Testing the limits of data-driven learning: language proficiency and training

ALEX BOULTON

*CRAPEL – ATILF/CNRS, Nancy-Université, 3 place Godefroi de Bouillon,
BP 3397, 54015 Nancy – cedex, France
(email: alex.boulton@univ-nancy2.fr)*

Abstract

The potential for corpora in language learning has attracted a significant amount of attention in recent years, including in the form of data-driven learning (DDL). Careful not to appear to over-promote the field, enthusiasts have urged caution in its application, in particular with regard to lower-level learners, and have argued that extensive learner-training in corpus techniques is an essential condition for DDL to be successful. Such limits seem eminently reasonable, but there is a notable dearth of empirical studies to support them. This paper describes a simple experiment to see how lower-level learners cope with corpus data with no prior training.

The language focus here is on linking adverbials in English, which are renowned to be difficult to teach using traditional methods. The subjects are 132 first-year students at an engineering college in France of roughly intermediate and lower levels of English. They were divided into random groups to compare their ability to deal with the target items using traditional sources (extracts from a bilingual dictionary or a grammar/usage manual) or corpus data (short contexts or truncated concordances). Performance was tested prior to the experiment, subsequently to check ability to use the different information sources as a reference, and later to test recall.

No evidence was found that traditional sources promote better recall, and corpus data seemed to be more effective for reference purposes. While the results of any single experiment must be treated with caution, these findings suggest the need for more empirical studies to complement the theoretical arguments and qualitative data which currently dominate the discussions of DDL.

Keywords: Data-driven learning, corpus, training, level, empirical evidence, linking adverbials

1 Background

Teaching and learning are not symmetrical activities: without learners, the teacher is redundant, but learning may occur without a teacher. It may even be, in some cases, that learning is more effective without a teacher, i.e. when learners discover things for themselves. This is the basic premise of data-driven learning (DDL), where learners examine naturally-occurring language and discover patterns on their own. DDL is alleged to have many advantages – to foster learner autonomy, increase language awareness, improve ability to deal with authentic language, and so on.

Unfortunately, empirical evidence to support such claims is rare indeed (Chambers, 2007), which makes it difficult to examine claims and counter-claims alike (Cresswell, 2007: 269). Proponents of DDL have gone to some pains to avoid the extreme position of claiming that it is appropriate for all situations, for all learners, and for all language items; rather, they see it either as a complement to the arsenal of techniques and strategies at the teacher's and learner's disposal, or as a synonym for some kind of task-based or "discovery" learning. In particular, it is widely held that DDL is only appropriate for sophisticated, advanced learners, and that extensive training is essential. This paper sets out to test these two assumptions as part of a series of experiments with the same population (see also Boulton, 2008; Boulton, 2007b).

2 Empirical research

Chambers (2007) provides a detailed survey of twelve papers claiming to evaluate the efficiency of DDL, and finds them for the most part small-scale and essentially qualitative in nature. In a survey of the methodologies used in 39 empirical studies of DDL, Boulton (2007a) adds that many are primarily concerned with attitudes towards the approach, or examine the processes involved, with limited attention to language learning as such. These are of course important areas, but even if learners like DDL and are capable of using some of the techniques, this is not evidence that it will lead to efficient learning. Similarly, among the quantitative studies which do look at language use, the prime focus is not necessarily learning itself. Gaskell and Cobb (2004) and Todd (2001), for example, are mainly interested in the use of corpora as a reference source – specifically, whether learners can use concordances to correct errors in their written productions. While learning may happen here, there is no attempt to test this. In all, Boulton (2007a) found only five studies which aim to empirically evaluate language learning using DDL: Allan (2006), Cobb (1999a), Horst and Cobb (2001), Koosha and Jafarpour (2006) and Sun and Wang (2003).

The results of these studies are definitely promising, but mitigated in most cases, hence the frequent observation that data-driven learning should not be seen as a panacea. The limits remain to be tested in order to answer questions such as the type of learners it is appropriate for, and how it can best be implemented. At the moment, the consensus seems to be that DDL is most appropriate for advanced, sophisticated learners in a hands-on approach with significant training. These default assumptions are examined in more detail in the following sections.

