning. In: L. Burnard & T. McEnery (eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 207-223). Frankfurt: Peter Lang.
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. In: J. Sinclair (ed.), *How to use corpora in language teaching* (pp. 271-299). Amsterdam: John Benjamins.
- 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*, 64, 4, 445-455.
- Sun, Y-C. (2003). Learning process, strategies and web-based concordancers: a case-study. *British Journal of Educational Technology*, 34, 5, 601-613.

- Swan, M. (2005). Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics*, 26, 3, 376-401.
- Taylor, J. (2012). *The mental corpus: how language is represented in the mind*. Oxford: Oxford University Press.
- Thomas, J. (2008). Impatience is a virtue: students and teachers interact with corpus data – now. In: A. Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 463-469). Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Thompson, G. (2001). Corpus, comparison, culture: doing the same things differently in different cultures. In: M. Ghadessy, A. Henry & R. Roseberry (eds.), *Small corpus studies and ELT* (pp. 311-334). Amsterdam: John Benjamins.
- Thurstun, J. & Candlin, C. (1997). *Exploring academic English: a workbook for student essay writing*. Sydney: CELTR.
- Tognini Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 14-27). London: Routledge.
- Tribble, C. (2012). Teaching and language corpora: quo vadis? *10th Teaching and Language Corpora (TaLC) International Conference*. Warsaw: Uniwersytet Warszawski, 11-14 July.
- Tribble, C. & Jones, G. (1997). *Concordances in the classroom (2nd edition)*. Houston: Athelstan.
- Turnbull, J. & Burston, J. (1998). Towards independent concordance work for students: lessons from a case study. *ON-CALL*, 12, 2, 10-21.
- Tyne, H. (2009a). Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL*, 31, 91-111. http://www.atilf.fr/IMG/pdf/melanges/05_Tyne.pdf (16.05.2013).
- Tyne, H. (2009b). Style in L2: the icing on the cake? In: E. Labeau & F. Myles (eds), *The advanced learner variety: the case of French* (pp. 243-268). Bern: Peter Lang.
- Tyne, H. (2012). Corpus work with ordinary teachers: data-driven learning activities. In: A. Boulton & J. Thomas (eds.), *Teaching and language corpora: input, process and product* (pp. 114-129). Brno: Masaryk University Press.
- Varley, S. (2009). I'll just look that up in the concordancer: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22, 2, 133-152.
- Weber, J.-J. (2001). A concordance- and genre-informed approach to ESP essay writing. *ELT Journal*, 55, 1, 14-20.
- Widdowson, H. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21, 1, 3-25.
- Willis, J. (1998). Concordances in the classroom without a computer: assembling and exploiting concordances of common words. In: B. Tomlinson (ed.), *Materials development in language teaching* (pp. 44-66). Cambridge: Cambridge University Press.
- Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12, 2, 31-49.
- Yoon, H. & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, 13, 4, 257-283.

Websites and corpus resources / tools

AntConc: <http://www.antlab.sci.waseda.ac.jp> (18.12.2012).

Backbone: <http://134.2.2.16:8080/backbone-search/faces/search.jsp#> (18.12.2012).

BootCat: <http://bootcat.sslmit.unibo.it> (18.12.2012).

British National Corpus: <http://www.natcorp.ox.ac.uk/> or via BYU site (19.05.2013).

BYU Corpora (BNC, CoCA, Corpus del español, Time Magazine Corpus): <http://corpus.byu.edu> (21.12.2012).

Compleat Lexical Tutor: <http://www.lextutor.ca/> (18.12.2012).

IFA Concordancer: <http://ifa.amu.edu.pl/~ifaconc> (18.12.2012).

Leeds Internet Corpus: <http://corpus.leeds.ac.uk/internet.html> (18.12.2012).

SACODEYL: <http://www.um.es/sacodeyl> (18.12.2012).

SketchEngine: <http://www.sketchengine.co.uk> (18.12.2012).

Virtual Language Centre: <http://vlc.polyu.edu.hk> (18.12.2012).

Web Concordancer: <http://webascorpus.org> (18.12.2012).

WebCorp: <http://www.webcorp.org.uk> (18.12.2012).

WordSmith Tools: <http://www.lexically.net> (18.12.2012).

Wortschatz: http://wortschatz.uni-leipzig.de/ws_fra (18.12.2012).

Compte rendu

Berruto, G. (2012).

***Saggi di Sociolinguistica e Linguistica* (a cura di Giuliano Bernini, Bruno Moretti, Stephan Schmid e Tullio Telmon).
Alessandria: Edizioni dell'Orso.**

"Saggi di sociolinguistica e linguistica", a cura di Giuliano Bernini, Bruno Moretti, Stephan Schmid e Tullio Telmon, raccoglie venti saggi di Gaetano Berruto scritti e pubblicati nell'arco di più di vent'anni, dal 1983 al 2006. La raccolta, ideata come omaggio per i sessantacinque anni dell'autore, nell'impossibilità oggettiva di dar conto della vastissima produzione dello stesso nella sua interezza (353 titoli alla data dell'uscita del libro, più di 370 alla data odierna), sceglie i più rappresentativi, nell'intento di dare una panoramica, se non completa, quantomeno il più ampia possibile dell'attività scientifica dello studioso.

I saggi sono raggruppati in quattro sezioni, corrispondenti ad altrettanti filoni di indagine. Alla prima, titolata "Fondamenti della sociolinguistica", vengono ricondotti sette articoli usciti tra il 1987 e il 2006, riguardanti vari temi di interesse della disciplina. Si parte con le riflessioni su concetti sociolinguistici fondamentali di "Lingua, dialetto, diglossia e dilalia" (apparso originariamente in Holtus & Kramer 1987: 57-81), in cui l'autore propone per la prima volta e "provvisoriamente" (p. 12) l'etichetta dilalia (ma, è ben noto, dopo 25 anni e più il termine è ormai invalso). "Le varietà del repertorio" (Sobrero (a c. di) 1993: 3-36) definisce il repertorio linguistico italo-romanzo medio insieme con l'architettura dell'italiano contemporaneo (già presente in Berruto (a c. di) 1987), nonché la questione del "dialetto italianizzato" (p. 45). "Scenari sociolinguistici per l'Italia del Duemila" (Holtus & Radtke (a c. di) 1994: 23-45) propone alcune "riflessioni futurologiche" (p. 52) sulla situazione sociolinguistica italiana, soffermandosi sui possibili scenari per la sorte dei dialetti. Il tema dei dialetti è mantenuto nell'articolo successivo, "Dialetti, tetti coperture – Alcune annotazioni in margine a una metafora sociolinguistica" (Iliescu *et al.* (a c. di) 2001: 23-40), in cui si esplora la nozione di lingua tetto sottolineandone implicazioni e punti problematici, mentre "Sul parlante nativo (di italiano)" (Radatz & Schlösser (a c. di) 2003: 1-14) riprende l'articolo di Acquaviva (2000) che poneva la questione dell'esistenza di "variazione sistematica fra le competenze di parlanti nativi della stessa lingua" (p. 87). "Quale dialetto per l'Italia del Duemila? Aspetti dell'italianizzazione e risorgenze dialettali in Piemonte (e altrove)" (Sobrero & Miglietta (a c. di) 2006: 101-127) ridiscute a distanza di più di dieci anni quanto proposto in "Scenari sociolinguistici per l'Italia del Duemila", portando dati empirici e

osservazioni qualitative al fine di trarre conclusioni "sullo stato di salute del dialetto all'inizio del Terzo Millennio" (p. 101). Infine, "Sul concetto di lingua mista" (Bombi *et al.* (a c. di) 2006: 153-159) esplora uno degli esiti estremi del contatto linguistico attraverso un'ampia carrellata di esempi.

