



**HAL**  
open science

## Data-driven learning: reasonable fears and rational reassurance

Alex Boulton

► **To cite this version:**

Alex Boulton. Data-driven learning: reasonable fears and rational reassurance. Indian Journal of Applied Linguistics., 2009, 35 (1), pp.XX. hal-00326990v1

**HAL Id: hal-00326990**

**<https://hal.science/hal-00326990v1>**

Submitted on 6 Oct 2008 (v1), last revised 19 Jun 2009 (v2)

**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.

## Data-driven learning: Reasonable fears and rational reassurance.

ALEX BOULTON  
*CRAPEL-ATILF/CNRS, Nancy Université, France*

### ABSTRACT

*Computer corpora have many potential applications in teaching and learning languages, the most direct of which – when the learners explore a corpus themselves – has become known as data-driven learning (DDL). Despite considerable enthusiasm in the research community and interest in higher education, the approach has not made major inroads to mainstream language teaching. This paper explores some of the reasons for this, with the intention of demystifying DDL for use with ordinary learners and teachers in ordinary classrooms.*

**Keywords:** Data-driven learning, corpora, obstacles, roles, teacher, resources, materials

### 1. BACKGROUND

There has been continual interest in foreign language pedagogy since time immemorial, but the last 50 years or so has seen particular creativity and diversity as practitioners seek more efficient ways to go about it. Most remarkable perhaps were the “designer methods” (H. Brown et al. 2007: 9) of the 1970s, such as Suggestopedia, the Silent Way or Total Physical Response. Their limited adoption world-wide is perhaps partly due to dogmatic adherence to ideology which remains impervious to evidence or experimentation, and insufficiently able to adapt to local cultures. Indeed, their existence has left a certain wariness towards any claim of “revolution” or “panacea” in the field. The most successful recent methodology globally has undoubtedly been the very broad church of the communicative approach (CA). While this implied a fundamental rethink of certain underpinnings, it has remained highly eclectic, retaining or adapting many existing tried and tested practices.

This makes CA hard to pin down (Hadley 2002), and many would be hard put to see the “communicative” nature of many self-proclaimed teachers, materials and practices.

One of the more traditional aspects apparent in many instances of CA is the emphasis on the teacher – not for nothing do Richards and Rodgers (2001: chap. 14) dub it “communicative language *teaching*” (emphasis added). The following caricature could as easily apply to many “communicative” classrooms today as to grammar-translation a century ago: “Your teacher is the guide and mentor, who will show you what to learn and how to learn it. Listen to your teacher and do as you are told” (D. Willis 2003: 167).

In general, however, CA has seen increased interest in the learner and the learning process. Concomitantly, the advent of information and communication technology (ICT) has inspired attempts to reduce the role of the teacher: we talk of computer-assisted language *learning* (CALL) rather than computer-assisted language *teaching*. It is a commonplace however that this is a naïve view, as much CALL software merely replaces the teacher with an even more rigid guide; it would not be out of place to replace the word *teacher* with the word *computer* in the quotation from Willis above. The perpetual question with new technologies is whether we are genuinely doing new things, or merely rehashing old things in new ways (cf. Noss & Pachler 1999); or as Higgins and Johns (1984: 10) put it: “the usual reaction from language teachers is that [CALL materials] contain nothing which cannot be done already with pencil and paper, and that the gains... do not justify the expense and trouble.” Sadly, the observation remains relevant 25 years later.

One particular use of ICT which claims to focus on learning rather than teaching is data-driven learning (DDL), to use the expression coined by Tim Johns. He summarises it as “the attempt to cut out the middleman as far as possible and to give the learner direct access to the data” (1991b: 30). The “middleman” refers of course to the teacher, but the computer is not seen as “a surrogate teacher or tutor, but as a rather special type of informant” (Johns 1991a: 1). DDL typically involves exposing learners to large quantities of authentic data – the electronic corpus – so that they can play an active role in exploring the language and detecting patterns in it. They are at the centre of the process, taking increased responsibility for their own learning rather than being taught rules in a more passive mode. Although many of the basic concepts are widespread in CA (learner-centred, discovery learning, autonomisation,

authentic language, etc.), DDL nonetheless strikes many as quite revolutionary, and therefore to be treated with caution.

The aim of the present article is to demystify or demythologise DDL, to examine a number of objections or fears that potentially interested parties may have. Some are cited by hostile sceptics (e.g. Dellar 2003), others by enthusiastic practitioners (e.g. Farr 2008; Sun 2003). The aim is not to ridicule these difficulties, but to discuss them rationally and, ideally, help sceptics to suspend their doubts long enough to experiment with the techniques for themselves. It is argued that DDL is well within the reach of regular teachers and learners in ordinary language teaching contexts, and that a small investment in terms of time and effort can lead to immediate and, more importantly, long-term language learning benefits.

The paper is not intended as a “how-to” introduction to DDL, as the ground has been covered excellently elsewhere for a variety of learning contexts. Several introductory articles are available on line, including Lamy and Klarskov Mortensen (2007), Gabrielatos (2005), Rüschoff (2004), Tan (2003), Hadley (2002), and Thomas (2002). There are also some excellent collections of research papers reporting on classroom uses of DDL, most notably perhaps Kübler (in press), Hidalgo et al. (2007), Sinclair (2004), Aston (2001), and Burnard and McEnery (2000). There is as yet no general manual devoted to DDL (the absence in itself highlights the recent and innovative nature of DDL, and the lack of instant recipes the responsiveness to local cultures); a number do however include sections on DDL alongside other applications of corpora in language teaching and learning not covered here (e.g. the use of learner corpora, syllabus and materials design, etc.), recently including O’Keeffe et al. (2007), Adolphs (2006), Gavioli (2005) and Hunston (2002). The bibliography of the present article also contains a number of key references in the field.

## 2. LEARNING

Clearly a prime concern is pedagogical: does it work, when, and how? Traditionally, teaching represents an attempt to simplify things as far as possible for the learner by encapsulating complex data in simple rules to be taught, reproduced and manipulated to ensure learning. Few would argue that such rules should be abandoned altogether: they can help to draw learners’ attention to features they might not otherwise notice in a clear and simple way. This is particularly the case for the “big themes” of grammar, where DDL is only occasionally applied

(Hunston 2002: 184). Where DDL seems to be most useful is for extending or deepening knowledge of existing language items, distinguishing close synonyms, detecting patterns of usage, collocation, colligation, morphology, and so on. It can sensitise learners to issues of frequency and typicality, register and text type, discourse and style, as well as the fuzzy nature of language itself.

DDL has a number of advantages over a rule-based approach. Firstly, many rules derived from intuition simply do not describe actual usage, as any number of corpus studies have shown over the last 20 years or more. Secondly, rules and exceptions do not provide an accurate picture of language in general, which adheres to patterns, tendencies and generalisations of prototypical usage rather than rigid right or wrong. Thirdly, rules rarely give an idea of frequencies: one may teach beginners the use of perfect and continuous aspects, but forget to point out that 90% of all verb phrases are not marked for aspect (Biber et al. 1999: 461). Finally, rules tend to be rather abstract, as they attempt to account for general language use: corpus investigations have shown substantial differences in the use of grammar in different registers or text types, or in speech and writing. Working on a specific corpus can help learners to identify the parts of the language which are relevant to them, to work on the forms frequently used in the registers and text types they need.