2.1 Level

DDL was largely initiated by Johns, who originally had in mind "a particular type of student (adult: well motivated: a sophisticated learner with experience of research methods in his subject area)" (1986: 161). He later reported that fellow teachers often objected that DDL "may be all very well for students as intelligent, sophisticated, and well-motivated as ours at Birmingham University, [but] it would not work with students as unintelligent, unsophisticated and poorly-motivated as theirs" (Johns, 1991: 12). But he goes on: "what I suspect, however, is that most students given the opportunity to show what they are capable of might be (almost) as remarkable."

Work by Sealey and Thompson (e.g. 2007) shows how even primary school children can exploit corpora in their mother tongue, suggesting that, in the right conditions, no great level of sophistication is necessary. But when it comes to learning a foreign or second language, very little DDL research ventures outside the university environment (Braun, 2007). Virtually all of the 39 empirical studies of DDL surveyed in Boulton (2007a) focus on university students, the two exceptions being Ciezielska-Ciupek (2001) and Sun and Wang (2003), conducted in secondary schools in Poland and Thailand respectively.

Similarly, it is also often assumed that DDL will not work with learners at lower levels – even to the extent that the objection seems to apply to level of sophistication and language proficiency simultaneously, as if the two were inextricably connected. Again, the objection seems to be based on gut reaction rather than on empirical evidence: Hadley (2002), for example, was assured by colleagues in Japan that any attempt to use DDL with beginners was "doomed to failure", even though these same colleagues had never tried. Boulton's (2007a) survey found only a handful of studies which claim to evaluate DDL at lower levels. As at other levels, the results are mixed, although in no case is it found to be completely useless. In one instance reported by Yoon and Hirvela (2004), DDL was apparently more successful with intermediate than advanced learners, although the experimental design may be partly responsible for this (O'Sullivan & Chambers, 2006: 61).

One commonly accepted objection is that learners at lower levels might simply not have sufficient analytical and linguistic skills to cope with the complexity and fuzziness of authentic data of a foreign or second language – "authentic" in Holec's (1990) sense that they were not originally created for purposes of language teaching or learning. On the other hand, learners might benefit from exposure to that complexity at an early stage rather than living in the false expectation that clear and simple rules can always be devised. Given the strong arguments that learners can and should be confronted with authentic documents from the beginning (e.g. Holec, 1990), the case against using corpora with lower levels is severely weakened in the absence of evidence to the contrary.

One might also wonder why a new approach should be reserved for learners who are already advanced, those who are already successful using existing methods. Perhaps DDL might be appropriate for "less successful" learners as an alternative to traditional methods which, by definition, have not worked in their case. But in the face of failure, traditional methods tend to inflict more of the same, perhaps on the assumption that poor learners will be poor learners whatever the method, so there is little point in attempting anything new. Responsibility and initiative (in DDL or otherwise) are thus seen almost as rewards for learners who have proved they can cope without; poor learners are not to be trusted with them.

Traditional teaching, especially at lower levels, tries to take as much of the burden off the learners as possible, partly by pre-formulating the rules to learn. One problem though is that such explicit rule-learning is an 'artificial' and 'difficult' process demanding considerable intellectual rigour. DDL, on the other hand, exploits processes that humans have evolved to be naturally good at: exposure to information, detection of patterns, extrapolation to other cases (e.g. Scott & Tribble, 2006: 6; Gaskell & Cobb, 2004). It might be argued that this is particularly appropriate

at lower levels when learners already have considerable cognitive demands made upon them.

One major aim of this paper is, then, to see to what extent lower level learners can benefit from aspects of DDL.

2.2 Training

It has frequently been claimed that lack of sufficient training is a major barrier to the implementation of DDL, whether for teachers or for learners, or indeed both (e.g. Breyer, 2006). Lack of training has on occasion been cited as a factor where empirical research has not produced the hoped-for results, but the training itself has generally been an incidental element rather than a main research question in such studies. More often still, it is simply taken as given. We are perhaps beginning to see something of a retreat on this strong insistence on training. For Bernardini (2001: 243), "the difficulties should not be overestimated; learners should quickly acquire the skills needed"; and for Sinclair (2004: 297), "any teacher or student can readily enter the world of the corpus and make the language useful in learning." Of course, the more experience and training learners have, the more responsibility and autonomy they can assume, and the more sophisticated the use of corpora can be, but "both teacher and student can make use of a corpus right away, with only a modest few hours of orientation" (Sinclair, 2004: 288). A number of studies, however (e.g. Boulton & Wilhelm, 2006), have shown that learners can prove quite sophisticated, even with complicated tasks such as building and analysing their own corpora, with comparatively little training.