La seconda parte, "Descrizioni dell'italiano", è composta da cinque articoli, usciti tra il 1983 e il 1986, concernenti l'analisi e la descrizione delle strutture dell'italiano; in particolar modo, sono presi in esame gli aspetti grammaticali sintattici della lingua parlata, il cui studio era allora agli albori. Il primo saggio, "L'italiano popolare e la semplificazione linguistica" (*Vox Romanica* 42, 1983, 38-79), prende in considerazione alcuni tratti morfosintattici della varietà con lo scopo di verificare se e in che modo essi costituiscano semplificazione linguistica, nonché se gli stessi tratti "siano presenti o non nei *pidgins* e/o creoli propriamente detti" (p. 147). "Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica?" (Holtus & Radtke (a c. di) 1985: 120-153), elenca una lista di caratteri trattati in letteratura come "tipici o chiaramente emergenti nel parlato" (p. 185) per tentare di interpretarli alla luce della caratterizzazione della grammatica del parlato. "«Dislocazioni a sinistra» e «grammatica» dell'italiano parlato" (Franchi de Bellis & Savoia (a c. di) 1985: 59-82) introduce, "per dar conto dei fenomeni a cavallo fra sintassi e pragmatica racchiusi nelle dislocazioni a sinistra" (p. 227), il concetto di Centro di interesse o Focus empatico. "Le dislocazioni a destra in italiano" (Stammerjohann (a c. di) 1986: 55-69) analizza il fenomeno speculare al precedente, "frequente soprattutto nel parlato" (p. 231) e relativamente poco studiato. Infine, in "Un tratto sintattico dell'italiano parlato: il c'è presentativo" (Lichem *et al.* (a c. di) 1986: 61-73), il fenomeno indicato dal titolo viene spiegato con un occhio all'*il y a* francese.

La terza parte, "La commutazione di codice", raggruppa alcuni studi, distribuiti sull'arco di quasi vent'anni (1985-2004), sui fenomeni di commutazione di codice ed enunciazione mistilingue. Nei primi due saggi, "«l pulman l-è nen ch-a cammina tanto forte». Su commutazione di codice e mescolanza dialetto-italiano" (in *Vox Romanica* 44, 1985: 59-76) e "Italiano regionale, commutazione di codice e enunciati mistilingui" (Cortelazzo & Mioni (a c. di) 1990: 105-130), si analizza l'alternanza di italiano e dialetto nell'interazione quotidiana in *corpora* di parlato spontaneo di italiano e dialetto, raccolti perlopiù tra Piemonte e Lombardia. In "Struttura dell'enunciazione mistilingue e contatti linguistici nell'Italia di Nord-Ovest (e altrove)" (Wunderli *et al.* (a c. di) 2001: 263-283) vengono stabiliti chiaramente i confini tra commutazione di codice, enunciazione mistilingue e ibridismi. In "Su restrizioni grammaticali nel *codemixing* e situazioni sociolinguistiche. Notazioni in margine al modello MLF" (*Sociolinguistica* 18, 2004: 54-72) si discute il modello teorico sul *codeswitching* elaborato da Carol Myers Scotton (1993, 2002), calandolo anche nella realtà sociolinguistica italiana e italo-romanza.

La quarta e ultima parte raccoglie infine quattro articoli su "L'italiano in Svizzera" usciti tra il 1984 e il 1993 (l'autore ha insegnato all'Università di Zurigo dal 1981 al 1995). "Appunti sull'italiano elvetico" (*Studi Linguistici Italiani* 10, 1984: 76-108) illustra alcune caratteristiche dell'italiano parlato nella Svizzera germanofona "nella sua qualità di terza *Landessprache* e *Amtssprache* della Confederazione elvetica" (p. 346). In "Note sul repertorio linguistico degli emigrati italiani in Svizzera tedesca" (*Linguistica* 31, 1991: 61-79) si analizzano i repertori linguistici degli emigrati italiani di prima e di seconda generazione. "*Fremdarbeiteritalienisch*: fenomeni di pidginizzazione dell'italiano nella svizzera tedesca" presenta e analizza per la prima volta la varietà di italiano pidginizzata in uso tra i lavoratori immigrati in Svizzera, battezzata così da Berruto stesso. Infine, "Italiano in Europa oggi: 'Foreigner Talk' nella Svizzera tedesca" (AA. VV. 1993: 2275-2290) si interroga sulle caratteristiche della varietà di italiano semplificato rivolto a stranieri nella Svizzera alemannica.

In conclusione, "Saggi di Sociolinguistica e Linguistica" ha l'indubbio pregio di raccogliere in un unico volume articoli che, come sottolineato nell'introduzione dai curatori, potrebbero essere al giorno d'oggi di difficile reperibilità. Il campione di saggi selezionato, per quanto giocoforza ristretto, permette di avere un'idea dell'ampio respiro dell'attività scientifica di Gaetano Berruto; inoltre, la divisione in aree tematiche (per quanto di grana grossa) consente di seguire gli sviluppi nel tempo delle idee dell'autore: ne è un esempio il caso degli scenari sulla sorte dei dialetti ipotizzati in "Scenari sociolinguistici per l'Italia del Duemila", ripresi e sviluppati a distanza di più di dieci anni con "Quale dialetto per l'Italia del Duemila? Aspetti dell'italianizzazione e risorgenze dialettali in Piemonte (e altrove)". Molto interessante la scelta di presentare alcuni contributi scelti tra quelli dedicati dallo studioso all'italiano nella Svizzera tedesca, tema piuttosto trascurato all'epoca della loro redazione, tra cui spicca il fondamentale studio sul *Fremdarbeiteritalienisch*.

Necessariamente, la scelta obbligata di privilegiare determinati filoni di ricerca dell'autore va a discapito di altri temi cari allo stesso, come la semantica e la linguistica applicata; mancanza che i curatori stessi non mancano di sottolineare con rammarico nella prefazione. Come già detto, però, dar conto di una produzione tanto vasta in un unico volume sarebbe risultato alquanto ostico. Sempre per ragioni di spazio, non è stato possibile riportare alla fine del volume l'elenco delle pubblicazioni scientifiche dell'autore.

