

How can DDL be used in language teaching?

Gaëtanelle Gilquin¹ and Sylviane Granger²

¹ F.N.R.S. – Centre for English Corpus Linguistics
Université catholique de Louvain
Gaetanelle.Gilquin@uclouvain.be

² Centre for English Corpus Linguistics
Université catholique de Louvain
Sylviane.Granger@uclouvain.be

1. THE PEDAGOGICAL FUNCTIONS OF DDL

Data-Driven Learning (DDL) consists in using the tools and techniques of corpus linguistics for pedagogical purposes. This type of approach presents several advantages. The first obvious one is that it brings authenticity into the classroom. Not only do corpora make it possible to expose learners to authentic language, but they can actually present them with a large number of authentic instances of a particular linguistic item. This ‘condensed exposure’ (Gabrielatos 2005: 10) can, among others, contribute to vocabulary expansion or heightened awareness of language patterns.

Secondly, DDL has an important corrective function. Learners, by comparing their own writing with data produced by (native) expert writers or by consulting a learner corpus where errors have been annotated (see Section 2 on error-tagged learner corpora), can find the help they need to correct their own interlanguage features (misuse, overuse, underuse) and thus improve their writing. As pointed out by Nesselhauf (2004: 140), this is ‘particularly useful for points which have already been covered in the classroom, possibly even repeatedly, but which the learners nevertheless still get wrong’, that is for so-called fossilised errors.

The DDL approach also has the advantage of including an element of discovery which arguably makes learning more motivating and more fun. In the DDL literature, learners are described, alternatively, as travellers (Bernardini 2001: 22), researchers (Johns 1997: 101), or detectives with John’s (1997: 101) slogan ‘Every student a Sherlock Holmes’. By means of various activities (see Section 3), learners are encouraged to observe corpus data, make hypotheses and formulate rules in order to gain insights into language (inductive approach) or check the validity of rules from their grammar or textbook (deductive approach). They thus become more involved, more active and, ultimately, more autonomous in the learning process. The learner is ‘empowered’ (Mair 2002), which has the effect of boosting his/her confidence and self-esteem.

More generally, learners can acquire (or at least refine) a number of crucial learning skills through the use of DDL. O’Sullivan (2007: 277) lists the following: ‘predicting, observing, noticing, thinking, reasoning, analysing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorising, hypothesising, and verifying’. These skills can be used to explore language, but since they are general cognitive skills, they may also be transferred to other fields of study.

2. THE DDL MATERIAL

In order to adopt a DDL methodology, two main resources are needed: a corpus and a tool to exploit the corpus (concordancing software). The choice of the corpus is crucial for, as rightly emphasised by Whistle (1999: 78), '[t]he value and usefulness of the concordance will be largely determined by the corpus'. It would probably not be wrong to say that any type of corpus may be used in DDL, and indeed, the literature on DDL mentions quite a large range of corpora: written, spoken or multimodal, monolingual or bilingual, general or specialised, native or non-native, tagged or untagged, etc. As can be expected, however, particular corpora are best suited for certain purposes. Consider the example of bilingual corpora. Such corpora may be used for two main purposes. They may help translation trainees, by 'drawing [their] attention to (un)typical solutions for typical problems found by *mature*, expert translators' (Bernardini 2004: 20; see also the chapters by Kenning and Aston & Kübler, this volume). They may also be used, with students who share the same mother tongue, to establish equivalences between the mother tongue and the target language, and especially, as Johns (2002: 114) puts it, to 'wean (...) students from the myth of one-to-one correspondence between first and second language'.

Whatever the type of corpus chosen, one important issue is that of its authenticity. Of course, in a way, corpora are always authentic in the sense that they contain naturally-occurring language data. However, scholars like Widdowson (2000), making a distinction between text production and text reception, argue that corpora may lack authenticity at the receptive end, even though they were initially authentic. Thus, learners may find it hard to relate to texts that were produced in a different culture, within a context that is not necessarily familiar to them. The texts are therefore likely to 'remain an "anonymous mass" to the learners' (Braun 2006: 26). In fact, Sripicharn (2004) demonstrates that, while native speakers are able to contextualise concordance lines (by identifying the setting of the concordances and the text type from which they are extracted), learners usually fail to show such ability. Several solutions have been proposed in the literature to solve this problem and help learners "authenticate" (Widdowson 2003: 66) the materials they are working with, for example involving learners in the creation of the corpus (Aston 2002), or using a corpus of recent news items (Chambers 2005: 120) or a small specialized corpus drawn from the 'genres which have relevance to the needs and interests of the learners' (Tribble 1997: 2).