Insistence on training is particularly associated with hands-on corpus work, but may be less important when working with prepared materials. However, it could be argued that students who are used to a deductive approach will require "extensive guidance" in using inductive learning strategies needed for interpreting concordances (Sun, 2003: 612). But the same could be said for dictionaries, for example: "The dictionary is an excellent tool in the hands of a skilled learner. An unskilled user wastes time and comes away frustrated from dictionary consultations" (Roby, 2005: 59). Indeed, Frankenberg-Garcia (2005) found that learners were just as good with minimal training on concordances as they were with dictionaries they had been using all their lives. With corpora as with dictionaries, it seems obvious that learners cannot be expected to gain maximum advantage by merely being given access to the tools. Yet for either resource, that is not to say that learners cannot benefit from at least some aspects with little or perhaps even no training.

One potential problem lies in the "sometimes startling physical appearance of concordances" (Lamy & Klarskov Mortensen, 2007), with the key word in the middle of truncated texts for "vertical" reading; this may be confusing for some, hence the need for a book such as Sinclair's *Reading Concordances* (2003). Specifically regarding L2 learners, Johns (1986: 157) first cited the appearance of "unfinished sentences" as a common learner complaint over twenty years ago. Gavioli (2005: 29) claims that the concordance "is not a type of text whose reading and digesting can be taken for granted... The type of (linguistic) information a concordance gives to the student-analyst is not obvious." All three major problems cited in Koosha and Jafarpour

(2006: 206) relate to the difficulty of interpreting concordances. Yoon and Hirvela (2004: 270) are among the few to quantify this, with 62% of their intermediate learners citing the ''cut-off sentences in concordance output'' as a difficulty. Otherwise the evidence remains largely anecdotal: Johns' (1986: 157) experience is that ''students fairly soon overcome this first aversion'', while for Lamy and Klarskov Mortensen (2007), ''our experience is that you cannot overestimate the students' need for familiarity with the appearance of concordances, and their need for guidance as to how to derive conclusions from lists of citations.''

The second main aim of the present study is thus to see whether learners can derive useful information from concordances without training.

3 Method

This study aims to address the two main issues mentioned above: whether lower level learners may be able to derive some benefit from a DDL approach in the form of a concordance print-out, and whether this may occur in the absence of training. In brief: university students requiring English for specific purposes were tested on a specific language point, that of linking adverbials. For each of the target items, different groups were provided with either corpus data (KWIC concordances or short contexts) or traditional pedagogical information (bilingual dictionary entries or grammar/usage notes). The tests were a simple multiple-choice gap-fill of concordance and sentence-length questions. A first test was conducted prior to the experiment as a control of existing knowledge and ability. A second test was conducted with the information to hand in order to see how learners fared using it for reference purposes. Recall of the different information types was tested at a later date using the same test format. This experiment design allowed comparison of the results between the three test sessions for the four different information types, as well as by level as measured by an in-house placement test.

3.1 Population

The participants were first-year students enrolled at an engineering college in the north east of France. The vast majority were male (86%) and native speakers of French (94%), average age 18.5. They had been studying English for an average of 6.6 years at school before coming to the college, though few could be qualified as advanced, as can be seen from the results of an in-house placement test organised for streaming purposes. Based on a full-length TOEIC of 100 listening and 100 grammar/ reading questions, their average score was 51.29%, which corresponds to about 450 on the TOEIC scale, towards the lower end of their ''intermediate'' band (405–600).[1] Motivation levels for English are generally quite low in this highly traditional environment (see Brown, 2007): the classes are compulsory, and seen by most as something to be endured rather than as preparation in a useful vocational skill. Discounting data from repeat students and those who missed any of the test sessions left 132 participants.

---

[1] Test of English for International Communication. Can-do levels table.

3.2 *Linking adverbials*

The language point chosen for the present study was linking adverbials, described as follows by Biber, Johansson, Leech, Conrad, and Finegan (1999: 875) in the *Longman Grammar of Written and Spoken English*: "the primary function of linking adverbials is to state the speaker/writer's perception of the relationship between two units of discourse". Such a general definition would appear appropriate as it derives from a corpus-based pedagogical description of English, and has the advantage of avoiding terminological disputes: the same items are referred to under many names, including linkers, connectors, conjuncts, conjunctions, cohesive markers, connecting adverbials, and so on.