Il volume risulta dunque, nonostante i limiti inevitabili in quanto intrinseci al formato della raccolta, uno strumento tanto utile quanto prezioso per avere sottomano un estratto della produzione in ambiti diversi della linguistica e della sociolinguistica di uno dei più importanti linguisti italiani.

BIBLIOGRAFIA

- AA.VV. (1993). *Omaggio a Gianfranco Folena*. Padova: Editoriale Programma.
- Acquaviva, P. (2000). La grammatica italiana: il lavoro comincia adesso. *Lingua e Stile* 2/2000, 249-272.
- Berruto, G. (1987). *Sociolinguistica dell'italiano contemporaneo*. Roma: La Nuova Italia Scientifica.
- Bombi, R., Cifoletti, G., Fusco, F., Innocente, L. & Orioles, V. (a c. di) (2006). *Studi linguistici in onore di Roberto Gusmani*. Alessandria: Edizioni dell'Orso.
- Cortelazzo, M. A. & Mioni, A. M. (a c. di) (1990). *L'italiano regionale. Atti del XVIII Congresso della S.L.I. (Padova - Vicenza, 14-16 settembre 1984)*. Roma: Bulzoni.
- Franchi de Bellis, A. & Savoia, L. M. (a c. di) (1985). *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive. Atti del XVII Congresso della S.L.I. (Urbino, 11-13 settembre 1983)*. Roma: Bulzoni.
- Holtus, G. & Kramer, J. (a c. di) (1987). *Romania et Slavia adriatica. Festschrift für Zarko Muljačić*. Amburgo: Buske.
- Holtus, G. & Radtke, E. (a c. di) (1985). *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübinga: Narr.
- Holtus, G. & Radtke, E. (a c. di) (1994). *Sprachprognostik und das 'italiano di domani'. Prospettive per una linguistica 'prognostica'*. Tübinga: Narr.
- Iliescu, M., Plangg, G. A. & Videsott, P. (a c. di) (2001). *Die vielfältige Romania. Dialekt – Sprache – Überdachungssprache. Gedenkschrift für Heinrich Schmid (1921-1999)*, Vigo di Fassa/Innsbruck: Institut Cultural Ladin/Institut für Romanistik.
- Lichem, K., Mara, E. & Knaller, S. (a c. di) (1986). *Parallela 2. Aspetti della sintassi dell'italiano contemporaneo*. Tübinga: Narr.
- Myers-Scotton, C. (1993). *Duelling languages. Grammatical structure in codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, C. (2002). *Contact linguistics. Bilingual encounters and grammatical outcomes*, Oxford: Oxford University Press.
- Radatz, H.-I. & Schlösser, R. (a c. di) (2003). *Donum Grammaticorum. Festschrift für Harro Stammerjohann*. Tübinga: Niemeyer.
- Sobrero, A. A. (a c. di) (1993). *Introduzione all'italiano contemporaneo. La variazione e gli usi*, Roma-Bari: Laterza.
- Sobrero, A. A. & Miglietta, A. (a c. di) (2006). *Lingua e dialetto nell'Italia del duemila*. Galatina: Congedo.
- Stammerjohann, H. (a c. di) (1986). *Tema-rema in italiano / Theme-rheme in Italian*. Tübinga: Narr.
- Wunderli, P., Werlen, I. & Grünert, M. (a c. di) (2001). *Italica - Raetica - Gallica. Studia linguarum litterarum artiumque in honorem Ricarda Liver*. Tübinga-Basilea: Francke.

Ilaria Fiorentini

Università di Pavia / Freie Universität Bozen

ilafiore@libero.it

Compte rendu

Pugnière-Saavreda, F.; Sitri, F.; Veinard, M. (2012).

L'analyse du discours dans la société. Engagement du chercheur et demande sociale.

Paris: Honoré Champion.

On le sait depuis longtemps, l'analyse du discours est moins une discipline qu'un carrefour disciplinaire. En ce sens, la parution de "l'Analyse du discours dans la société" apporte une pierre de plus dans un édifice intrinsèquement fragile, car subissant des tiraillements entre diverses tendances d'analyse voire certaines chapelles. A cette difficulté d'ordre épistémologique, il faut ajouter celle de l'inscription de l'analyse du discours (AD) dans la société. Cette dernière est source de diverses tensions, car l'AD partage avec d'autres disciplines un regard sur la société qu'elle enrichit et qui l'enrichit réciproquement, adopte parfois une posture critique qui fait prendre au chercheur le risque d'être un acteur dans cette même société, subit des demandes sociales impliquant, à divers titres, le chercheur.

Ce nœud de tensions mais aussi de dialogues est au final l'objet de ce recueil de 26 chapitres divisés en trois parties et six sections. C'en est même le dénominateur commun, tant les contributions apparaissent éclatées entre non seulement différentes approches et différents genres de discours analysés, mais aussi entre des chercheurs de différents pays, dont certains sont confirmés et d'autres en devenir, et dont les visions de l'engagement, sur le plan éthique ou idéologique, sont également variables. "Un panorama riche et varié des pratiques de l'analyse de discours aujourd'hui", promettent les auteurs de cet ouvrage (p. 20): ce pari est réussi, mais les interrogations sur la cohérence de l'ensemble se posent dans le même temps.

Parvenir à mêler des questions d'engagement, de demande sociale, d'idéologie et d'analyse de corpus allant des règlements d'ateliers à la fin du XIXe siècle jusqu'à l'émission de télévision *l'Hebdo du médiateur* a au moins le mérite de montrer à quel point le champ est actif dans la volonté de se donner les moyens de comprendre le social. A ce titre, l'ouvrage recensé ici vaut moins par sa collection hétéroclite de chapitres – saluons toutefois le travail des éditeurs qui ont restitué par de courtes sections des thèmes communs à deux, trois ou quatre contributions – que par son potentiel en tant qu'ouvrage de consultation sur l'un ou l'autre des aspects de l'analyse du discours. On notera que la question de l'engagement du chercheur promise en sous-titre est nettement moins abordée que la question de la demande sociale.

La première partie de l'ouvrage est consacrée aux débats de nature épistémologique. L'ambition des uns – peut-on proposer, comme le fait un auteur, un protocole d'analyse de discours en quelques pages? – le militantisme des autres, le caractère opaque de certains usages référentiels cabreront certains lecteurs tout en témoignant de certaines tendances en AD: une lecture inspirée de Pêcheux et de Foucault (Guigue) ou une autre incarnée dans la lutte politique (Orlandi). On relèvera des mises au point sur l'important courant de l'analyse du discours au Brésil (Scherer, Petri) et sur le non moins important courant de la Critical Discourse Analysis ou CDA (Petitleclerc).