Two types of corpora that appear particularly helpful for the process of authentication are the 'pedagogic corpus' (Willis 2003: 163) and the 'local learner corpus' (Seidlhofer 2002). The former consists of the texts used in the classroom to support teaching (texts from the learners' coursebooks, plus any additional texts that the teacher may have brought into the classroom). Although such a corpus may partly consist of concocted texts (in cases where some of the texts used in class were invented), these texts have already been processed for meaning by the learners, and are therefore better contextualized and more directly relevant to them (Granger forthcoming). Alternatively, a corpus may also be created that comprises transcriptions of the lectures attended by the students, as experimented by Flowerdew (1993) in an English for Specific Purposes course – although building such a pedagogic corpus will naturally be more time-consuming.

The second type of corpus that seems promising in terms of authenticity is the local learner corpus. Learner corpora, which include data produced by non-native speakers of the language, have remained relatively discreet on the DDL scene up to now (Nesselhauf 2004: 140). Yet, learner corpora can be extremely useful for form-focused instruction (see e.g. Granger & Tribble 1998 or Seidlhofer 2000), because they present students with typical interlanguage features, especially when the data were produced by learners from the same mother tongue background as the students. Local learner corpora go even one step further, as they contain data produced by the very same students who will be using the corpus. They are thus 'both participants in and analysts of their own language use' (Seidlhofer 2002: 213), and

the interlanguage features represented in the corpus are the features of their own interlanguage. This also means that the teacher can provide ‘tailor-made feedback’ (Mukherjee 2006: 19) to the learners, either as a group or individually.

Another important issue when it comes to the choice of a corpus is annotation. Corpora can be used as raw text, i.e. with no annotation of any kind, or they can be tagged with additional information such as part-of-speech (POS-tagging), syntactic structure (parsing) or, in the case of learner corpora, errors (error-tagging). Raw corpora offer numerous possibilities for the exploration of language by learners. However, they also have their limitations. For example, a raw corpus may involve a lot of editing (by the teacher or learner) to get rid of unwanted concordance lines, whereas a POS-tagged or parsed corpus may help refine the search query (for example, selecting *to* as a preposition and not as an infinitive marker) and thus reduce the amount of necessary editing. As for error-tagging, it makes it much easier to notice interlanguage features and often comes with possible corrections. It should, however, be borne in mind that the annotation of tagged corpora may sometimes be problematic (mistagging, inconsistencies) and that it always reflects a certain theoretical perspective that may not be shared by the teacher. Moreover, as Gabel (2001: 284) suggests, tagging is not easy to implement within a school context – at least when, as is often the case, the teacher creates his/her own corpus rather than using a ready-made corpus.

Corpora are of little help if they are not combined with a tool to exploit them. Breyer (2006: 173) rightly remarks that the role of the concordancing software ‘has hitherto been somewhat overlooked in the discussion about the application of a corpus technology in language pedagogy’. Yet, like the corpus, it is of crucial importance for the ultimate success of DDL. Higgins (1991) makes a distinction between ‘research’ concordancers and ‘classroom’ concordancers. While such a distinction is not clear-cut, since research concordancers can be used in the classroom and vice versa (as Higgins himself notes), it is still true that in order to be used in the classroom, a concordancer should ideally possess a number of characteristics. Stevens (1995: 2) gives the following list: the concordancer should be fast and responsive, it should load quickly and allow interruption at any point (with the option to work with the data already loaded at that point), it should be possible to look for more than one word at the same time and for strings of words, to use Boolean operators (AND, OR, NOT, etc) and wild cards (to indicate any unspecified characters), and to sort the output instantaneously. To this wish list, we would like to add the possibility of easily creating exercises from the concordances. This is the case, for example, of Multiconcord (King & Woolls 1996), which allows the user to produce several types of cloze tests, or WordSmith Tools (Scott 2004), which has an option for blanking out the search-words in concordances. Last but not least, a classroom concordancer should be user-friendly. While this is an important feature for any concordancer, this is even more essential in the case of a concordancer intended for learners, who have to ‘get to grips with new material (the corpora), new technology (the software) and a new approach (DDL) all at once’ (Boulton 2008a: 38). After having dealt with the corpora and the software, we now turn to the DDL approach proper, and show how corpora can be used in language teaching.

3. THE OPERATIONALIZATION OF DDL

The range of activities that are possible in DDL is wide and, as Breyer (2006: 162) puts it, ‘limited only by the imagination of the user’. Space prevents us from reviewing all these possibilities, but in this section, we give a broad overview of the way DDL may be operationalized and show how the choice of presentation and activity depends on the learning context (e.g. one-to-one consultation vs. classroom activity), the level of the learners

(language proficiency and familiarity with DDL) and the topic investigated (e.g. vocabulary or discourse). Note that we will not deal with cases where the results of corpus analysis are used to ‘inform teaching decisions’ or ‘prepare teaching materials’ (Johns 1988: 20), although this is sometimes also referred to as DDL. Here, we will only consider cases where the learner directly interacts with the corpus.