A number of corpus studies (e.g. Conrad, 1999; Aijmer, 2003) have been devoted to linking adverbials as their use is markedly complex. This also suggests they are likely to be difficult to teach, support for which can be found in several studies of learner corpora showing that they continue to pose problems even at advanced levels, as measured by overuse, underuse, and misuse (Aarts & Granger, 1998; Altenberg & Tapper, 1998; Cheng & Warren, 2000; Crewe, 1990; de Haan & Van Esch, 2007; L. Flowerdew, 1998; Granger & Rayson, 1998; Granger & Tribble, 1998; Granger & Tyson, 1996).

Probably in part because they are difficult to describe, linking adverbials seem to be one of the areas "generally ignored, neglected or misrepresented in standard works of reference and course materials", and thus most suited for DDL (Johns, 1997). J. Flowerdew (2001) finds differences between corpus data and the way linking adverbials are traditionally taught, and Garton (1996: 8) uses them as an illustration of a "gap between simplified textbook models and authentic native-speaker usage". For successful mastery of such items, learners would seem to need something more than what can currently be found in standard materials.

It may be possible to learn by analogy from data even without being able to put that knowledge into words in the form of clear and simple rules; or as Widdowson (1998: 715) puts it, "a lot of time is wasted in trying to teach things that can only be learned by experience". In this way, the learners' findings "may show a far greater degree of abstraction and subtlety" (Johns, 1991: 3) than that found in attempts at formal descriptions. For all of these reasons, linking adverbials would seem to be an appropriate language area for investigation in DDL.

Linking adverbials as a whole, according to Biber *et al.* (1999: 880), are "considerably more common in conversation and academic prose than in fiction and news". As the experiment described here was based on newspapers for reasons outlined below, the focus was on the most common type of linking adverbials in this register, namely contrast and concession (Biber *et al.*, 1999: 882). Table 1 shows the ten items selected for the experiment, along with their frequency per million words in the news and spoken registers compared to the entire 100 million words of the British National Corpus.[2]

For this experiment, these items represent a suitably wide variety of linking adverbials of contrast. They have varying rates of frequency overall, with the least

---

[2]   Using Mark Davies' freely available on-line interface: http://corpus.byu.edu/bnc/

Table 1 *Frequency per million words of linking adverbials in the British National Corpus: in the news and spoken registers, and in the entire 100 million word corpus*

linking adverbial	news	spoken	overall
but	5167.97	6384.16	4452.45
however	387.85	89.31	597.30
actually	88.46	1236.48	254.54
in fact	78.49	289.70	162.63
anyway	35.53	504.41	116.52
whereas	18.52	63.76	61.69
on the other hand	19.36	23.90	53.11
besides	11.37	7.35	24.62
nonetheless	6.30	4.06	12.96
on the contrary	2.44	1.26	7.97

frequent more typical of written registers. Even these were likely to be familiar to the participants in this study, as all feature explicitly in the relevant section[3] of the usage manual available to our learners (Swan, 2005: 138–157); however, none were in the list of "130 common mistakes" (p. xxvi–xxix). Some of them are apparently quite simple (e.g. *but*), others more complex (e.g. *besides*). Some are frequently classed as "*faux amis*" (e.g. *actually* vs. *actuellement*), and others are deceptively transparent: *on the contrary* is formally similar to French *au contraire*, but its use is sufficiently complex for Lake (2004: 138) to advise learners to avoid it altogether. All of these items have posed problems for our learners in the past.

3.3 Authentic language

The language items for the tests and the corpus information sheets were gathered on a single day from WebCorp (Renouf, Kehoe & Banerjee, 2007). The main disadvantages are slow download time (a minor drawback for only ten items), and lack of stability as the corpus changes from day to day. This is true of any internet "corpus", and although it might be desirable to know the size of the data-base and the relative frequency of the items, these were not essential for this experiment. On the other hand, WebCorp has a number of relevant advantages: in particular, it is freely available on the internet, and allows domain-specific searches of up-to-date sources sufficient for current purposes.

The chosen options were for Google, plaintext output and thirty words to left and right. This last criterion allowed for easier downloading, and meant that the same data could be used for different question formats. Although linking adverbials are not particularly frequent in newspapers (Conrad, 1999), the search was limited to this genre

---

[3]   The section is titled "discourse markers", although we avoid this term to avoid confusion with features typical of spoken language. For Schiffrin (1987: 31), for example, discourse markers are "sequentially dependent elements which bracket units of talk" – "talk" being the key word, as can be seen through her choice of items: *oh*, *well*, *and*, *but*, *or*, *so*, *because*, *now*, *then*, *y'know*, *I mean*.