Rule-based learning is extremely demanding – one reason perhaps why it is so beloved in traditional educational environments as a serious intellectual activity. Teachers find rules comforting and reassuring, easier to present and to test, but a false comfort nonetheless. A large literature, most recently in evolutionary psychology (e.g. Cosmides & Tooby 1992), demonstrates how and why human beings have evolved to be good at noticing regularities in nature, interpreting them and extrapolating to other cases, the very processes which DDL brings to the fore (Scott & Tribble 2006: 6). Learners can be surprisingly capable: they may not always be accurate in their conclusions, but neither are rules generally assimilated completely and accurately at first go – *all* learning is a process of gradual approximation to the target (Aston 2001: 13). Furthermore, it has frequently been observed that learners' observations are more accurate and complete than traditional grammar rules; at the very least, their inferences are likely to be relevant and comprehensible to them. After all, as Gaskell and Cobb (2004: 304) remind us, foreign languages are mainly learned “through enormous amounts of brute practice in mapping meanings and situations to words and structures. These mappings... lead over a very

large number of episodes... to the slow extraction of patterns that are rarely articulated.” Such a picture of massive exposure is virtually impossible for most students, whose main contact with the target language is in the classroom in their L1 environment. And this is of course precisely the advantage of DDL, as it provides opportunities for substantial amounts of targeted practice on selected items which otherwise would only be met on occasion or through invented and impoverished contexts.

The *process* of language learning is thus paramount – “every learner a Sherlock Holmes”, as the ever quotable Johns puts it (1997a: 1). Inductive learning may be more motivating and relevant, and the discovery process itself may lead to deeper cognitive processing, and hence better understanding as well as better retention (Laufer & Hulstijn 2001). O’Sullivan (2007: 277) provides an impressive list of cognitive skills liable to be refined through corpus use: “predicting, observing, noticing, thinking, reasoning, analysing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorising, hypothesising, and verifying.” Detecting patterns and regularities also allow learners to realise that much of language use is highly fuzzy, with typical or frequent uses rather than rules and exceptions. Indeed, this is one reason why it is so difficult to formulate rules which are at the same time accurate, complete and easily comprehensible. Often they remain abstract and abstruse, difficult for the learners to understand, let alone remember and apply when needed.

More delicate perhaps is the question of whether DDL actually works, or how effective it is. Chambers (2007) examines 12 DDL studies, mostly small-scale and qualitative in nature, while Boulton (2008a) surveys 50 with some claim to empirical analysis, although he notes that the majority are mainly concerned with annex questions such as what learners do or whether they like doing it, or how effective corpora can be as a reference tool in writing, translating or error-correction rather than as a learning tool. While it is not possible to discuss the results of all these studies in detail, the overall picture is certainly complex: partly because of the vast number of variables as DDL is experimented around the world with different types of learners, in different cultures, in different learning environments; partly because of the flexibility of the approach, meaning that each study uses different tools and techniques; partly because each analysis has its own procedure and often a very precise focus. The majority of these points are not exclusive to DDL but apply equally to any other methodology;

applied to CA, the question “does it work” seems almost nonsensical – the point is to adapt it to suit the local environment.

The overall pattern is certainly encouraging, especially regarding the qualitative studies. The few which do attempt some kind of quantitative evaluation of learning outcomes per se produce more mitigated results – positive, and yet often not as substantial or as statistically significant as might be hoped (Boulton 2008d). But again, this is typical of empirical studies in most fields of language learning. In the particular case of DDL, it may be that the real benefits lie less in short-term gains on targeted items, and more in incidental learning from exposure to the large numbers of other items, greater sensitivity to language and the processes of language learning, better noticing skills for items relevant to their own needs, and so on. All of these lead to increased autonomy outside the classroom (Johns 1991b: 31), and the mastery of tools and techniques which can be used long after instruction has finished. Most studies report that learners who have had experience of working with corpora intend to continue doing so in the future (e.g. Allan 2006; Lee & Swales 2006; Chambers & O’Sullivan 2004; Gaskell & Cobb 2004; Yoon & Hirvela 2004). Evidence for such benefits is likely to be difficult to obtain; fortunately, teachers tend to accept or reject particular tools, materials and techniques not on the basis of research evidence, but on their own pragmatic experience – whether it works for them in their particular situation.

### 3. LEARNERS

The empirical evidence, we have seen, is encouraging yet not world-shattering. Of course, all learners are different, and it is likely that quantitative analyses conceal considerable variation, with some learners benefiting enormously, others not – just as with any methodology. Learners may have difficulty adapting as they are “asked to abandon deeply rooted norms of classroom behaviour” (Bernardini 2001a: 23). Many learners may prefer to be told what to do, accepting that it is the teacher’s role as expert to show them, and resent having to take any responsibility for their own learning. But coming to terms with the new roles may be more a problem for teachers than for learners, as we shall see in Section 4.

Background culture no doubt has a part to play, as cultural conditions vary tremendously around the world, from “the staunchly individualistic (essentially Anglo-Saxon) to the patriarchal collectivistic (essentially Oriental)” (D. Brown 2007: 61). Clearly it is essential to

remain sensitive to background cultures: “any supposedly general principles have to be interpreted with reference to local settings, or otherwise they are doomed to remain meaningless” (Seidlhofer 2002: 220). Further, culture itself is a generalisation: different cultures may exist at more local levels in different regions, different institutions, even at the level of the individual classroom, each with its own dynamic. Flowerdew (2001: 376), for example, found that science and engineering students took to DDL quite easily, while business students from the same institution had more difficulty mastering the approach and the software.

Statistical results from empirical studies conceal individual differences from learner to learner in any part of the world. For example, regarding her own DDL experiments, Chambers (2005: 119) speculates that “differences in motivation or learning styles may explain the considerable variation in the success of the activity.” Cultures which attach particular value to certain characteristics may encourage them; but it is possible to create a local classroom culture which is different from the background: allowing learners the opportunity to be different, to be themselves, will find a certain number of adherents anywhere. The present author’s DDL research is conducted in France, which according to D. Brown (2007) is more towards the “patriarchal collectivist” end of the spectrum, and yet learners’ reactions on the whole tend to be extremely positive (e.g. Boulton 2008d). It is certainly true that “learners... may live within culturally diverse pedagogic traditions not compatible with [DDL]” (G. Cook 1998: 60); but it would seem ethically dubious to deny learners the opportunity even to try a potentially useful set of tools and skills on the assumption that they will all adhere to the precepts of that culture.

Very little is known about the types of learners who take most readily to DDL or extract most benefit from it. One of the few to venture an idea is Flowerdew (2008: 117), who notes that it:

may not appeal to students with different cognitive styles. Field-dependent students who thrive in cooperative, interactive settings and who would seem to enjoy discussion centering on extrapolation of rules from examples may benefit from this type of pedagogy. However, field-independent learners who are known to prefer instruction emphasizing rules may not take to the inductive approach inherent in corpus-based pedagogy.