It is probably fair to say that most DDL activities involve concordances of some sort. Concordances, however, may be presented in various ways. The concordance lines may be truncated (so-called KWIC – Key-Word-In-Context – view) or take the shape of a complete sentence, the whole concordance may be provided to the learner or just a selection of it, the concordance lines may be edited or presented in their original form, and they may be shown on screen or printed on a handout. Each presentation has its pros and cons. The KWIC view may be confusing, especially for beginners. Johns (1986: 157) observes that learners’ first reaction is often to complain about the ‘unfinished sentences’. However, the KWIC view, with all the occurrences of the search word aligned under one another, makes patterns more visible than the sentence view, and Boulton (2009a) actually reports an experiment with lower-intermediate learners where KWICs provided better results than full sentence contexts. The next question is whether the concordancer output should be used as is or whether it should be manipulated in some way. Manipulation may involve the selection of a subset of the concordance, often with the aim of reducing the data to manageable quantities. Several criteria have been proposed in the literature to perform this selection, including readability (the most difficult concordance lines are discarded, cf. Kuo et al. 2001), frequency (only the concordance lines illustrating the most frequent uses are kept, cf. Levy 1990: 180) and usefulness (only those concordance lines that are judged useful are kept, cf. Tribble 1997: 4). However, Gabrielatos (2005: 18) rightly points out that ‘[t]his manipulation should be carried out with the understanding that the adapted samples are not good guides to the frequency of a language item’. Random selection, as opposed to principled selection, may help avoid this bias and maintain some semblance of fidelity to the data (cf. Johns 2002: 110). Manipulation may also consist of editing the concordances, and in particular, simplifying them (cf. Gabrielatos 2005: 18). Boulton (2009b) is not in favour of such a practice, as this, according to him, undermines the ‘authenticity’ advantage of DDL and does not prepare learners ‘for the realities of the authentic language we are presumably preparing them for’. Yet, Nesselhauf’s (2004: 143–44) discussion of a concordance of the verb *suggest* in the LOCNESS corpus (a corpus of argumentative essays by British and American students) and the German subcorpus of the International Corpus of Learner English is a good example of why manipulation is sometimes necessary in DDL (especially with beginners):

As the lines are now, they could be confusing for learners in many respects: there is at least one typographical error (*suggested than* instead of *that* in LOCNESS); one of the occurrences of *to* after *suggest* does not constitute wrong complementation (*could suggest to her two colleagues*); and, as *suggest + ing* only occurs in the learner but not in the native speaker corpus, learners might even come to the conclusion that this construction is not possible in English.

Finally, concordances may be presented on screen or on a handout (what Gabrielatos [2005: 13] describes as the hard vs. soft version of DDL). The choice of the mode of presentation depends on the availability or not of the necessary hardware and software (see Section 5), but it is also a function of learners’ level. Thus, Boulton (2008a: 38) suggests that ‘DDL in early stages can eliminate the computer from the equation by using prepared materials on paper’ (see also Whistle 1999: 78). An interesting alternative, suggested by Charles (2007), is to let the students work on a computer in class and to then provide them with a printout of the concordances for further study at home, or simply as a record of what has been done in class.

DDL activities may be located along a cline ranging from teacher-led to learner-led (Mukherjee 2006: 12; see also Gabrielatos 2005: 11). At the teacher-led end, we find relatively controlled tasks such as cloze tests and fill-in exercises. At the learner-led end, we find what Bernardini (2004: 22) calls ‘discovery learning’, which consists in ‘brows[ing] large and varied text collections in open-ended, exploratory ways’. As we move from one end of the cline to the other, learners have more and more freedom and bear more and more responsibility for their own instruction, deciding, for example, what they are going to investigate and how they want to go about it. This explains why there should ideally be a gradual shift from one end to the other, and why teacher-led activities tend to be better suited to beginners, whereas discovery learning is often claimed to be most appropriate for ‘very advanced learners who are filling in gaps in their knowledge rather than laying down the foundations’ (Hunston 2002: 171). In-between totally teacher-led DDL and totally learner-led DDL, there is a whole range of activities, with various types of ‘filters’ exercised by the teacher (Gavioli 2005: 30). By way of illustration, here are a few activities that one could propose to students. Learners could be shown a concordance sorted alphabetically and be encouraged to notice the repetition of certain lexical chunks or be asked to group the patterns in a meaningful way. They could also be given a series of blanked out concordances illustrating different contexts of a word and be required to find the missing word. Alternatively, the concordances could come from a bilingual corpus, so that learners can use the translations to help them find the missing word more easily. The potential of DDL for editing one’s own work has already been mentioned earlier, and many studies in the literature show how students can use corpora to revise their work, either by correcting problems underlined by the teacher or by deciding themselves what they want to check with the help of the concordancer (Kennedy & Miceli 2001: 81). One use that in our opinion tends to be neglected in the DDL literature is the use of the concordancer as a ‘sleeping resource’ (Johns 1988: 22), to help learners when the need arises. In the same way as it has become natural to have a dictionary in most language classrooms and to consult it in case of doubt, we would like, one day, to see every classroom equipped with a computer, and students using it to query a corpus in order to answer a question that has suddenly arisen during the lesson.