Table 2 *Frequency of target items in corpus*

item	occurrences retrieved
a) but	975
b) actually	581
c) however	375
d) anyway	366
e) whereas	311
f) in fact	223
g) besides	161
h) nonetheless	143
i) on the other hand	111
j) on the contrary	69

in English, partly as newspapers are familiar to most students, partly to gain up-to-date samples, and partly to ensure a certain minimum quality. Specifically, the search was limited to ''British broadsheets'', as students in France are less likely to be familiar with American newspapers and tabloids. The total results for each are given in Table 2.

To keep the language authentic it was not altered or edited in any way: each concordance line retained was kept exactly as it was (except for formatting for pagination purposes). However, the 30 concordance lines needed were selected from a random pool of 50, firstly to eliminate doublets and non-cohesive text – an essential stage with the typically messy web-as-corpus (Kilgarriff, 2001). Of the remainder, the most apparently transparent or useful for the experiment were retained – the ''most illustrative'', in Stevens' words (1991: 51). Such selection of concordances is common practice – Kennedy and Miceli, for example, talk of ''quality control'' in their small samples (2002: 187) – and can be justified on the grounds that a concordance consists of *samples* not *examples* (Gavioli, 2005: 7): teachers routinely select examples of all kinds. There is no reason to believe it should affect the basic research questions, viz whether lower level learners with no previous corpus training can make sense of a concordance and draw useful information from it.

3.4 Experimental condition

For the experimental condition itself, the students were randomly assigned to one of four groups (Table 3) to test their ability to interpret, apply and recall different information types.

The information sheets, each about three pages long, were distributed after the first test, and collected at the end of the session:

a) The short context (SC) sheets consisted of five short contexts for each test item gathered using WebCorp, grouped together and with the test item highlighted for ease of reference. Where the search had returned more than one full sentence within the specified span (thirty words either side), all full orthographic sentences were accepted, giving an average of 2.4 full sentences for each sample, or nearly 40 words.

Table 3 *Group distribution*

no. of students	group type
34	SC: short context data
34	KW: key word in context data
32	BD: bilingual dictionary data
32	GU: grammar/usage data

b) The keyword in context (KW) sheets featured eight concordance lines for each test item gathered using WebCorp, grouped together and with the test item highlighted for ease of reference. The concordance format provided an average of 46 characters either side, or just over eight whole words.

c) The bilingual dictionary (BD) entries for each test item were taken from the *Collins-Robert Senior*, a fairly large desk dictionary of nearly 2,000 pages of entries. Like most bilingual and unlike most monolingual dictionaries, it is not substantially corpus-based, but is popular in France, familiar to many learners, and large enough to provide sufficient information for our purposes. The information was presented in alphabetical order for the entries in exactly the format of the dictionary.

d) The grammar/usage (GU) notes for each test item came from Swan's *Practical English Usage* (2005), which uses "simple everyday language" (p. ix). Although not explicitly corpus-based, the recent edition used claims to be "thoroughly checked against large electronic databases ('corpora') of authentic spoken and written English" (p. ix). The relevant entries from the section on "discourse markers" formed the basis, supplemented by other entries for the target items where they existed. The information was presented in exactly the format of the manual.

3.5 Test instrument

The participants completed three separate tests: (a) a pre-test; (b) a test where they could consult the information sheets; (c) a recall test conducted ten days later. All students had the same testing instrument in each test regardless of which group they were assigned to; the format was identical between the three testing sessions, only the actual language content was new each time. Only answers which corresponded to the original data were marked correct; other responses might have been judged acceptable in many of the full-sentence contexts, but this necessarily involves an element of subjectivity, and more importantly would bias the scoring as alternative responses would be less likely in the multiple concordances. For the purposes of the experiment, it was considered sufficient that all groups completed both question types and were subjected to the same scoring procedure.