La deuxième section, sur l'engagement des chercheurs, offre des contributions souvent plus claires; parmi elles, celle de Diane Vincent permet d'étudier l'opposition entre analyse linguistique des discours et analyse sociale reposant sur des discours. Cette habile comparaison illustrée par quelques exemples représentatifs permet de montrer des rapprochements entre des usages de l'AD. La pertinence sociale de la recherche trouve un écho dans la contribution de Saint-George revenant par ailleurs sur la controverse entre AD et CDA et plaidant pour une linguistique engagée.

La deuxième partie du livre est nettement plus orientée sur des pratiques discursives et leurs rapports avec une demande sociale: "L'hebdo du médiateur" (Garric) dont l'analyse montre la récupération par les journalistes des critiques faites à l'encontre de la chaîne ou la question délicate de la représentation des enfants en danger dans les médias (Ramos *et al.*) font l'objet de deux contributions intéressantes en soi. Il n'en demeure pas moins que le rapport avec la demande sociale, s'il existe bien, est un peu ténu, en particulier pour une contribution sur le dispositif technique à l'œuvre dans des blogs (Jeanne-Perrier). La deuxième section de cette partie, sur les discours de formation, présente l'analyse de différents genres discursifs répondant à des demandes sociales: l'e-conférence en relation avec certaines philosophies de l'éducation (progressive, behavioriste, humaniste, radicale) (Rakotonoelina); la lecture dialoguée en classe revenant sur la dichotomie entre approches formelles et approches interactionnelles du récit avant d'établir une typologie des "mouvements" à l'œuvre dans la lecture dialoguée pour aider à mettre en mots et aider à suivre le fil narratif (Carcassonne & Froment); la lettre de candidature enfin avec un accent porté sur la question de la politesse et sur l'analyse des manuels d'instruction pour ce genre discursif (Avram, Salinas). La richesse de ces contributions éclaire bien l'importance de l'articulation entre genres de discours et demande ou pratique sociale.

Avec la troisième partie, portant sur les lieux institutionnels, la question de l'ancrage de l'AD en contexte est abordée essentiellement sous l'angle dialogique. Le domaine juridique, peu exploré à notre connaissance par l'AD,

est pris en charge par Ross Charnock qui montre l'intérêt du corpus des jugements de la Common Law pour l'analyse du dialogisme – les précédents juridiques étant au centre des rendus de jugement, mais aussi, peut-être de manière moins frontale, l'intérêt de la linguistique pour les juristes dans l'évolution diachronique du langage. La notion de dialogisme, dont on sait qu'elle est un de piliers de l'analyse de discours pour Jean-Michel Adam en particulier, est donc, sans surprise, au centre de plusieurs contributions de cette section – fondée sur l'approche des interactions (Carcassonne & Servel, Moulinou, Demaître-Lahaye) – tant elle permet d'abriter l'intrication entre langage, société et mémoire socio-langagière.

Plus technique, la partie suivante, composée de deux contributions, interroge le recours à l'informatique au service de l'AD. D'une part, l'analyse lexicométrique est appliquée au discours judiciaire (Brunel *et al.*), d'autre part, une étude sur l'anaphore résomptive dans l'écrit scientifique (Lundquist *et al.*) est alimentée par les résultats d'un logiciel de navigation textuelle. La question de l'engagement du chercheur comme celle de la demande sociale n'apparaît ici que de manière indirecte.

Enfin, la dernière partie revient sur un concept qui a fait couler de l'encre au sein de l'AD depuis Althusser: l'idéologie. Thierry Guilbert prend en charge de front la question de la demande sociale – posant le problème de "son application [...] et donc de son implication" (p. 385). Défendant une approche multidisciplinaire et critique, Guilbert analyse en fin de parcours des aspects de l'écriture éditoriale en montrant l'évidence avec laquelle se présentent l'idéologie. Cependant, les autres contributions évitent la question de l'idéologie pour mobiliser celle de la représentation au sens grizéen du terme – l'image de la parentalité dans des guides parentaux français et allemands valent le détour (Von Münchow) ou celle de la norme – normes de genre en l'occurrence (Greco).

Malgré des contributions inégales, la richesse du volume édité par Frédéric Pugnière-Saavreda, Frédérique Sitri et Marie Veinard à la suite d'un colloque tenu en 2008 ne fait pas de doute dans sa volonté de faire un état du champ. On demeure frappé par la diversité des approches, ici ancrées dans une perspective clairement internationale, qui incite, pour les optimistes, à se réjouir de la multiplication des analyses de discours, et, pour les pessimistes, à craindre un dialogue impossible, par la faute d'une perte de certaines références et positions théoriques communes.

Thierry HERMAN

Universités de Neuchâtel et de Lausanne
thierry.herman@unine.ch

Compte rendu

Ferrari, A.; Cignetti, L.; De Cesare, A.-M.; Lala, L.; Mandelli, M.; Ricci, C.; Roggia, C. E. (2008).

L'interfaccia lingua-testo. Natura e funzioni dell'articolazione informativa dell'enunciato.

Alessandria: Edizioni dell'Orso.

Il volume recensito costituisce il risultato di un progetto di ricerca pluriennale coordinato da Angela Ferrari (Università di Basilea), che ha visto la partecipazione di numerosi collaboratori afferenti all'Istituto di Italianistica dell'Università di Basilea. L'obiettivo del progetto è stato l'elaborazione di un modello della struttura informativa dell'enunciato scritto, con particolare attenzione all'italiano contemporaneo non letterario (cfr. Ferrari 2003b). I risultati esposti in questo volume assumono un respiro più ampio: viene qui presentato un modello dell'organizzazione semantico-pragmatica del testo scritto, che assegna grande importanza all'articolazione informativa dell'enunciato e alla sua funzione di interfaccia tra il sistema della lingua e il sistema del testo. La prospettiva di analisi della testualità scritta proposta in questo lavoro è emersa inizialmente in diversi lavori della studiosa svizzera (cfr. Ferrari 1995, 2003a, Ferrari & Zampese 2000) e ha trovato una sistemazione coerente nei successivi volumi collettivi, sempre a cura di Angela Ferrari (cfr. Ferrari 2004, 2005a, 2006), legati al progetto di ricerca citato.

L'ipotesi alla base del modello proposto (p. 14) considera diversi aspetti del rapporto tra lingua e testo e della costituzione interna delle unità di riferimento. Il punto centrale dell'ipotesi è la codificazione di valori testuali in alcune strutture linguistiche, più precisamente nella componente semantica di lessico, sintassi e punteggiatura: i sistemi linguistico e testuale sono visti come "due sistemi autonomi governati da regole diverse" (p. 62), che però possono comunicare attraverso un'interfaccia. Questa funzione è svolta, per l'appunto, dall'articolazione informativa dell'enunciato, che è costituita da diversi livelli concettualmente autonomi ma con valori in associazione preferenziale. Attraverso l'azione dell'articolazione informativa dell'enunciato, i valori testuali iscritti, con densità diseguale, in tutti gli ambiti di organizzazione della lingua trovano un'attualizzazione comunicativa. Il testo che così si costruisce è un'entità di natura intrinsecamente semantico-pragmatica, le cui unità di riferimento sono organizzate secondo più dimensioni concettuali.