In what precedes, we have mainly dealt with activities involving concordances. While this is probably the main component of DDL (and in fact, according to some strict definitions, its only component, cf. Johns & King 1991: iii), other possibilities exist. Thus, frequency lists, which list all the words of the corpus in descending order of frequency, may prove to be a valuable resource as well. Aston (2001) suggests using them in (literary) text-analysis, as a means of learning more about the subject of the text and its meaning. It is also possible to compare two frequency lists built on the basis of different corpora. The comparison could involve two varieties of English (e.g. British vs. American English) or two genres or text types (e.g. fiction vs. journalese, speech vs. writing), with the aim of making learners more sensitive to language variation. Learners could also compare a frequency list representing learner production (ideally, their own production) and one representing expert production, which would make them aware of the words that tend to be underused or overused by learners. The words from the frequency lists can then be used as starting points for concordance analysis. Most concordancers nowadays also make it possible to automatically extract collocates and word clusters. These, like frequency lists, may serve as a good starting point for the further analysis of language – in this case, language in its phraseological dimension (cf. Cheng 2007). Another way to exploit corpora in the classroom is to read entire portions of the corpus. This is what Charles (2007: 295) recommends for the study of rhetorical functions: a particular search item is used as ‘a probe to locate the part of a text in which a given rhetorical function may occur’, and the context is then expanded to the whole paragraph or the whole text to see how the function is expressed.

The choice of presentation and activity depends on a number of factors, among which the learning context, the level of the learners and the topic investigated. Thus, actual use of the corpus ('hard' version of DDL) and focus on the learner's individual needs may be easier in the context of a one-to-one consultation (such as described by Johns 2002: 111ff) than in the context of a classroom activity – and virtually impossible to implement if a sufficient number of computers is not available in the classroom. An activity meant to be carried out as part of a homework assignment or in distance education (Collins 2000) may have to contain more explicit instructions and 'signposts' than one which takes place in class, with the teacher as a guide and the fellow students as 'travel companions'. The learner's level (both in terms of language proficiency and familiarity with DDL) has already been shown to be important when deciding on a particular presentation of the data and a specific activity. This probably explains why scholars do not seem to agree on the level of the learners for whom DDL is appropriate. Depending on the methods they have in mind, DDL may be suitable for a given audience or not. In fact, it has been argued that DDL is possible with all learners (even beginners), but that (1) the method has to be adapted to the learners' level (Hadley 2002) and (2) results may vary, with beginners 'draw[ing] relatively low-level conclusions about the structuring of the language' and more advanced learners 'mak[ing] more subtle high-level inferences' on the basis of one and the same concordance (Johns 1986: 159). Another aspect that should be taken into account when choosing a way of approaching DDL is the topic to be investigated. Many authors claim that DDL is most effective 'on the "collocational border" between syntax and lexis' (Johns 2002: 109; see also Levy 1990 or O'Sullivan & Chambers 2006). However, it works with other topics too (for example, some aspects of discourse or grammar), although, here again, some adaptation may be required.

A final note is that, despite the attraction of some methods (as noted above, the concordance is particularly popular among DDL specialists) and some topics (Kennedy & Miceli [2001: 83], for example, refer to the 'lure of prepositions'), a key word in DDL is variety. Not only does variety make it possible to prevent tediousness among learners (a problem often highlighted in the literature, cf. Chambers 2007: 12), but it also caters to learners' different preferences and learning styles. Similarly, since it is precisely one of the goals of DDL to develop a more autonomous learning style, the teacher should avoid conformity as far as possible and agree to let the students approach corpora in the way they feel most comfortable with (Hunston 2002: 193).

4. ASSESSING THE EFFECTIVENESS OF DDL

One important question to ask regarding DDL is whether it works and actually facilitates language learning. It must be admitted that at this stage, very little is known about the effectiveness of DDL, and it is a recurrent theme in the DDL literature that more empirical studies are needed to validate this approach (see, among many others, Bernardini 2001: 247, Hadley 2002: 120 or Mukherjee 2006: 21). Often, the claims about the effectiveness of DDL are more of an act of faith, sometimes relying on subjective observation or informal testing, but usually engaging in pure speculation. Three types of evaluation can be carried out (Boulton 2008b: 41): evaluation of the attitudes (what do participants think about DDL?), practices (how well are the learners doing with DDL?) and efficiency (can learners gain benefit from DDL?). While attitudes and practices are important, the criterion that, ultimately, should be decisive in determining whether DDL is worth doing or not is efficiency. Learners may enjoy DDL and be good at it, but if they do not learn anything from it, there is no point in adding this to a curriculum that is already overloaded. Despite the crucial importance of this criterion, Boulton (2008b: 42) observes that very few studies seek to quantitatively assess the

efficiency of DDL, and that in most of the studies that do, no control group is involved, which seriously calls into question the validity of these studies.