It is important to bear in mind that learning styles are not static, but are subject to change along with the various learning experiences. Cresswell (2007: 279) takes this to suggest that learners who are reticent may be won over by a gentle introduction via teacher-mediated



paper-based materials to check rules (what he calls “deductive DDL”) rather than full-blown autonomous, hands-on “inductive DDL”.

Considerable research is needed before any definite conclusions can be reached. This becomes particularly apparent as increasing quantities of empirical research are starting to question the traditional assumption that DDL is only useful for advanced, sophisticated, adult learners. The vast majority of published research unsurprisingly concentrates on this type of public as they are to be found in the researchers’ own university environments, going right back to Johns (1986: 161), who was working with:

a particular type of student (adult: well motivated: a sophisticated learner with experience of research methods in his subject area) with particular needs (fairly closely specifiable in terms of target texts) in a particular learning / teaching situation (in which a great deal of emphasis is placed on developing students’ learning strategies and on their responsibility for their own learning).

But this does not preclude others: his following sentence points out that “it remains to be seen how far the ‘research methodology’ outlined above would be suitable for other learners.” It also depends greatly on the activities assigned: DDL is not an all-or-nothing affair, and teachers should not be put off if they feel their learners are not up to the hands-on serendipitous learning reported in many papers (Mukherjee (2006: 14). Teachers who are wary of losing too much control may find inspiration in some of the less radical implementations mentioned in Section 5.

DDL researchers are increasingly working in high school environments (e.g. Braun 2007; Sun & Wang 2003; Ciezielska-Ciupek 2001). In general, these researchers find their students to be enthusiastic, with DDL providing substantial benefits. This echoes findings from studies using similar techniques with even younger learners in their L1, most notably the work of Sealey and Thompson (e.g. 2004). In his survey of 50 empirical studies of DDL, Boulton (2008a) found eight further studies working with lower levels, including two ostensibly with beginners. Although the aims and procedures were in most cases fairly limited, all of these studies report success.

While the overwhelming majority of all studies do find most learners enthusiastic about DDL, there are occasionally more negative findings (notably Estling Vannestål & Lindquist 2007; Whistle 1999). Even positive reports cite some learner dissatisfaction, especially that the work can be mechanical, laborious, and even tedious (Chambers 2007). Allan (2006) found that her students tired after more than 30 minutes a week of DDL outside class, and others have suggested in-

class DDL activities should not be prolonged more than this (e.g. Whistle 1999). Clearly “a variety of tasks is important, and an over-reliance on concordancing should be avoided” (Allan 2006: 9). Numerous examples are given in the materials listed in Section 5; the possibilities are “limited only by the imagination of the user” (Breyer 2006: 162). Many software packages allow user-friendly interfaces for various types of tasks: comparing varieties, registers or text-types; looking for collocates and chunks; comparing frequencies; and so on.

Motivation can be increased by allowing learners greater involvement in creating the corpus, deciding what goes into it, or using their own productions (cf. Seidlhofer 2000). This helps them to see the relevance of what they are doing, which can also be achieved by working on language areas they know they have problems with. Johns’ approach was largely “reactive, responding to the difficult questions that intelligent students put... the concordancer allowing the teacher to say ‘I’m not sure: let’s find out together’.” This can be done in the form of prepared materials, or simply having a computer ready in the classroom (cf. Tribble 1997). Learners may also be encouraged to pursue their own enquiries individually – so-called serendipitous learning (e.g. Bernardini 2000), even in the form of corpus-based projects out of class, whether with a linguistic or other focus (e.g. Boulton in press; Römer 2006; Kettemann & Marco 2004).

Occasionally it is claimed that learners may have difficulty with the authentic language found in corpora, especially in interpreting the truncated concordance lines of key words in context (KWICs). This may be a problem in some cases, especially when dealing with “messy” data self-compiled from the Internet (Tribble 1997). But it is perhaps overstated – more a teachers’ worry than one expressed by the students themselves; Boulton (2009) for example reports lower-intermediate learners scoring higher with KWICs than with full sentence contexts. The important point is that the learner does not need to understand everything in each line, as the multiplicity of lines provides more contexts from richer, more varied sources (cf. Stevens 1991). KWICs require a new kind of “vertical” reading, which can be facilitated by encouraging learners to focus on a few words either side of the node; Sinclair (2003) provides extensive tips and techniques.

The language may be made more accessible if it is possible to “grade” the texts within a corpus (e.g. Chujo et al. 2007), or the concordance output (Wible et al. 2002). It has also been suggested that simplified readers may provide one solution (e.g. Cobb 2006), though this might be argued to undermine one advantage of DDL, namely its

use of authentic text. The heated discussions over the use of invented sentences (e.g. Carter 1998 vs. G. Cook 1998; Widdowson 2000 vs. Stubbs 2001; G. Cook 2001 vs. V. Cook 2002) have contrasted the rich nature of authentic text with the focusing nature of invented sentences. Learners, like teachers, might find the messy nature of real language in use to be destabilising at first, preferring the teacher to have all the answers. But it would seem disingenuous to coddle learners with simplified language, disempowering them and leaving them unprepared for the realities of the authentic language we are presumably preparing them for. Widdowson has argued that a major problem with authentic text is that it is taken out of context (especially in the case of concordances), and so by definition loses its authenticity of purpose. Its relevance “must depend on whether learners can *make it real*” (2000: 7). This is not a new issue: Johns (1988: 10) argued that:

text... and the learner’s engagement with text should play a central role in the learning process. In that engagement, a key concept is that of authenticity, viewed from three points of view – authenticity of script, authenticity of purpose, and authenticity of activity.

More recently, Braun (2005: 53) agrees that “real-language texts... are only useful insofar as the learner is able to authenticate them, i.e. to create a relationship to the texts,” but this can be achieved in several ways. She herself suggests using multi-modal corpora; another possibility is to use small corpora (e.g. Aston 1997), especially in ESP contexts (Gavioli 2005), or corpora of learners’ textbooks (e.g. Mparutsa et al. 1991), or to allow learners to choose or create the corpus as argued earlier. Moreover, Mishan (2004) makes the important point that corpus consultation itself an authentic activity: learners are authentically engaged in a research activity that the corpus was compiled for and the software designed for.

Learners interacting with the corpora directly on computer sometimes claim it is frustrating (e.g. Farr 2008), as they have difficulty thinking of appropriate questions, formulating them appropriately, choosing relevant corpora, interpreting the results, and refining their queries with subsequent searches (e.g. Kennedy & Miceli 2001). Training is of the essence here for hands-on DDL to be effective and efficient; as Frankenberg-Garcia (2005a) points out, learners can always benefit from further training even with such familiar tools as dictionaries. Some research recommends several hours of initial training (e.g. Aston 1996), but this tends to be for use of software designed for research linguists, especially earlier generations of software which were considerably slower and less user-friendly. Corpora with integrated interfaces for on-line access today may require

as little as five minutes' introduction (Boulton 2008d). In any case, the introduction of corpora is probably best conducted piece-meal rather than plunging the learners straight in at the deep end. "The difficulties should not be overestimated; learners should quickly acquire the skills needed" (Bernardini 2001b: 243).