L'ipotesi della testualità integrata porta con sé importanti corollari epistemologici legati al concetto di linguistica del testo e agli obiettivi della

disciplina, riassunti nella Premessa al volume (pp. 13-19). Ragionare in termini linguistico-testuali significa fare della linguistica pragmatica, ovvero interessarsi ai valori comunicativi iscritti nel sistema della lingua. Il punto di riferimento storico di questo paradigma di ricerca è la *stylistique de la langue* promossa da Charles Bally (cfr. Bally 1944), che viene percepita non come lo studio di un livello di analisi particolare della lingua, ma come una diversa modalità di osservazione dell'intero sistema linguistico. L'obiettivo di ricerca peculiare della linguistica del testo sarà quindi "la costruzione del senso nel testo" (p. 15), che comprende due fuochi di interesse: il modo in cui alcuni contenuti linguistici determinano la componente testuale del senso e il modo in cui il testo acquisisce la coerenza complessiva che consente di definirlo come tale.

Da un punto di vista semantico, questa prospettiva implica un'indagine del significato linguistico di tipo non denotativo, ovvero di quella porzione di significato che, tra le altre cose, consente agli elementi linguistici di costruire un'architettura testuale. Formulazioni equivalenti dal punto di vista denotativo possono essere inserite in strutture sintattiche e interpuntive diverse, che indicano diversi rilievi comunicativi e gerarchie informative, sfruttabili a livello testuale (p. 26). La costruzione del testo non può prescindere, ad ogni modo, dall'interazione tra i dati linguistici e i dati contestuali pertinenti, che possono confermare o smentire l'interpretazione veicolata dalla lingua. Più in generale, la concezione semantica alla base del volume è una concezione minimalista, con risvolti a livello sia linguistico sia testuale: da una parte, i significati linguistici sono visti come sottospecificati e arricchiti dinamicamente in contesto; dall'altra parte, anche gli sfruttamenti testuali delle configurazioni linguistiche non sono determinati una volta per tutte, ma dipendono in modo cruciale dai contenuti ospitati (denotativi o procedurali) e dal contesto. L'uso di un modello minimalista e flessibile è visto come "particolarmente idoneo a rendere conto della specificità dell'italiano scritto" (p. 178), a causa della ricchezza di forme e di strutture che la scrittura italiana ha recepito da varie fonti. Esso permette inoltre di analizzare in modo più puntuale oggetti sfaccettati come la lingua e il testo, che sono dotati di un "ordine non immediatamente percepibile ma fortemente effettivo" (p. 178; cfr. anche la citazione in esergo da Bally 1944).

Il volume è articolato in tre parti: la prima parte (*La lingua, il testo, l'interfaccia lingua-testo*, 21-69) ha la funzione di introdurre le grandezze fondamentali in gioco nel modello, con particolare attenzione al ruolo mediatore svolto dall'articolazione informativa dell'enunciato; la seconda parte (*L'organizzazione informativa dell'Enunciato e l'organizzazione semantico-pragmatica del testo*, 71-175), cuore teorico del volume, elabora nel dettaglio la descrizione della struttura informativa plurilivellare dell'enunciato e delle principali dimensioni di organizzazione del testo; la terza parte (*Aspetti*

linguistici delle organizzazioni informativo-testuali, 177-387), infine, presenta numerose forme e strutture linguistiche dal punto di vista della pertinenza testuale. L'apparato bibliografico del volume è organizzato come segue: oltre alla consueta bibliografia finale (pp. 389-407), sono presenti, al termine della prima parte e dei capitoli di seconda e terza parte, due riquadri bibliografici dedicati, rispettivamente, agli studi pertinenti per le analisi proposte nella parte o capitolo in esame e agli studi degli autori sull'argomento affrontato. Nei seguenti capoversi della recensione, le tre parti del volume verranno descritte nel dettaglio, prestando particolare attenzione agli aspetti dotati di una funzione centrale nel modello elaborato.

Nella prima parte, interamente a cura di Angela Ferrari, viene in primo luogo precisata la prospettiva di indagine del volume dal punto di vista semantico (pp. 22-31) e vengono introdotte le nozioni principali del modello testuale (pp. 31-43): la porzione testuale oggetto di interesse è il Capoverso, costituito da Unità Comunicative che, quando sono realizzate esplicitamente, assumono la forma di Enunciati. Lo spazio in cui le Unità Comunicative sono situate è tridimensionale ed è formato da un piano principale e da più piani secondari; è in questi ultimi che si collocano le Unità Comunicative con funzione di Inciso, solitamente delimitate nella superficie del testo da parentesi o trattini. Nei paragrafi successivi (pp. 43-66), viene introdotto il livello dell'organizzazione informativa dell'Enunciato, di cui viene data una prima caratterizzazione sommaria, precisata nella seconda parte del volume; vengono inoltre precisate le linee dell'interazione tra questo livello e le dimensioni organizzative del testo, nonché le motivazioni che spingono verso un'interpretazione del livello in questione come interfaccia tra lingua e testo.

La seconda parte è a cura di Angela Ferrari, con l'eccezione di un paragrafo e un sottoparagrafo redatti da Anna-Maria De Cesare (*L'organizzazione "cognitiva" dei referenti*, pp. 72-79; *Il Nucleo*, pp. 91-99, all'interno del paragrafo dedicato al livello gerarchico-informativo) e di un sotto-paragrafo a cura di Luca Cignetti (*L'Appendice come Unità Informativa e l'Inciso come Unità Comunicativa*, pp. 111-115, sempre all'interno del paragrafo sopra citato). La parte è divisa in due capitoli autonomi, che si occupano rispettivamente dell'organizzazione informativa dell'Enunciato (pp. 71-115) e dell'organizzazione semantico-pragmatica del testo (pp. 119-172).

L'organizzazione informativa dell'Enunciato viene descritta come una realtà costituita da diversi livelli, i più significativi dei quali sono l'organizzazione "cognitiva" dei referenti (distinzione tra referenti Attivi, Semi-Attivi e Non Attivi, a seconda del grado di salienza nella Memoria Testuale; pp. 72-79), il livello Topic-Comment (definito, sulla base di Lambrecht 1994, a partire da una relazione di *aboutness*; pp. 79-87) e il livello gerarchico-informativo (pp. 88-115). Quest'ultimo livello svolge un ruolo cruciale nel modello elaborato: la segmentazione gerarchica delle parti dell'Enunciato permette di individuare tre

tipi di Unità Minimali, dotate di una specifica pertinenza testuale, che prendono il nome di Unità Informative.