As regards attitudes, a survey of the literature reveals extremely mixed results. While Ilse (1991: 107) points out that learners found the approach ‘fascinating’, Whistle (1999: 77) notes that it was ‘fairly unpopular with a majority of students’. Some authors report both positive and negative attitudes among the same learners. Kennedy & Miceli’s (2001: 80) students, for example, found DDL helpful and confidence-boosting, but sometimes also discouraging, time-consuming and frustrating. The same mixed results appear when considering learners’ capacities to do DDL (cf. Hadley 2002, Kennedy & Miceli 2001 and Sripicharn, this volume). As for efficiency, which we consider crucial for the future of DDL, most studies report some gain from DDL, though usually not very substantial (see e.g. Cobb 1997). As Boulton (2009b) rightly points out, however, DDL may have little impact on the knowledge of the question at hand, but have longer-term effects on the development of more general skills.

While these results may not be as outstanding as one may have wished or hoped for on the basis of some of the claims found in the literature, there are encouraging signs that DDL may deserve a place in the language classroom. Yet, one must recognize along with Barbieri & Eckhardt (2007: 320), and several other scholars, that ‘[d]espite the wealth of existing publications on classroom concordancing (...), the impact of concordancing and DDL in LT [language teaching] has been relatively inconspicuous’. This is because the implementation of DDL in the classroom poses a number of practical problems, which may discourage teachers from adopting this methodology, despite its numerous theoretical advantages. The major problems and limitations of DDL are examined in the next section.

5. THE PROBLEMS AND LIMITATIONS OF DDL

We will consider four aspects of DDL that may be problematic: the logistics, the teacher’s point of view, the learner’s point of view and the content of DDL. Logistics is often cited as one of the biggest problems of DDL. If learners are to actually use corpora in the classroom, they need computers (ideally one per student, but at least one for every two or three students), but also corpora and text retrieval software. All this costs a lot of money, which schools and universities are not always able to afford – or are not ready to invest without a clear guarantee that this will be profitable (cf. Hadley 2002: 110). It is true that some resources are freely available: certain concordancers and corpora, or data that can be compiled into a corpus (data from the web, essays produced by one’s students, etc). However, it should be borne in mind that these resources may have more limited functionalities than expensive tools and that for some data, copyright clearance might be necessary. The soft version of DDL requires slightly less technological equipment. Sometimes it is possible to get hold of ready-made DDL worksheets, but such resources are very rare, especially on the publishing market (Thurstun & Candlin 1997 and Tribble & Jones 1997 are two exceptions). Most of the time, therefore, the teacher will have to create his/her own materials, which still implies access to a computer (but just one), a corpus and a concordancer. Creating one’s own materials, however, takes time. Time, precisely, is another obstacle to the implementation of DDL. Not only is it time-consuming for the teacher to prepare the teaching materials, but it is also time-consuming to train students to deal with corpora and, perhaps more importantly, to complete a DDL task (Díez Bedmar [2006] describes an activity that took one hour and a half for just one word). And according to some, the results ‘might not repay the time taken’ (Willis, cited in Hunston 2002: 178).

If we consider the teacher's point of view, one reason for not doing DDL might simply be that the teacher does not know enough about corpora and the possibility of using corpora in the classroom. There would therefore be a need for 'in-service teacher training programmes' (Mukherjee 2006: 10). As Mauranen (2004: 100) points out, '[b]efore learners can be introduced to good corpus skills, their teachers need to possess them in the first place'. But it could also be that the teacher is familiar with corpora and DDL but prefers, for some reason, not to adopt such a methodology with his/her students. Logistic reasons may account for this choice (see above), or scepticism about the efficiency of DDL (see Section 4), but as pointed out by Boulton (2009b), it might be that these practical objections are actually 'camouflage for more profound theoretical concerns about the nature of learning, and more especially of teachers' and learners' roles'. Although, as noted above, DDL activities may be situated along a continuum from teacher-led to learner-led, it is nonetheless true that teachers have a less central role in DDL than in traditional teaching. They tend to have relatively little control over what happens during the lesson, which may be 'incompatible with the "minimum risk" scenario which can be found in many teaching cultures' (Boulton 2009b). In addition, since the computer becomes the main source of knowledge, this may be experienced as a 'loss of expertise' by the teacher (Hunston 2002: 171). DDL, therefore, requires that teachers take risks, and agree to "let go" and let the student take pride of place in the classroom.