Teachers may sympathize with Whistle's (1999: 77) students, some of whom "could not see why the concordances could not be prepared in advance and handed out in class." Indeed, the whole point of rules is to avoid wasting time by having learners work them out for themselves, and time spent on the computer may be considered as time not spent on the real issue of language learning. There are a number of points to be made here. Firstly, induction is more likely to lead to long-term retention than simply being told – the process of discovery itself is important (cf. Laufer & Hulstijn 2001). Secondly, "corpus skills constitute a learning task in themselves... Once acquired, they facilitate learning greatly and need not be constantly refreshed" (Mauranen 2004a: 99). Each time learners think of questions or try to interpret the data, they become better at it; the slow process in early stages contributes to more efficient learning later on. Furthermore, not only are learners acquiring language skills, but are becoming better, more autonomous learners. They are acquiring language as well as ICT skills and life skills at the same time (cf. Inkster 1997), skills which can cross over to other domains of study. A number of papers report on the interdisciplinary nature of corpus linguistics, encouraging learners to apply them to literature, cultural studies, and personal interests such as song lyrics and film transcripts (e.g. Boulton in press). Römer's (2006: 105) attitude is that we are "equip[ping] our students with a tool box, containing skills that are transferable from problem to problem across sub-disciplines." Similarly, not only does DDL enable learners to export skills to other fields, it also enables them to import them, making use of ICT skills and others they already have (e.g. using Internet search engines).

In other words, time spent on DDL is not time wasted, even if the process seems disproportionate to the immediate gains on the targeted items. For many learners, of course, time is not a luxury, as they have a syllabus to cover in an already tight schedule. Teachers might find it difficult to motivate them if they do not look beyond the short-term benefits, especially as regards their grades; as Milton (1996: 239-240) remarks, learners may lose interest in anything which is not explicitly exam-oriented. Lee and Swales (2006) provide an example course

outline, but in most cases it is likely to be preferable to integrate DDL into other course work.

#### 4. TEACHERS

From the teacher's point of view, if DDL has yet to make real inroads to mainstream teaching practices and environments, the problem could lie at any one of three stages: a) teachers might not know about DDL; b) they might know but be unwilling or unable to put it into practice; c) they might try it and then reject it. The major problem rests perhaps with the very first stage: DDL has simply not yet penetrated the consciousness of the teaching profession world-wide. For example, a recent survey among nearly 250 high school teachers in Germany found that approximately 80% were entirely unaware of corpus applications in language learning (Mukherjee 2004). In Britain, questionnaires sent to higher education institutions showed that corpus use remained exceptional (Thompson 2006). The research interest is certainly there, with numerous articles, websites, conferences, and so on, but more is clearly needed to break out of the research environment.

Awareness would increase if major publishers were to produce DDL materials. As yet, very little exists exclusively devoted to DDL, and while corpora are used to inform many textbooks and other materials, they are deliberately hidden with no DDL-style activities in sight. McCarthy (2004: 15), a major figure behind pedagogical uses of language corpora as well as many language teaching materials, remarks of one recent course: "teachers and learners should expect that, in most ways, corpus informed materials will look like traditionally prepared materials. The presentation of new language and activity types will be familiar." Informal discussions suggest that publishers are reticent to produce DDL materials, believing there to be no market for them; but until they exist, there will be no demand – a Catch 22 situation.

Conrad (2000: 556) has argued that "the strongest force for change could be a new generation of ESL teachers" introduced to corpora in their pre-service training. A number of attempts have been made to promote this, usually meeting with considerable enthusiasm on the part of the teachers (e.g. Farr 2008; Tsui 2005; O'Keeffe & Farr 2003; Seidlhofer 2000; Renouf 1997). However, in the case of pre-service training in particular, such courses are unlikely to attract much interest until such time as they become fully integrated to the training programme and examination requirements (cf. Davis & Russell-Pinson 2004; O'Keeffe & Farr 2003). Too short an introduction may leave

teachers sceptical (e.g. in Boulton 2008d), and this scepticism may endure even after a training course. Mukherjee (2004), for example, finds that teachers on his in-service training course quickly see the interest for themselves (as a source of authentic examples, creating tests, checking usage, etc.), but are loath to give their learners direct access to corpora. A further problem is time: as long as DDL is seen as an optional extra, it may be resented as an unnecessary burden on the teacher (cf. Mauranen 2004b: 197).

Johns (1991a: 12) reports similar scepticism, with teachers saying it “may be all very well for students as intelligent, sophisticated, and well-motivated as ours..., it would not work with students as unintelligent, unsophisticated and poorly-motivated as theirs.” These teachers may be right, as they are basing their reaction on their own personal teaching experience. Nonetheless, as Johns goes on to say, it is difficult to know what learners are capable of until they try; denying them the opportunity of acquiring skills would seem a short-term and defeatist position to adopt. The negativity expressed by some (though of course that Johns’ quotation is something of a caricature) suggests another problem, namely the teachers themselves. DDL is quite incompatible with the “minimum risk” scenario which can be found in many teaching cultures (Johns 1988: 11), “in which the teacher ploughs through a textbook reading out the explanations and checking students’ answers in the key.” DDL is dangerous. This is no doubt one of the “reasons for teachers to be hesitant to introduce their students to DDL activities even if they are aware of the full range of concordance-based learning methods” (Götz & Mukherjee 2006: 51).

The whole mindset of DDL – and indeed of our ICT era – is completely at odds with the traditional teacher-oriented paradigm:

The instructor [plays] a more Socratic role, posing questions and guiding the learning process, rather than taking an ecclesiastical approach, providing ‘the word’ on a subject that the student is to ‘learn’ (memorize) and repeat back in some format. (Frans 2000: 24)

The potential threat to face is obvious, and it is not surprising that teachers are reluctant to make themselves psychologically if not literally redundant, whatever lip-service is paid to learner-centredness. Teachers are traditionally at the centre of the stage, and may not enjoy taking a back seat. They have been trained to be the knower, the *fons et origo* of language and pedagogy in the classroom. In many cultures, the teacher is not allowed not to know: admitting ignorance is unthinkable, and rather than doing so teachers invent a spurious answer on the spur of the moment. Similarly, it can be difficult having one’s authority questioned, something which DDL actively encourages. Teachers may

actually find themselves knowing less on particular language points than their students, as learners' findings can be quite sophisticated, contradicting traditional rules: "one student told me that the best thing... was that she felt able to contradict her teacher" (Aston 1997: 52). The same applies to technical expertise, and teachers may feel it undermines their role when the unexpected happens in the computer laboratory. This can be especially face-threatening when the learners are more technically sophisticated than the teacher, but being better than the students is not enough: the teacher is expected to be perfect.