L'unica Unità Informativa necessaria e sufficiente alla costruzione di un Enunciato è il Nucleo (pp. 91-99), che "contiene l'informazione di primo piano dell'Enunciato, quella che ne definisce la funzione illocutiva e testuale" (p. 92); la componente più rilevante del Nucleo, dal punto di vista comunicativo, è il Fuoco Informativo, definito dal criterio del dinamismo comunicativo (di origine pragmatista: cfr. Firbas 1974) e non dalla relazione di *aboutness* che caratterizza, invece, le funzioni di Topic e Comment. La centralità testuale del Nucleo si esplica nella sua capacità, non condivisa dalle altre Unità Informative, di definire le relazioni logiche e topicali che strutturano il testo: sono gli elementi contenuti nei Nuclei degli Enunciati a gestire le relazioni semantico-pragmatiche del piano principale del testo. Il Quadro (pp. 99-105) è un'Unità Informativa facoltativa che si posiziona sempre in apertura di Enunciato. Le funzioni testuali del Quadro agiscono con un ampio raggio: esso può essere utilizzato per assicurare la coerenza dell'Enunciato ospitante nei confronti del cotesto sinistro o per definire domini di riferimento la cui validità può oltrepassare i confini dell'Enunciato. L'Appendice (pp. 105-110) è un'Unità Informativa facoltativa di tipo informativamente subordinato: essa è sempre legata a un'altra Unità Informativa (con qualità di Nucleo, Quadro o Appendice) e la sua azione informativa ha una portata limitata a quell'elemento. Il raggio d'azione strettamente locale dell'Appendice è la caratteristica principale di questa Unità Informativa, che consente sfruttamenti testuali in senso correttivo e riformulativo, *in minore* rispetto al corpo comunicativo dell'Enunciato. Le tre Unità Informative descritte sono gli elementi che adempiono la funzione di interfaccia tra lingua e testo postulata dal modello: esse sono fondate (almeno in parte) su criteri linguistici (e.g. presenza di segni interpuntivi o di costruzioni sintattiche del tipo "clausola") e contribuiscono alla costruzione dell'architettura testuale globale, fornendo istruzioni interpretative decisive per la coerenza del testo.

L'organizzazione semantico-pragmatica del testo è descritta a partire da due dimensioni essenziali: la dimensione logica e la dimensione topicale. La dimensione logica (pp. 120-152) consiste nell'insieme di relazioni logico-semantiche che articolano il contenuto del testo, a tutti i livelli delle unità semantiche che lo costituiscono (dalla semplice entità con denotazione eventiva all'intero Capoverso). Vengono distinte relazioni logiche *de re*, coincidenti con connessioni di tipo fattuale tra gli eventi descritti, e relazioni logiche *de dicto*, che riguardano la composizione del testo da parte del locutore, più utili ai fini della costruzione di un modello testuale. L'argomentazione di questa sezione del capitolo dedica particolare spazio alle differenze legate alla natura delle unità connesse: le relazioni tra Unità Comunicative proiettano gerarchie diverse rispetto alle relazioni tra Unità

Informative, e la scelta di calare il termine di una relazione logica in Unità di Quadro o di Appendice comporta importanti conseguenze testuali.

La dimensione topicale (pp. 152-172) coglie la prospettiva organizzativa del testo dettata dalla scelta e dall'ordinamento dei Topic di Enunciato. Come già anticipato, il Topic di Enunciato tende a coincidere con il Topic dell'Unità Informativa nucleare, a meno che esso non esaurisca un'Unità di Quadro. Il concetto di progressione del Topic utilizzato nel volume si ricollega alla nozione praghese di progressione tematica (cfr. Daneš 1974), definita dall'incrocio di criteri quali lo statuto informativo, la coreferenza e la distanza lineare tra i Topic collegati. La prospettiva topicale permette, parallelamente a quanto accade in prospettiva logica, di gerarchizzare il testo, attraverso l'identificazione di Topic di livello superiore all'Enunciato, sino al Topic di fondo che governa l'intero testo. L'articolazione gerarchico-informativa dell'Enunciato ha funzioni meno decisive in ambito topicale rispetto a quanto accade in ambito logico: le Unità di Quadro e di Appendice, solitamente, non danno un contributo essenziale alla progressione topicale, ma, in alcuni casi, "possono tuttavia collaborare a rafforzarla e in parte a delinearla" (p. 168) (si veda, a titolo di esempio, il caso di Topic in Quadro sopra citato).

La terza parte riunisce contributi curati da tutti gli autori del volume, con l'obiettivo di osservare i valori informativo-testuali di numerose configurazioni linguistiche. La suddivisione in capitoli ricalca l'individuazione di altrettante strutture sintattiche (o categorie lessicali), all'interno delle quali viene osservato il paradigma di configurazioni che si forma attraverso variazioni sintattiche, interpuntive o lessicali. Il fenomeno su cui si concentra maggiormente l'argomentazione in questi capitoli è l'anti-orientamento semantico, ovvero la presenza di configurazioni linguistiche dai valori testuali non congruenti: si pensi, ad esempio, alle cosiddette relative paratattiche, che mostrano a un tempo integrazione sintattica (si agganciano a un elemento della clausola precedente) e frammentazione sintattica (sono separate dalla clausola precedente da un segno di interpunzione forte), proiettando in tal modo sia linearizzazione sia autonomia informativa.

Dopo l'introduzione a cura di Angela Ferrari (pp. 177-181), i due capitoli successivi si concentrano su configurazioni particolari di clausola: il capitolo *Gli Enunciati nominali* (pp. 182-204), curato da Letizia Lala, analizza gli Enunciati la cui funzione predicativa centrale non è svolta da un verbo, suddividendoli in Enunciati nominali autonomi ed emarginazioni (strutture sintatticamente legate al cotesto precedente ma separate da esso da un segno interpuntivo, che assumono valore di Enunciato o di Unità Informativa); Carlo Enrico Roggia studia invece *La clausola, l'ordine dei costituenti* (pp. 206-232), con particolare attenzione alle clausole marcate dal punto di vista gerarchico-informativo (dislocazioni, anteposizioni e posposizioni sintattiche, focalizzazioni, frasi scisse). Seguono tre capitoli dedicati alle proprietà della

