

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