As Johns (1991b: 36) points out, "one of the most striking aspects of the development of computer-assisted learning over the past 20 years has been the change in the assumptions made about the role of the teacher." This is just as true, if not more so, with DDL, as has been apparently since the very early stages. In particular, it "entails a shift in the traditional division of roles between student and teacher, with the student now taking on more responsibility for his or her learning, and the teacher acting as research director and research collaborator rather than transmitter of knowledge" (Johns 1988: 14). This partial transfer of power is not to be confused with an abnegation of responsibility, as the teacher assumes new roles instead. The teacher "has to learn to become a director and coordinator of student-initiated research" (Johns 1991a: 3), by "abandoning the role of expert and taking on that of research organiser" (Johns 1991b: 31). Some teachers may take to the changes more easily than others, but even those who are doubtful may be surprised how "liberating" it can be (Bernardini 2001a: 23), dropping the mask of perfect knower, passing an increasing measure of responsibility to the learners, finding out new things about the language along with them. The teacher is not replaced by the corpus, which is merely a source of data. The teacher's role in facilitating the interface and in fostering the appropriate kind of "researcher attitude" (Bernardini 2001a: 21) is crucial – a teacher who is sceptical to the core is unlikely to create the necessary atmosphere for a new approach to work.

Teachers may have self-doubts about issues other than face. As learners may seek comfort in rules, this can be even more important for teachers; even suggesting fuzziness can be taken as an admission of ignorance. Non-natives may feel particularly insecure in the face of variation (Kaltenböck & Mehlmauer-Larcher 2002). However, just as corpora can give learners the confidence to challenge their teachers, so they can give teachers the confidence to challenge received ideas about language by providing "access the combined intuitions of literally

thousands of native speakers together” (Frankenberg-Garcia 2005b: 192). In any case, it has become apparent that the expert or “successful user of English” (Prodromou 2003) may be a more relevant than the native speaker for many purposes.

Teachers may also be worried about their lack of expertise – not just in the target language and ICT in general, but in the specific way they come together in corpus linguistics. Teachers certainly need to be at ease with using corpus data before asking their students to do the same (Mauranen 2004b: 100), though personal experience suggests that most teachers using DDL are largely self-taught. For teachers as for learners, the important thing is to “get your hands dirty”, the very spirit of DDL itself (O’Keeffe & Farr 2003). Mention has already been made of introductions on-line, and training is being introduced in some courses. For a more thorough grounding, a number of teacher-training courses exist on line, such as Heinle’s *ELT Advantage with An Introduction to Corpora in English Language Teaching* by McCarthy, O’Keeffe and Walsh.<sup>1</sup>

## 5. RESOURCES

The lack of resources is a commonly cited problem. While money, as in all fields, provides access to some wonderful facilities, surprising things can be achieved with limited technology and freely available resources, especially via the Internet.

Among the better known corpora is the Bank of English (BoE), currently standing at around 500 million words, used in the COBUILD projects. Though expensive to buy and intended mainly for research purposes, a free interface (Collins WordbanksOnline<sup>2</sup>) to 56 million words allows a number of interesting interactions for learners, outlined extensively in Thomas (2002). Another large corpus of British English is the British National Corpus (BNC), 100 million words collected in the early 1990s and carefully prepared. This can be purchased for use with dedicated software (Xaira), but such uses tend to favour research rather than learning applications. As with the BoE, there is an official site which allows a number of interesting queries<sup>3</sup>; more useful for learning purposes perhaps is the interface<sup>4</sup> created by Davies at Brigham Young University. Davies has also created the useful Time corpus: the entire collection of *Time* magazine from 1923 to 2006, searchable by date. A recent addition is the 360-million-word Corpus of Contemporary American English, compiled directly from the Internet and updated twice yearly. The disadvantage of such automatic



collection is that it tends to include more background noise than in corpora such as the BNC, and may not be as representative of spontaneous speech in particular.

Some of these large corpora have been marked up to help with part-of-speech queries, and are searchable by genre or text type, comparing for example speech and writing, or legal and journalistic English – all highly desirable for teaching purposes. There are also a number of specialised corpora, especially in the fields of academic English; these include the Michigan Corpus of Academic Spoken English (MICASE), also with an on-line interface<sup>5</sup>, and corpora of British Academic Written English (BAWE) and British Academic Spoken English (BASE)<sup>6</sup>. Use may be found for parallel corpora, where texts exist alongside their translations in one or more languages. One commonly used is EuroParl, contrasting 11 languages of the European Union; available for download<sup>7</sup> or for on-line searches<sup>8</sup>.

British and American English unsurprisingly dominate, especially in the public domain. Where other varieties (or indeed other languages) are required, an alternative is to use the Internet as a corpus itself. Search engines such as Google are not without their appeal, but are not ideal as they are intended for content rather than form-based searches: this limits the kind of query that can be formulated and, just as importantly, the presentation of the results. Other tools have been developed specifically to exploit the web as corpus. WebCorp<sup>9</sup> can produce concordances as the output format, and restrict searches by date, textual domains, to British or American newspapers, and so on. Rather faster is Fletcher's WebConcordancer<sup>10</sup> software for direct searches in 34 languages. He is also in the process of compiling the very large (one billion word) Web Corpus of English from the Internet. WebBootCat (available with SketchEngine<sup>11</sup> in a free 30-day trial) allows the user to "seed" the Internet with specific search terms; it then automatically trawls the web for documents which contain all of these to create an "instant corpus".

The web-as-corpus approach is notoriously messy, and many prefer to create their own small corpus. This is particularly appropriate where learners have particular needs (cf. Braun 2005 on "pedagogically relevant corpora"), and it is not difficult nowadays to construct small, home-made corpora for specific purposes; Gavioli (2005) provides in-depth discussion of this. Without mark-up the possibilities will be reduced, but many software packages provide the basics of frequency lists, collocates, concordancers, and so on. One of the most widely cited in research papers is WordSmith Tools<sup>12</sup>; the free demonstration

version severely limits search possibilities and output, but the full version is relatively inexpensive. The tool is probably more complicated than most learners require (Kosem 2008), and simpler packages may be more appropriate. AntConc<sup>13</sup> is completely free and more accessible for non-specialist use. A number of other sites such as LexTutor<sup>14</sup> offer software packages which include, but are not limited to, corpus analysis tools.

Most of the research in DDL supposes the existence of computer labs; this makes all kinds of activities possible. However, this is not always the reality in every institution: there may be no computer room at all, it may be regularly unavailable at appropriate times, it may have insufficient computers, they may be old and slow, with no chance to download software, limited or no access to the Internet, no available technical support, and so on. Even in the best of material conditions, many teachers (and learners) may be reluctant to use computers due to all sorts of unseen eventualities – technical (insufficient expertise to cope with breakdowns or simply the unexpected), pedagogical (CALL in general is seen as inefficient, or an interruption to the “serious” learning), dynamic (abuse, lack of motivation), and so on.

First of all, there are a number of semi-technical solutions, the most obvious being to assign activities out of class, either at school or at home. DDL and corpora have even been used successfully in distance education (e.g. Boulton, in press; Collins 2000), although guidance is essential. An in-class alternative is to have a single focal point. J. Willis (1998) describes a series of activities using concordances on the blackboard; an overhead projector or a slide presentation is probably more practical in most cases (e.g. Estling Vannestål & Lindquist 2007). The teacher may also use a single computer and projector to demonstrate techniques and answer questions reactively (e.g. Tribble 2007). Where a small number of computers are available, students may work in pairs or small groups; not only is the collaborative aspect motivating for many, but pairing linguistically advanced learners with more ICT-literate partners may prove particularly fruitful, ensuring opportunities for each to contribute in their own way.