frase complessa: Magda Mandelli si occupa della *Frase complessa per coordinazione* (pp. 235-256), in particolare degli effetti di variazioni sintattiche, interpuntive e lessicali in frasi coordinate dalla congiunzione e; Angela Ferrari analizza invece la frase complessa per subordinazione in due capitoli dedicati rispettivamente alla *Relativa appositiva* (pp. 259-282) e alla *Subordinata circostanziale* (pp. 285-313): il trattamento delle relative appositive permette di approfondire il ruolo testuale delle Unità Informative di Appendice, proiettate da questo tipo di clausola, e le funzioni di strutture dal forte valore informativo-testuale, quali la relativa paratattica e l'apposizione grammaticalizzata con relativa; i punti focali di interesse nell'analisi delle subordinate circostanziali sono, invece, la presenza o assenza di integrazione sintattica, che proiettano valori testuali diversi, e l'anti-orientamento semantico che caratterizza le subordinate precedute dal punto fermo. Il penultimo capitolo (pp. 316-338), redatto da Luca Cignetti con un intervento di Angela Ferrari (pp. 334-335), si occupa di *Altre forme di complessità sintattica*, in particolare di fenomeni di estrazione e inserzione sintattica: la scelta di queste operazioni sintattiche comporta solitamente la creazione di un'Unità di Inciso, che si situa su un piano secondario del testo, ma, nel caso dell'utilizzo di virgole al posto di parentesi o trattini, essa può anche proiettare un'Unità Informativa di Appendice. L'ultimo capitolo, infine, sposta la prospettiva sulla testualità integrata in elementi lessicali ed è diviso in due sezioni: Anna-Maria De Cesare si occupa di *Avverbi paradigmaticizzanti* (pp. 340-360), osservandone le funzioni gerarchico-informative e di organizzazione testuale, mentre Claudia Ricci analizza l'uso dei *Connettivi nel testo* (pp. 362-385) e ne mette in luce i differenti valori di impiego nelle Unità di Nucleo, Quadro e Appendice.

Le finalità del modello descritto sono di tipo sia teorico sia applicativo: i due obiettivi del modello, esposti nella Premessa al volume, sono "proporre una rappresentazione astratta della testualità scritta" ed "essere uno strumento per l'analisi della scrittura funzionale (non letteraria) contemporanea" (p. 13). Uno dei punti di forza del modello è la doppia limitazione dell'oggetto di studio che emerge da questi obiettivi: il fuoco di interesse è posizionato sul testo scritto (limitazione diamesica) e, in particolare, sulle sue manifestazioni di tipo funzionale (limitazione tipologica). Questa doppia limitazione consente agli autori, in primo luogo, di trattare con il giusto grado di approfondimento il ruolo dei contrassegni peculiari dello scritto, come la punteggiatura, senza l'obbligo di cercare corrispettivi orali dei valori informativi descritti (le differenze tra scritto e parlato sono invece al centro di lavori precedenti di Angela Ferrari: cfr. Ferrari 2005b). I segnali di autonomia enunciativa e informativa forniti, rispettivamente, dal punto e dalla virgola sono un esempio del ruolo essenziale rivestito dalla punteggiatura nell'elaborazione del modello; molto spesso, è proprio grazie ai segnali (semanticamente co-orientati o anti-orientati) forniti dalla punteggiatura che la scrittura acquisisce valori informativi essenzialmente diversi da quelli dell'oralità, che contribuiscono alle più

generalmente differenze pragmatiche tra scritto e parlato. In secondo luogo, la limitazione tipologica comporta l'analisi di tipi testuali le cui strategie di composizione oscillano tra valori pratici e obiettivi stilistici: numerosi esempi sono tratti da "generi di scrittura che si vogliono brillanti e stilisticamente attraenti, come la stampa quotidiana contemporanea" (p. 336), di cui si sottolineano gli sfruttamenti testuali di strategie linguistiche (e.g. catafora tematica p. 159, nominalizzazione sintagmatica p. 189, enunciati nominali a picco informativo p. 193, inserzioni tra virgole p. 336).

Un'ulteriore limitazione dai risvolti positivi, evidente nel volume ed esplicita nel titolo del progetto che vi ha dato origine, è quella che riguarda la lingua oggetto di analisi: l'italiano contemporaneo. Il fuoco di attenzione sull'italiano contemporaneo è alla base di alcune osservazioni che dimostrano una considerazione attenta delle peculiarità retoriche della tradizione italiana: le strutture informative disegnate dalle configurazioni linguistiche vanno considerate nell'ambito di "una scrittura che, come quella italiana, tende per tradizione retorica a evitare la ripetizione" (p. 172) e che, parallelamente, tiene in grande considerazione "la ricerca di *variatio* – ben radicata in tutte le lingue romanze" (p. 316). Queste osservazioni dimostrano come siano ben presenti agli autori del volume l'importanza delle scelte stilistiche, fondate nella tradizione retorica di una comunità linguistica, e il ruolo culturalmente specifico che esse possono avere. La scelta di marcare la coreferenza tra due entità con la strategia dell'accordo verbale o con un'anafora lessicale "fedele" è un perfetto esempio di fenomeno in cui l'articolazione gerarchico-informativa dell'Enunciato risente delle specificità retoriche: l'anafora lessicale comporta in italiano l'attribuzione al referente di un forte rilievo comunicativo, che va considerato nella segmentazione informativa dell'Enunciato.

L'attenzione alle motivazioni stilistiche si traduce anche in giudizi di appropriatezza più sfaccettati e attenti ai vincoli imposti dalle diverse tipologie testuali (cfr. Sabatini 1999 per la nozione di testi più o meno vincolanti, richiamata in alcuni passi del volume). Un fenomeno come l'emarginazione di Enunciati nominali, ad esempio, proietta gerarchie informative diverse a seconda della scelta del segno di interpunzione: il punto fermo ha un effetto totalizzante sui contenuti testuali antecedenti, testualmente più marcato, mentre i due punti sono rivolti verso il cotesto destro e annunciano un complemento a ciò che precede. Le due possibilità "non sempre sono appropriate per le stesse tipologie testuali e relativi registri stilistici" (p. 198), e "[i]l paragone tra le due varianti varrà [...] per quelle tipologie di testo, e per quegli stili, all'interno dei quali siano ammesse entrambe le possibilità" (p. 199). È così che le scelte linguistiche dalla forte testualità integrata vengono valutate in modo più raffinato, consapevole delle differenze che sussistono tra le diverse manifestazioni della scrittura funzionale.

Il modello presentato nel volume può essere uno strumento utile per chiunque sia interessato alla testualità scritta, specie se l'interesse sia rivolto, in particolare, alle molteplici manifestazioni dell'italiano non letterario. Dal punto di vista della strutturazione del volume, le prime due parti, curate da Angela Ferrari, hanno il pregio di presentare in modo molto chiaro gli elementi di base del modello e gli strumenti di analisi che vengono poi visti all'opera nella terza parte. La scelta di dedicare la prima parte alla presentazione delle grandezze in gioco e la seconda parte all'approfondimento degli aspetti centrali e dei tecnicismi del modello giova alla comprensibilità dell'argomentazione. Inoltre, il volume si segnala per la grande ricchezza di esempi forniti e commentati (se ne contano 917), tratti in buona parte dai lavori precedenti degli autori, che talvolta vengono ripetuti nelle diverse sezioni: la ripetizione degli esempi permette di cogliere i diversi punti problematici insiti nel caso in esame, attraverso l'utilizzo di diverse nozioni esplicative proposte dal modello.