Each test sheet included two exercises, each comprising ten questions corresponding to the ten target items. One set of questions presented ten sets of four concordances; the other presented ten short contexts, each of between one and three full sentences (Figure 1). In both cases, the target words were blanked, and

Example concordance question for BUT:

I can't sleep, partly because it is just too hot, _____ partly because I think the 'anniversary eff
es' sense of living lives that run on a parallel _____ separate track to that used by local peopl
   start trying to mount her. "Cows look calm, _____ really they are gay nymphomaniacs," he
   im". The timing could not have been worse, _____ the job offer couldn't have been better. It

Example sentence-length question for BUT:

Total US advertising spend this year is expected to be about $280bn. Spending on internet
advertising is expected to grow at 25% this year _____ will still only be a fraction of the total,
some $12bn. At Microsoft's annual meeting, chief executive Steve Ballmer insisted, "We will
catch up and we will surpass Google."

Fig. 1. Example test questions.

participants had to choose from the list of target items (*a* to *j*); a completed example
was provided for each exercise type. It seemed plausible that the concordance group
might have an unfair advantage if the entire test instrument was based on KWICs,
while the others might do better given only full sentences; the combination of the two
question types was designed to reduce bias in favour of one or the other.

4 Results and discussion

One of the most immediately striking aspects was that the scores were very low, an
average of only 13.92% per student over the three tests (Table 4). This is in large part
due to the number of blank responses – 30.26% overall – suggesting that linking
adverbials as a whole are indeed difficult for these learners.

   While unexpected, the low absolute scores are not in themselves a problem: the
important consideration is that they allow us to distinguish between the different
variables. The data can be analysed in three main ways:

   a)   test: comparing the changes between tests 1, 2 and 3, and between the
        concordance and sentence question types;
   b)   level: comparing the scores between three bands of proficiency in English;
   c)   group: comparing the scores between the groups according to the type of
        information they received during test 2 (SC, KW, GU, BD).

4.1 Results by test and question type

Figure 2 shows some of the clearest findings: the lowest scores were obtained on test 1;
there was significant improvement in test 2; and although the scores in test 3 were

Table 4 *Correct responses, all groups combined*

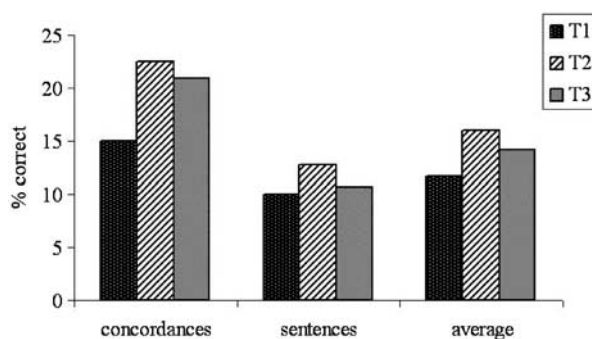| %	T1	T2	T3	concordances	sentences	total
ave	11.67	16.04	14.07	19.47	11.15	13.92

T1/T2/T3 = successive tests.

Fig. 2. Correct responses by test (T1/T2/T3) and question type
Concordance questions; full sentence questions; average of both question types.

lower than in test 2, they were still significantly higher than in test 1 ($p < 0.05$). Overall, more than half of all participants scored better in tests 2 (59.0%) and 3 (50.8%) than in test 1. This is perhaps not surprising, partly because test 2 was administered immediately after test 1 while test 3 took place ten days later, but more importantly because the information sheets were available for consultation during test 2 but not in test 3. It seems, then, that these learners can use the various types of information as a reference (test 2), as well as for recall at a later date (test 3).

As regards the test design, it is significant ($p < 0.01$) that students scored higher on the concordance questions than they did on the full sentence questions: 19.47% and 11.15% respectively over the three test sessions. This confirms Stevens' (1991) results, and the pattern holds however we look at the data, for each level and for each group, so the overall low scores cannot be attributed solely to the use of concordances as a test instrument. A more likely cause is the use of authentic language, as it was used in both test types, agreeing with the findings of a number of researchers (e.g. Koosha & Jafarpour, 2006; Cobb, 1999a).

4.2 Results by level

As it is often alleged that corpus data are only appropriate at higher levels, we also need to examine the effect of proficiency. The sample population was divided into three equal groups according to level determined by their placement test scores, the highest averaging 64.49% (at the lower end of TOEIC's "basic working proficiency"), the others 50.34% ("intermediate") and 40.27% ("elementary") respectively. Figure 3 shows that the highest level group fared best in the experiments on all counts, though it should be remembered that even they were only intermediate in level.