Secondly, the basic activities, procedures and techniques can be conducted using printed materials alone – what Gabrielatos (2005) calls “soft” DDL. Certainly it seems that “DDL activities can be plotted on a cline of learner autonomy, ranging from teacher-led and relatively closed concordance-based activities to entirely learner-centred corpus-browsing projects” (Mukherjee 2006: 12). While prepared materials tend to be seen mainly as a stepping-stone to full-blown hands-on

concordancing, they have advantages in themselves, reducing the cognitive burden and allowing learners to gain an insight into the techniques involved using uncluttered, selected data and without the technological difficulties. A number of papers show learners using paper-based materials successfully as a reference source (Boulton 2008b, 2009) as well as for learning (Allan 2006; Koosha & Jafarpour 2006).

There is currently a dearth of published materials of this nature available ready to use (cf. Boulton 2008c). Although groundbreaking use has been made of corpora as a source of examples and to inform reference materials and coursebooks (e.g. Biber et al. 1999; McCarthy et al. 2006), publishers have yet to take up 25-year-old suggestions to incorporate DDL activities into teaching materials themselves (Higgins & Johns 1984: 93). Occasional exercises can be found in materials produced by Athelstan, such as *Business Phrasal Verbs and Collocations* (Burdine & Barlow 2007), but to date only two DDL textbooks exist: *Exploring Academic English: A Workbook for Student Essay Writing* (Thurstun & Candlin 1997) and *Concordances in the Classroom* (Tribble & Jones 1997). The fact that both of these are over 10 years old shows the difficulties involved in preparing general-purpose “off the peg” materials, and while they are still widely cited, this tends to be as sources of example activities rather than for use in their own right (Boulton 2008c).

Many teachers and researchers prefer to produce their own materials to target particular language points in ways relevant to their own learners; papers can be found describing courses and materials of this nature from Johns (1991a, 1991b) to Boulton (2008d). One major problem, as these authors point out, is that they are extremely time-consuming to produce for one-off usage. Fortunately, the DDL community is such that there are a number of on-line sources where materials can be downloaded ready for use or for inspiration. The first port of call for many is Johns’ DDL page<sup>15</sup>, as well as his “kibitzers”<sup>16</sup>, based on individual language points encountered in learners’ written texts. Barlow’s CorpusLab<sup>17</sup> allows teachers and researchers to upload their own materials for sharing. Other teachers have their own sites with downloadable materials, such as Sripicharn in Thailand<sup>18</sup>, or Estling Vannestål and colleagues in Sweden<sup>19</sup>.

It has not been possible in the context of this short section to describe the relative merits or uses of different tools, materials and approaches; nor is it possible to cover all the resources which are

available. As so often, and in the spirit of DDL, the best solution is simply to explore for oneself.

## 6. CONCLUSION

Dramatic claims for new methodologies generally cause uneasiness in the teaching profession, which has seen many pendulum swings over the years. Gabrielatos (2005: n.p.) therefore recommends that “the use of corpora should not be treated as an alternative to, or rival of, existing teaching approaches, but as a welcome addition.” DDL does not reject past practice, it builds on it, drawing on existing skills highly prized in the communicative classroom, and adapting them to cutting-edge technology; the combination provides not only “new materials but also with a whole new range of things to observe as well as a new way to observe them” (Gavioli 2005: 40). Two decades ago, its founding father described it as “innovative and possibly revolutionary” (Johns 1991b: 27), while Butler (1990: 344), even though cautioning against dramatic claims for new CALL technologies, nevertheless claimed that “the hyperbole in this case [DDL] is perhaps more justified.” Even then, Johns (1988: 9) divided the field into enthusiasts and sceptics, a situation which still prevails today. The users are found mainly in research environments, while regular teachers, if they are aware of learning applications of corpora at all, tend to remain sceptical for some of the reasons discussed in this article. And yet they are the ones we need to convince – “ordinary teachers and learners in ordinary classrooms” (Mauranen 2004b: 208).

Many of these reasons, it has been suggested here, are comprehensible. But each individual worry can be countered, and “any teacher or student can readily enter the world of the corpus and make the language useful in learning” (Sinclair 2004: 297). The fact that the “trickle-down” effect from research to teaching practices has not become the “torrent” predicted by Leech (1997: 2) suggests a deeper malaise, leaving the feeling that the practical objections are perhaps camouflage for more profound theoretical concerns about the nature of learning, and more especially of teachers’ and learners’ roles. Such fears are therefore not to be dismissed lightly. Nevertheless, as teachers, we are ultimately here for our learners, not for ourselves. It certainly requires time and effort, and a little perseverance, but more importantly a willingness to experiment with hands-on concordancing oneself. These are an investment for the future – our learners’ and our own: as

Conrad (1999: 3) puts it, “practising teachers and teachers-in-training... owe it to their students” – and also, ultimately, to themselves.

#### NOTES

1. <http://eltadvantage.ed2go.com/cgi-bin/eltadvantage/oic/newcrsdes.cgi?course=3ce&name=eltadvantage&departmentnum=EL>
2. <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
3. <http://sara.natcorp.ox.ac.uk/lookup.html>
4. <http://corpus.byu.edu/bnc/x.asp>
5. <http://quod.lib.umich.edu/m/micase/>
6. <http://www2.warwick.ac.uk/fac/soc/al/research/projects/resources/>
7. <http://www.statmt.org/europarl/>
8. <http://www.let.rug.nl/tiedeman/OPUS/lex.php>
9. <http://www.webcorp.org.uk/wcadvanced.html>
10. <http://webascorpus.org/>
11. <http://www.sketchengine.co.uk/>
12. <http://www.lexically.net/wordsmith>
13. <http://www.antlab.sci.waseda.ac.jp/software.html>
14. <http://www.lextutor.ca/>
15. <http://www.eisu2.bham.ac.uk/johnstf/timconc.htm>
16. <http://www.eisu2.bham.ac.uk/johnstf/timeap3.htm#revision>
17. <http://www.corpuslab.com/>
18. [http://www.geocities.com/tonypgnews/units\\_index\\_pilot.htm](http://www.geocities.com/tonypgnews/units_index_pilot.htm)
19. <http://www.vxu.se/hum/utb/amnen/engelska/kig/>

#### REFERENCES

- Adolphs, S. 2006. *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. London: Routledge.
- 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.
- Aston, G. 1996. The British National Corpus as a language learner resource. In S. Botley, J. Glass, A. McEnery & A. Wilson (Eds.), *Proceedings of TALC 1996. UCREL Technical Papers*, 9, 178-191.
- Aston, G. 1997. Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & P. Melia (Eds.), *Practical Applications in Language Corpora (PALC 97)* (pp. 51-62). Lodz: Lodz University Press.
- Aston, G. (Ed.) 2001. *Learning with Corpora*. Houston: Athelstan.