For learners too, DDL may sometimes appear rather off-putting. Working with corpora is not straightforward and necessitates quite some training to acquire the basic skills – what Mukherjee (2002: 179) calls 'corpus literacy' (see also Sripicharn, this volume on the importance of preparing learners for using language corpora). Whistle (1999: 77) reports the case of some students who 'failed to see anything' and 'failed to formulate any clear rules'. Students may have 'difficulty devising effective search strategies' (O'Sullivan & Chambers 2006: 60) because of faulty spelling, for example, or simply because of the complexity of the processes involved (see Sun [2003: 607-8] for a good example of ineffective search strategy), and they may draw wrong inferences on the basis of the evidence (cf. Sripicharn 2004). It must also be stressed that DDL (and, in particular, the inductive learning strategies that it often entails) may be suitable for certain learners only, depending on their learning style.

The final aspect that may explain some of the reluctance to apply DDL has to do with its content. We have already mentioned the problem of authentication (or lack thereof) of the corpus data. In addition, the output of the search query may contain too much or too little data, or no data at all – which, for learners, poses the question of the distinction between 'does not exist' and 'is not represented in the corpus' (cf. Kennedy & Miceli 2001: 86). It may include too much 'noise', i.e. irrelevant hits, or it may be too difficult for learners to understand because of insufficient knowledge of the target language (Koosha & Jafarpour 2006: 206). It could also be that the corpus shows more details than the student is expected to learn or contains language which the teacher does not want the students to imitate, for example non-standard forms, swear words or literary phrases (see also Aston 1999 and Aston & Kübler, this volume on the possible unreliability of translations in bilingual corpora). Another problem is that the DDL approach may not be effective for all aspects of language, with some questions more 'concordance-ready' than others (Johns 1988: 25). Someya (2000), for example, explains that errors in prepositions are more easily dealt with by means of concordances than errors in articles. More generally, Gabrielatos (2005: 21) warns against the dangers of 'corpus worship' and 'frequency worship', thus suggesting that corpora and DDL are no panacea and that other tools besides corpora and other factors besides frequency have to be considered for teaching to be successful.

All this does not mean that DDL should be abandoned altogether. It is a promising technique which, as noted earlier, brings learners into contact with (potentially) authentic language, motivates them by introducing an element of discovery, develops important

cognitive skills and, more generally, provides benefits which go well beyond the knowledge of the item under study. However, the above list of problems should encourage us to think more deeply about the best ways of implementing DDL and, in particular, about the importance of adapting DDL to the specific learning situation in which it takes place (learners' levels, subject of study, availability of equipment, etc). Equally important is the necessity of going through a phase of validation – as one would do for any new teaching method – to test the efficiency of DDL. As long as such validation has not been carried out, it might be better to start off with a modest introduction and to leave more radical incorporation of DDL into the curriculum for later, when its efficiency has been proved and when students have got used to it.

FURTHER READING

- Boulton, A. (2009) "Data-driven learning: reasonable fears and rational reassurance," *Indian Journal of Applied Linguistics* 35(1). (This paper deals with the main obstacles to the implementation of DDL and provides some possible solutions.)
- Cobb, T. (1997) "Is there any measurable learning from hands-on concordancing?," *System* 25(3): 301-15. (This is one of the few empirical studies that seek to test the efficiency of DDL.)
- Granger, S. and C. Tribble (1998) "Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning," in S. Granger (ed.) *Learner English on Computer*, London & New York: Longman, pp. 199-209. (This paper demonstrates the relevance of learner corpora to DDL.)
- Johns, T. (1988) "Whence and whither classroom concordancing?" in T. Bongaerts, P. de Haan, S. Lobbe and H. Wekker (eds) *Computer Applications in Language Learning*, Dordrecht: Foris, pp. 9-33. (This is a good overview of the main uses of concordancing in the classroom.)
- Sun, Y.-C. (2003) "Learning process, strategies and web-based concordancers: a case study," *British Journal of Educational Technology* 34(5): 601-13. (This paper gives a good example of how learners can go wrong in drawing inferences on the basis of corpus data.)

REFERENCES

- Aston, G. (1999) "Corpus use and learning to translate," *Textus* 12: 289-314.
- Aston, G. (2001) "Learning with corpora: an overview," in G. Aston (ed.) *Learning with Corpora*, Houston: Athelstan, pp. 7-45.
- Aston, G. (2002) "The learner as corpus designer," in B. Kettemann and G. Marko (eds) *Teaching and Learning by Doing Corpus Analysis*, Amsterdam & New York: Rodopi, pp. 9-25.
- Barbieri, F. and S.E.B. Eckhardt (2007) "Applying corpus-based findings to form-focused instruction: the case of reported speech," *Language Teaching Research* 11(3): 319-46.
- Bernardini, S. (2001) "'Spoilt for choice': a learner explores general language corpora," in G. Aston (ed.) *Learning with Corpora*, Houston: Athelstan, pp. 220-49.