L'apparato descrittivo del testo permette di analizzare con accuratezza tutte le unità che lo compongono, dai Capoversi alle Unità Informative Minimali. Queste ultime, in particolare, forniscono il contributo più originale del modello alla comprensione della testualità scritta: la segmentazione gerarchica dell'Enunciato in Unità di Nucleo, Quadro e Appendice permette di cogliere le conseguenze decisive a livello testuale che derivano dalla collocazione di specifiche configurazioni sintattiche in specifiche posizioni gerarchico-informative.

L'identificazione di Unità di Nucleo, Quadro e Appendice deriva, in ultima analisi, da criteri funzionali e contestuali, ma la forma linguistica del messaggio fornisce degli indizi spesso decisivi: "si rivelano significativi – *ex negativo* – i tratti linguistici che sono sintomatici di una connessione non pienamente riconducibile alle leggi formali sistemiche della *langue*" (p. 89). Assumono grande importanza, in questo quadro, i segni di interpunzione e i connettivi pragmatici: i primi hanno spesso un ruolo informativo, che va comunque valutato in opposizione alle possibili motivazioni alternative di natura sintattica, semantica e pragmatica; ai secondi è invece spesso affidata "l'espressione linguistica delle relazioni logiche" (p. 362), fondamentali ai fini della coerenza globale del testo. L'argomentazione del volume pone di frequente l'accento sulla forte pertinenza informativa e testuale di questi elementi, capaci di proiettare confini e collegamenti da considerare in ottica linguistico-testuale. Un aspetto centrale del modello presentato è la distinzione, ancora una volta risalente a Bally, tra "una *sintassi arbitraria*" – governata da principi formali, come l'accordo tra soggetto e predicato verbale – e una *"sintassi motivata"* da ragioni pragmatiche" (p. 30): le scelte testualmente pertinenti del locutore si situano obbligatoriamente a livello della sintassi motivata, e, d'altro canto, non tutti i risultati della sintassi motivata sono motivati da scelte testuali (e.g. virgole in coordinazioni asindetich).

L'attenzione riservata alle indicazioni interpretative veicolate dalla lingua è probabilmente la qualità principale del volume, che si traduce in analisi raffinate dei differenziali informativo-testuali insiti nelle strutture linguistiche. D'altra parte, Angela Ferrari ha insistito di recente sulla necessità di un ritorno alla *langue* per la linguistica del testo (cfr. Ferrari 2009 in stampa), che consenta di slegare la coerenza da generici principi di interpretabilità e di recuperare (e operazionalizzare) i fondamenti linguistici della coerenza testuale.

Anche la scelta di descrivere l'organizzazione del testo secondo due dimensioni fondamentali, la dimensione logica e la dimensione topicale, risulta particolarmente felice: specie nello scritto, la coerenza complessiva del testo ha una natura ibrida, che non si può ridurre alla sola coesione (o connessità) tra istanze di referenti testuali. Viene peraltro notato che queste dimensioni «non esauriscono la "sostanza" dell'organizzazione semantico-pragmatica del testo» (p. 42): anche altre prospettive (e.g. illocutiva, dispositiva) hanno un ruolo strutturante, ma le due considerate nel dettaglio "restano tuttavia centrali nello scritto" (p. 42). La considerazione congiunta di queste dimensioni potrà rivelarsi utile anche per studi rivolti a fenomeni micro-testuali (e.g. anafora discorsiva), che potranno così essere considerati in relazione allo specifico aspetto organizzativo del testo che aiutano a costruire, o verso il quale il loro contributo è prevalente.

BIBLIOGRAFIA

- Bally, C. (1944). *Linguistique générale et linguistique française*. Berna: Francke.
- Daneš, F. (1974). Functional Sentence Perspective and the Organization of the Text. In F. Daneš (a c. di.), *Papers on Functional Sentence Perspective* (pp. 106-128). Praga-Parigi: Academia-Mouton.
- Ferrari, A. (1995). *Connessioni. Uno studio integrato della subordinazione avverbiale*. Ginevra: Slatkine.
- Ferrari, A. (2003a). *Le ragioni del testo. Aspetti sintattici e interpuntivi dell'italiano contemporaneo*. Firenze: Accademia della Crusca.
- Ferrari, A. (2003b). La struttura informativa dell'enunciato scritto (nell'italiano contemporaneo non letterario). Presentazione del progetto FNRS P001—68675/1 [www.lisulb.unibas.ch/structure/files/Pro_FNRS_it.pdf] (27/01/2013).
- Ferrari, A. (a c. di) (2004). *La lingua nel testo, il testo nella lingua*. Torino: Istituto dell'Atlante Linguistico Italiano.
- Ferrari, A. (a c. di) (2005a). *Rilievi. Le gerarchie semantico-pragmatiche di alcuni tipi di testo*. Firenze: Cesati.
- Ferrari, A. (2005b). Tipi di testo e tipi di gerarchie testuali, con particolare attenzione alla distinzione tra scritto e parlato. In A. Ferrari (a c. di), *Rilievi. Le gerarchie semantico-pragmatiche di alcuni tipi di testo* (pp. 15-51). Firenze: Cesati.
- Ferrari, A. (a c. di) (2006). *Parole frasi testi, tra scritto e parlato* (=Cenobio LV/3).

- Ferrari, A. (2009). Maria-Elisabeth Conte e la linguistica del testo. Note introduttive. In F. Venier (a c. di), *Tra pragmatica e linguistica testuale: ricordando Maria-Elisabeth Conte* (pp. 239-250). Alessandria: Edizioni dell'Orso.
- Ferrari, A. (in stampa). La linguistica del testo. In G. Iannaccaro (a c. di), *La linguistica italiana all'alba del terzo millennio (1997-2010)*. Roma: Bulzoni.
- Ferrari, A. & Zampese L. (2000). *Dalla frase al testo. Una grammatica per l'italiano*. Bologna: Zanichelli.
- Firbas, J. (1974). Some aspects of the Czechoslovak approach to problems of Functional Sentence Perspective. In F. Daneš (a c. di), *Papers on Functional Sentence Perspective* (pp. 11-37). Praga-Parigi: Academia-Mouton.
- Lambrecht, K. (1994). *Information structure and sentence form. Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Sabatini, F. (1999). 'Rigidità-esplicitezza' vs 'elasticità-implicitezza': possibili parametri massimi per una tipologia dei testi. In G. Skytte & F. Sabatini (a c. di), *Linguistica testuale comparativa. In memoriam Maria-Elisabeth Conte* (pp. 141-172). Copenhagen: Museum Tusculanum Press.

Filippo Pecorari

Università di Pavia

filpec@alice.it