- 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.
- Bernardini, S. 2001a. Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15-36). Amsterdam: John Benjamins.
- Bernardini, S. 2001b. 'Spoilt for choice.' A learner explores general language corpora. In G. Aston (Ed.), *Learning with Corpora* (pp. 220-249). Houston: Athelstan.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Pearson.
- Boulton, A. 2008a. Evaluating corpus use in language learning: State of play and future directions. Paper presented at *AAFL 2008 (American Association for Corpus Linguistics)*. Provo, UT: Brigham Young University.
- Boulton, A. 2008b. Looking (for) empirical evidence for DDL at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Practical Applications in Language and Computers*. Frankfurt: Peter Lang.
- Boulton, A. 2008c. 'Off-the-peg' materials for data-driven learning. Paper presented at *New Trends in Corpus Linguistics for Language Teaching and Translation Studies: In Honour of John Sinclair*. Granada, Spain: University of Granada / University Jaume I.
- Boulton, A. 2008d. DDL: Reaching the parts other teaching can't reach? In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8<sup>th</sup> Teaching and Language Corpora Conference* (pp. 38-44). Lisbon, Portugal: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Boulton, A. 2009. Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21/1.
- Boulton, A. in press. Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Teaching and Language Corpora (TaLC 7)*. Amsterdam: Rodopi.
- 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.
- Brown, D. 2007. Language learner motivation and the role of choice in ESP listening engagement. *ASp*, 51-52, 159-187.
- Brown, H., Tarone, E., Swan, M., Ellis, R., Prodromou, L., Bruton, A., Johnson, K., Nunan, D., Oxford, R., Goh, C., Waters, A. & Savignon, S. 2007. Forty years of language teaching. *Language Teaching*, 40, 1-15.

- Burdine, S. & Barlow, M. 2007. *Business Phrasal Verbs and Collocations*. Houston: Athelstan.
- Burnard, L. & McEnery, T. (Eds.) 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.
- Butler, J. 1990. Concordancing, teaching and error analysis: Some applications and a case study. *System*, 18/3, 343-349.
- Carter, R. 1998. Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52/1, 43-56.
- 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 & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp. 3-16). Amsterdam: Rodopi.
- Chambers, A. & O'Sullivan, I. 2004. Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16/1, 158-172.
- Chujo, K., Utiyama, M. & Nishigaki, C. 2007. Towards building a usable corpus collection for the ELT classroom. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, p. 47-69.
- Ciezielska-Ciupek, M. 2001. Teaching with the Internet and corpus materials: Preparation of the ELT materials using the Internet and corpus resources. In B. Lewandowska-Tomaszczyk (Ed.), *PALC 2001: Practical Applications in Language Corpora* (pp. 521-531). Frankfurt: Peter Lang.
- Cobb, T. 2006. The case for computer-assisted extensive reading. *LexTutor*. Available online: <<http://tesl-ej.org/ej32/a1.html>>
- Collins, H. 2000. Materials design and language corpora: A report in the context of distance education. In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 51-63). Frankfurt: Peter Lang.
- Conrad, S. 1999. The importance of corpus-based research for language teachers. *System*, 27/1, 1-18.
- Conrad, S. 2000. Will corpus linguistics revolutionize grammar teaching in the 21<sup>st</sup> century? *TESOL Quarterly*, 34, 548-560.
- Cook, G. 1998. The uses of reality: A reply to Ronald Carter. *ELT Journal*, 52/1, 57-63.
- Cook, G. 2001. 'The philosopher pulled the lower jaw of the hen': Ludicrous invented sentences in language teaching. *Applied Linguistics*, 22/3, 366-387.
- Cook, V. 2002. The functions of invented sentences: A reply to G. Cook. *Applied Linguistics*, 23/2, 262-269.
- Cosmides, L. & Tooby, J. 1992. Cognitive adaptations for social exchange. In H. Barkow, L. Cosmides and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163-228). Oxford: Oxford University Press.
- Cresswell, A. 2007. Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda & J. Santana

- (Eds.), *Corpora in the Foreign Language Classroom* (pp. 267-287). Amsterdam: Rodopi.
- Davis, B. & Russell-Pinson, L. 2004. Concordancing and corpora for K-12 teachers: Project MORE. In U. Connor & T. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 147-169). Amsterdam: Rodopi.
- Dellar, H. 2003. What have corpora ever done for us? *DevelopingTeachers.com*. Available online: <[http://www.developingteachers.com/articles\\_tchtraining/corporapf\\_hugh.htm](http://www.developingteachers.com/articles_tchtraining/corporapf_hugh.htm)>
- Estling Vannestål, M. & Lindquist, H. 2007. Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19/3, 329-350.
- Farr, F. 2008. Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17/1, 25-43.
- Flowerdew, L. 2001. The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry & R. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and practice* (pp. 363-379). Amsterdam: John Benjamins.
- Flowerdew, L. 2008. Pedagogic value of corpora: A critical evaluation. In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8<sup>th</sup> Teaching and Language Corpora Conference* (pp. 115-119). Lisbon, Portugal: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Frand, J. 2000. The information-age mindset: Changes in students and implications for higher education. *EDUCAUSE Review*, 35/5, 14-24.
- Frankenberg-Garcia, A. 2005a. A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18/3, 335-355.
- Frankenberg-Garcia, A. 2005b. Pedagogical uses of monolingual and parallel concordances. *ELT Journal*, 59/3, 189-198.
- 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. Available online: <<http://tesl-ej.org/ej32/a1.html>>
- Gaskell, D. & Cobb, T. 2004. Can learners use concordance feedback for writing errors? *System*, 32/3, 301-319.
- Gavioli, L. 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
- Götz, S. & Mukherjee, J. 2006. Evaluation of data-driven learning in university teaching: A project report. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods* (pp. 49-67). Frankfurt: Peter Lang.
- Hadley, G. 2002. Sensing the winds of change: An introduction to data-driven learning. *RELC Journal*, 33/2, 99-124. Available online: <<http://www.nuis.ac.jp/~hadley/publication/windofchange/windsofchange.htm>>