- Bernardini, S. (2004) "Corpora in the classroom. An overview and some reflections on future developments," in J. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam & Philadelphia: Benjamins, pp. 15-36.
- Boulton, A. (2008a) "DDL: reaching the parts other teaching can't reach?" in A. Frankenburg-Garcia (ed.) *Proceedings of the 8th Teaching and Language Corpora Conference*, Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa, pp. 38-44.
- Boulton, A. (2008b) "Esprit de corpus: promouvoir l'exploitation de corpus en apprentissage des langues," *Texte et Corpus* 3: 37-46.
- Boulton, A. (2009a) "Testing the limits of data-driven learning: language proficiency and training," *ReCALL* 21(1): 37-54.
- Boulton, A. (2009b) "Data-driven learning: reasonable fears and rational reassurance," *Indian Journal of Applied Linguistics* 35(1).
- Braun, S. (2006) "ELISA: a pedagogically enriched corpus for language learning purposes," in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 25-47.
- Breyer, Y. (2006) "My Concordancer: tailor-made software for language teachers and learners," in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 157-76.
- Chambers, A. (2005) "Integrating corpus consultation in language studies," *Language Learning & Technology* 9(2): 111-25.
- Chambers, A. (2007) "Popularising corpus consultation by language learners and teachers," in E. Hidalgo, L. Quereda and J. Santana (eds) *Corpora in the Foreign Language Classroom*, Amsterdam & New York: Rodopi, pp. 3-16.
- 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: 289-302.
- Cheng, W. (2007) "Concgramming: a corpus-driven approach to learning the phraseology of discipline-specific texts," *CORELL: Computer Resources for Language Learning* 1: 22-35.
- Cobb, T. (1997) "Is there any measurable learning from hands-on concordancing?" *System* 25(3): 301-15.
- Collins, H. (2000) "Materials design and language corpora: a report in the context of distance education," in L. Burnard and T. McEnery (eds) *Rethinking Language Pedagogy from a Corpus Perspective*, Frankfurt: Peter Lang, pp. 51-63.
- Díez Bedmar, M.B. (2006) "Making friends with DDL: helping students enrich their vocabulary," *Humanising Language Teaching* 8(3). Available at <http://www.hltmag.co.uk/may06/mart02.htm>.
- Flowerdew, J. (1993) "Concordancing as a tool in course design," *System* 21(2): 231-43.
- Gabel, S. (2001) "Over-indulgence and under-representation in interlanguage: reflections on the utilization of concordancers in self-directed foreign language learning," *Computer Assisted Language Learning* 14(3-4): 269-88.
- Gabrielatos, C. (2005) "Corpora and language teaching: just a fling or wedding bells?" *TESL-EJ* 8(4): 1-35. Available at <http://www-writing.berkeley.edu/TESL-EJ/ej32/a1.html>.
- Gavioli, L. (2005) *Exploring Corpora for ESP Learning*, Amsterdam: Benjamins.
- Granger, S. and C. Tribble (1998) "Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning," in S. Granger (ed.) *Learner English on Computer*, London & New York: Longman, pp. 199-209.
- Granger, S. (Forthcoming) "From phraseology to pedagogy: challenges and prospects," to appear in T. Herbst, P. Uhrig and S. Schüller (eds) *Chunks in Corpus Linguistics and Cognitive Linguistics*.

