



HAL
open science

Corpus linguistics and data-driven learning: a critical overview

Alex Boulton, Henry Tyne

► **To cite this version:**

Alex Boulton, Henry Tyne. Corpus linguistics and data-driven learning: a critical overview. Bulletin suisse de Linguistique appliquée, 2013, Bulletin suisse de linguistique appliquée, 97, pp.97-118. hal-01208263

HAL Id: hal-01208263

<https://hal.science/hal-01208263v1>

Submitted on 5 Oct 2015

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

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

Corpus 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/Tim.Johns/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).