- Hidalgo, E., Quereda, L. & Santana, J. (Eds.), 2007. *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.
- Higgins, J. & Johns, T. 1984. *Computers in Language Learning*. London: Collins.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Inkster, G. 1997. First catch your corpus: Building a French undergraduate corpus from readily available textual resources. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 267-276). Harlow: Addison Wesley Longman.
- Johns, T. 1986. Micro-Concord: A language learner's research tool. *System*, 14/2, 151-162.
- Johns, T. 1988. Whence and whither classroom concordancing? In P. Bongaerts, P. de Haan, S. Lobbe & H. Wekker (Eds.), *Computer Applications in Language Learning* (pp. 9-27). Dordrecht: Foris.
- Johns, T. 1991a. Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom Concordancing. English Language Research Journal*, 4, 1-16.
- Johns, T. 1991b. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Johns & P. King (Eds.), *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 (Eds.), *Teaching and Language Corpora* (pp. 100-115). Harlow: Addison Wesley Longman.
- Kaltenböck, G. & Mehlmauer-Larcher, B. 2002. Teaching ESP: How text corpora can help. In A. Pulverness (Ed.), *IATEFL 2002: Dublin Conference Selections* (pp. 31-33). Whitstable: IATEFL.
- Kennedy, C. & Miceli, T. 2001. An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5/3, 77-90.
- Kettemann, B. & Marko, G. 2004. Can the L in TALC stand for Literature? In G. Aston, S. Bernardini & D. Stewart (Eds.), *Corpora and Language Learners* (pp. 169-193). Amsterdam: John Benjamins.
- 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.
- Kosem, I. 2008. User-friendly corpus tools for language teaching and learning. In A. Garcia-Garcia (Ed.), *Proceedings of the 8<sup>th</sup> Teaching and Language Corpora Conference* (pp. 183-192). Lisbon, Portugal: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Kübler, N. (Ed.) in press. *Teaching and Language Corpora (TaLC 7)*. Amsterdam: Rodopi.
- 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. Available online: <[http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm)>
- Laufer, B. & Hulstijn, J. 2001. Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22/1, 1-26.
- Lee, D. & Swales, J. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25, 56-75.
- Leech, G. 1997. Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 1-23). Harlow: Addison Wesley Longman.
- Mauranen, A. 2004a. Spoken corpus for an ordinary learner. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 89-105). Amsterdam: John Benjamins.
- Mauranen, A. 2004b. 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. Available online: <<http://www.cambridge.org/us/esl/Touchstone/teacher/images/pdf/CorpusBookletfinal.pdf>>
- McCarthy, M., McCarten, J. & Sandiford, H. 2006 *Touchstone 4. Teacher's edition*. Cambridge: Cambridge University Press.
- Milton, J. 1996. Exploiting L1 and L2 corpora for computer assisted language learning design: The role of an interactive hypertext grammar. In S. Botley, J. Glass, A. McEnery & A. Wilson (Eds.), *Proceedings of TALC 1996. UCREL Technical Papers*, 9, 233-243.
- Mishan, F. 2004. Authenticating corpora for language learning: A problem and its resolution. *ELT Journal*, 58/3, 219-227.
- 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.
- Noss, R. & Pachler, N. 1999. The challenge of new technologies: Doing old things in new ways, or doing new things? In P. Mortimore (Ed.), *Understanding Pedagogy and its Impact on Learning* (pp. 195-211). London: Paul Chapman.

- O'Keeffe, A. & Farr, F. 2003. Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights. *TESOL Quarterly*, 37/3, 389-418.
- O'Keeffe, A., McCarthy, M. & Carter, R. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, I. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19/3, 269-286.
- Prodromou, L. 2003. In search of the successful user of English: How a corpus of non-native speaker language could impact on EFL teaching. *Modern English Teacher*, 12/2, 5-14.
- Renouf, A. 1997. Teaching corpus linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 255-266). Harlow: Addison Wesley Longman.
- Richards, J. & Rodgers, T. 2001. *Approaches and Methods in Language Teaching*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Römer, U. 2006. Where the computer meets language, literature, and pedagogy: Corpus analysis in English studies. In A. Gerbig & A. Müller-Wood, (Eds.), *How Globalization Affects the Teaching of English: Studying Culture Through Texts* (pp. 81-109). Lampeter: Mellen Press.
- Rüschhoff, B. 2004. *Data-driven learning: The idea*. Available online: <[http://www.ecml.at/projects/voll/rationale\\_and\\_help/booklets/resources/menu\\_booklet\\_ddl.htm](http://www.ecml.at/projects/voll/rationale_and_help/booklets/resources/menu_booklet_ddl.htm)>
- Scott, M. & Tribble, C. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Sealey, A. & Thompson, P. 2004. What do you call the dull words? Primary school children using corpus-based approaches to learn about language. *English in Education*, 38/1, 80-91.
- 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.
- Seidlhofer, B. 2002. Pedagogy and local learner corpora: Working with learner-driven data. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 213-234). Amsterdam: John Benjamins.
- Sinclair, J. 2003. *Reading Concordances: An Introduction*. Harlow: Longman.
- Sinclair, J. (Ed.) 2004. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Stevens, V. 1991. Concordance-based vocabulary exercises: A viable alternative to gap-filling. In T. Johns & P. King (Eds.), *Classroom Concordancing*. *English Language Research Journal*, 4, 47-61.
- Stubbs, M. 2001. Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics*, 22, 149-172.

- Sun, Y-C. 2003. Learning process, strategies and web-based concordancers: A case-study. *British Journal of Educational Technology*, 34/5, 601-613.
- Sun, Y-C. & Wang, L-Y. 2003. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16/1, 83-94.
- Tan, M. 2003. Language corpora for language teachers. *Journal of Language and Learning*, 1/2, 98-105. Available online: <[http://www.shakespeare.uk.net/journal/jllearn/1\\_2/tan1.html](http://www.shakespeare.uk.net/journal/jllearn/1_2/tan1.html)>
- Thomas, J. 2002. *A Ten-Step Introduction to Concordancing through the Collins COBUILD Corpus Concordance Sampler*. Brno: Masaryk University. Available online: <<http://web.quick.cz/jaedth/Introduction%20to%20CCS.htm>>
- Thompson, P. 2006. Assessing the contribution of corpora to EAP practice. In Z. Kantaridou, I. Papadopoulou & I. Mahili (Eds.), *Motivation in Learning Language for Specific and Academic Purposes* [CDROM]. Macedonia: University of Macedonia.
- Thurstun, J. & Candlin, C. 1997. *Exploring Academic English: A Workbook for Student Essay Writing*. Sydney: CELTR.
- Tribble, C. 1997. Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk & P. Melia (Eds.), *Practical Applications in Language Corpora (PALC 97)* (pp. 106-117). Lodz: Lodz University Press.
- Tribble, C. 2007. Managing relationships in professional writing. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp. 289-308). Amsterdam: Rodopi.
- Tribble, C. & Jones, G. 1997. *Concordances in the Classroom*. 2<sup>nd</sup> edition. Houston: Athelstan.
- Tsui, A. 2005. ESL teachers' questions and corpus evidence. *International Journal of Corpus Linguistics*, 10/3, 335-356.
- Whistle, J. 1999. Concordancing with students using an 'off-the-web' corpus. *ReCALL*, 11/2, 74-80.
- Wible, D., Chien, F., Kuo, C-H. & Wang, C. 2002. Toward automating a personalized concordancer for data-driven learning: A lexical difficulty filter for language learners. In B. Kettemann, & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis* (pp. 147-154). Amsterdam: Rodopi.
- Widdowson, H. 2000. On the limitations of linguistics applied. *Applied Linguistics*, 21/1, 3-25.
- Willis, D. 2003. *Rules, Patterns and Words*. Cambridge: Cambridge University Press.
- Willis, J. 1998. Concordances in the classroom without a computer. In B. Tomlinson (Ed.), *Materials Development in Language Teaching* (pp. 44-66). Cambridge: Cambridge University Press.
- Yoon, H. & Hirvela, A. 2004. ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, 13/4, 257-283.

**ALEX BOULTON**  
CRAPEL-ATILF/CNRS, NANCY UNIVERSITÉ  
3 PLACE GODEFROI DE BOUILLON  
BP 3397  
54015 NANCY – CEDEX  
E-MAIL: BOULTON@UNIV-NANCY2.FR  
PHONE: (F) 383967130  
FAX: (F) 383967132