- Hadley, G. (2002) "An introduction to data-driven learning," *RELC Journal* 33(2): 99-124.
- Higgins, J. (1991) "Which concordancer? A comparative review of MS-DOS software," *System* 19(1-2): 91-100.
- Hunston, S. (2002) *Corpora in Applied Linguistics*, Cambridge: Cambridge University Press.
- Ilse, W.-R. (1991) "Concordancing in vocational training," in T. Johns and P. King (eds) *Classroom Concordancing*, special issue of *ELR Journal* 4: 103-13.
- Johns, T. (1986) "Micro-Concord: a language learner's research tool," *System* 14(2): 151-62.
- Johns, T. (1988) "Whence and whither classroom concordancing?" in T. Bongaerts, P. de Haan, S. Lobbe and H. Wekker (eds) *Computer Applications in Language Learning*, Dordrecht: Foris, pp. 9-33.
- Johns, T. (1997) "Contexts: the background, development and trialling of a concordance-based CALL program," in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds) *Teaching and Language Corpora*, London & New York: Longman, pp. 100-15.
- Johns, T. (2002) "Data-driven learning: the perpetual challenge," in B. Kettemann and G. Marko (eds) *Teaching and Learning by Doing Corpus Analysis*, Amsterdam & New York: Rodopi, pp. 107-17.
- Johns, T. and P. King (eds) (1991) *Classroom Concordancing*, special issue of *ELR Journal* 4.
- Kennedy, C. and T. Miceli (2001) "An evaluation of intermediate students' approaches to corpus investigation," *Language Learning & Technology* 5(3): 77-90.
- King, P. and D. Woolls (1996) "Creating and using a multilingual parallel concordancer," *Translation and Meaning* 4: 459-66.
- Koosha, M. and A.A. Jafarpour (2006) "Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL adult learners," *Asian EFL Journal* 8(4): 192-209.
- Kuo, C.-H., D. Wible, C.-C. Wang and F. Chien (2001) "The design of a lexical difficulty filter for language learning on the internet," in *Proceedings of the IEEE International Conference on Advanced Learning Techniques (ICALT'01)*, Madison, WI, 6-8 August 2001, pp. 53-4. Available at <http://www2.computer.org/portal/web/csdl/abs/proceedings/icalt/2001/1013/00/10130053abs.htm>.
- Levy, M. (1990) "Concordances and their integration into a word-processing environment for language learners," *System* 18(2): 177-88.
- Mair, C. (2002) "Empowering non-native speakers: the hidden surplus value of corpora in continental English departments," in B. Kettemann and G. Marko (eds) *Teaching and Learning by Doing Corpus Analysis*, Amsterdam & New York: Rodopi, pp. 119-30.
- Mauranen, A. (2004) "Spoken corpus for an ordinary learner," in J. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam & Philadelphia: Benjamins, pp. 89-105.
- Mukherjee, J. (2002) *Korpuslinguistik und Englischunterricht: Eine Einführung*, Frankfurt am Main: Peter Lang.
- Mukherjee, J. (2006) "Corpus linguistics and language pedagogy: the state of the art – and beyond," in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 5-24.
- Nesselhauf, N. (2004) "Learner corpora and their potential for language teaching," in J. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam & Philadelphia: Benjamins, pp. 125-52.
- O'Sullivan, Í. (2007) "Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy," *ReCALL* 19(3): 269-86.
- O'Sullivan, Í. and A. Chambers (2006) "Learners' writing skills in French: corpus consultation and learner evaluation," *Journal of Second Language Writing* 15: 49-68.
- Scott, M. (2004) *Wordsmith Tools version 4*. Oxford: Oxford University Press.

- Seidlhofer, B. (2000) "Operationalizing intertextuality: using learner corpora for learning," in L. Burnard and T. McEnery (eds) *Rethinking Language Pedagogy from a Corpus Perspective*, Frankfurt: Peter Lang, pp. 207-23.
- Seidlhofer, B. (2002) "Pedagogy and local learner corpora: working with learning-driven data," in S. Granger, J. Hung and S. Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, Amsterdam & Philadelphia: Benjamins, pp. 213-34.
- Someya, Y. (2000) "Online business letter corpus KWIC concordancer and an experiment in data-driven learning/writing," paper presented at the 3rd *Association for Business Communication International Conference*, Kyoto, 9 August 2000.
- Sripicharn, P. (2004) "Examining native speakers' and learners' investigation of the same concordance data and its implications for classroom concordancing with ELF learners," in G. Aston, S. Bernardini and D. Stewart (eds) *Corpora and Language Learners*, Amsterdam & Philadelphia: Benjamins, pp. 233-45.
- Stevens, V. (1995) "Concordancing with language learners: why? when? what?" *CAELL Journal* 6(2): 2-10. Available at <http://www.eisu2.bham.ac.uk/johnstf/stevens.htm>.
- Sun, Y.-C. (2003) "Learning process, strategies and web-based concordancers: a case study," *British Journal of Educational Technology* 34(5): 601-13.
- Thurstun, J. and C. Candlin (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," paper presented at the first international conference *Practical Applications in Language Corpora*, University of Lodz, Poland. Available at <http://www.tribble.co.uk/text/Palc.htm>.
- Tribble, C. and G. Jones (1997) *Concordances in the Classroom: A Resource Book for Teachers* (second edition), Houston: Athelstan.
- Whistle, J. (1999) "Concordancing with students using an 'off-the-Web' corpus," *ReCall* 11(2): 74-80.
- Widdowson, H.G. (2000) "On the limitations of linguistics applied," *Applied Linguistics* 21(1): 3-25.
- Widdowson, H.G. (2003) *Defining Issues in English Language Teaching*, Oxford: Oxford University Press.
- Willis, D. (2003) *Rules, Patterns and Words. Grammar and Lexis in English Language Teaching*, Cambridge: Cambridge University Press.