This basic pattern is to be expected, but the differences are fairly small: there was no significant difference between levels 2 and 3, and the difference with the highest level was only just significantly different at the $p < 0.05$ level. It is important too to note that all levels showed the same pattern of evolution over the three tests: lowest on test 1, highest on test 2, with a drop on test 3 (although the results were still significantly higher than in test 1; $p < 0.05$).
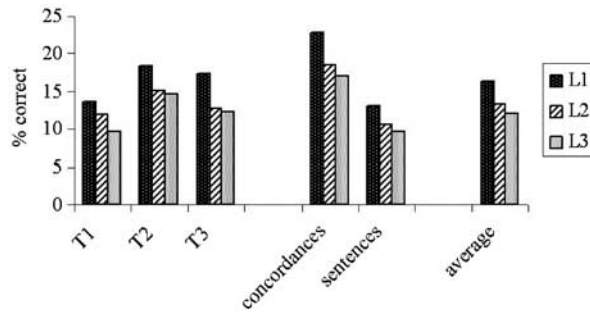
Fig. 3. Correct responses by level (L1=highest) over the three tests (T1/T2/T3)
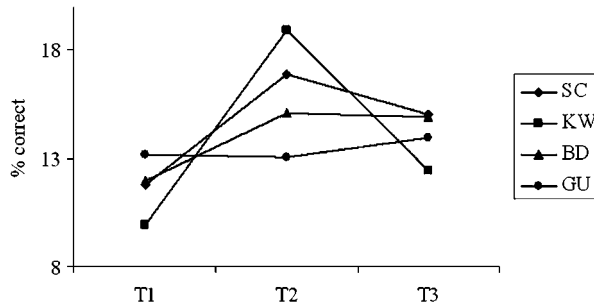Concordance questions; full sentence questions; average of both question types.

Fig. 4. Change in correct responses between tests (T1/T2/T3), by data type
SC = full-sentence group; KW = key-word-in-context group; BD = bilingual dictionary group;
GU = grammar/usage manual group.

4.3 Information types

The crucial stage is to examine the effect of the different types of information for tests 2 and 3. This can be seen clearly in Figure 4, which shows that the two groups that had corpus information (SC and KW) managed to use this information more effectively as a reference source in test 2 than the groups which had the traditional pedagogical information (BD and GU). When it comes to recall, however, in test 3, the differences are no longer significant between any of the four groups ($p > 0.05$): there is thus no indication that having had corpus data as a reference is better or worse than traditional sources for learning itself.

Table 5 provides the scores for each group in each test, along with the percentage rate of change between tests. On average, students scored 37% higher in test 2 than in test 1, with the corpus groups showing the biggest improvements: +43% for the SC group and +91% for the KW group. On the other hand, the group with the bilingual dictionary information (BD) showed a more modest improvement (+26%), while the group with the grammar/usage information (GU) actually decreased by –1%, although this last result is not significant ($p > 0.05$). Between tests 2 and 3, as we have already seen, there was a substantial drop: –12% on average over all groups. The biggest drop was in the two corpus data groups (SC and KW), which

Table 5 *Correct answers and percentage changes between tests (T1/T2/T3), by information type*

	T1	T2	T3	T1 > T2	T2 > T3	T1 > T3
SC	11.76	16.86	15.00	+43.33	−11.05	+27.50
KW	9.90	18.92	12.45	+91.09	−34.20	+25.74
BD	11.98	15.10	14.90	+26.09	−1.38	+24.35
GU	13.13	13.02	13.96	−0.79	+7.20	+6.35
ave	11.67	16.04	14.07	+37.45	−12.28	+20.56

SC = full-sentence group; KW = key-word-in-context group; BD = bilingual dictionary group; GU = grammar/usage manual group.

had previously recorded the biggest increases; the changes for the other groups were not significant. A more useful point of comparison for test 3 is against test 1. All groups show an improvement here, on average +21%. The GU group increased by only +6%; this insignificant difference may be accounted for by familiarity with the test design alone. The other groups increased by around +25%, a significant improvement ($p < 0.05$) although there was no significant difference between these three groups.

An alternative approach is to look at the number of participants in each group who improved their scores. The group where most scored better was the KW group: 79% of them scored higher in test 2 than in test 1; 65% also scored higher in test 3 than in test 1. In the GU group, however, only just over a third improved their scores.

4.4 Discussion

Overall, the results do suggest a number of conclusions. The students in this experiment as a whole were not at very advanced levels of language ability, and had had no prior training in using concordances or corpus data. However, the vast majority performed better on the concordance questions than on the full sentences. This result held over all levels and irrespective of the information type received prior to test 2, suggesting that there may have been some "test training" effect (i.e. the use of concordances in the first test helped the students with subsequent tests). Other factors may also have been at work; for example, the "novelty value" of the concordance questions may have induced students to spend more time on these. However, it seems unlikely that such explanations could account for a substantial part of the difference. Taken at face value, this finding directly contradicts the notion that concordances should only be used at higher levels of proficiency and after substantial training. Although the format is unfamiliar, it does have two immediate advantages. Firstly, the KWIC presentation draws attention to the target items, thus encouraging noticing (Hyland & Milton, 1997: 384). Secondly, more sets of evidence are provided (Stevens, 1991) at no extra "cost" – the number of words in a four-line concordance being roughly equivalent to that in one full-sentence context. These multiple contexts may be precisely what is necessary for the "broader perspective" of the target items (Levy, 1990: 184), i.e. to provide numerous paths to them in a more efficient way than would occur accidentally or with longer contexts. This is the reasoning adopted

by Thurstun and Candlin (1997) and which underlies their concordance-based textbook for academic writing (see also Thurstun & Candlin, 1998).

It might also be that the truncated lines actually contribute to success by reducing the information load – especially important perhaps for lower levels. They allow a focus on form and meaning in short, multiple contexts, showing various usages simultaneously and without the distraction of longer stretches of discourse (Cobb, 1999b). It does then seem that for some learners with some items and for some parts of the learning process, it may be that a more extended context is not necessarily desirable if it detracts from the target without contributing significantly greater comprehension (see also Wilson, 1997: 128).

The use of different types of information also provides some useful insights. Firstly, when using the information sheets in test 2 as a reference – or "informant" to use Johns' (1991) term – the two groups with corpus data fared significantly better than the groups with traditional reference information. The obvious implication is that corpus data can be useful as a reference source for untrained learners, especially in the form of KWICs – significantly more so than traditional pedagogical information. This supports a number of papers which study the use of corpora as a reference source for writing or error-correction (e.g. Gaskell & Cobb, 2004; Todd, 2001). On the other hand, the difference disappeared in the recall situation of test 3: corpus data were neither more nor less helpful than traditional pedagogical information in recall. One might imagine though that training on the unfamiliar corpus information would provide a significant advantage.

Of the two traditional information groups, it certainly seems that, for lower levels of language ability, grammar/usage notes are of very limited use: no statistical difference was detected in performance for this group over the three tests. Of the two corpus groups, the one with the KWIC data fared significantly better in test 2, and showed greater improvement overall. However, while statistically significant, it has to be pointed out that this group started out with a lower level in test 1, the pre-experimental task. This can only be described as an anomaly, as the students were assigned randomly to the four groups prior to this test, but any conclusions must therefore be tentative and in need of further research.

5 Conclusion

This study examined the ability of lower level learners to use authentic corpus data as a reference source and for learning. No training was provided in the use of corpus data. Used as a reference source in this study, corpus samples led to more successful results than traditional pedagogical resources of the type the learners were familiar with: a bilingual dictionary and a grammar/usage manual. For the purposes of recall, the corpus and pedagogical resources were found equally effective. Of the two types of corpus data, it seems that authentic contexts in the form of multiple KWIC concordances are more amenable to lower levels than longer contexts consisting of one or more full sentences.

In the light of these findings, it seems that data-driven learning could be appropriate for a wider range of learners than usually assumed. It may be most suited for advanced learners trained in using corpora, but it can bring something to the

learning process even at lower levels of ability. Subsequent informal feedback from the participants suggests that the main difficulty lay not in the DDL approach or in the KWIC presentation, but rather in the difficulty of using authentic language. Within a DDL approach, a number of potential solutions could be considered. It might be possible to simplify the corpus itself, for example using simplified readers (Allan, 2008); alternatively, it might be possible to grade the texts within the corpus automatically (Chujo, Utiyama & Nishigaki, 2007). At a later stage, the teacher might carefully select concordances for use with particular groups of students and for particular language points; Johns (1991: 4) admits to "a degree of 'rule-hiding' in the selection of citations, the categories adopted, and the sequencing of citations within each category". Some might even be tempted to edit the concordances, although Adolphs (2006: 108) points out that this would "change the nature of the data and run counter to the objective of exposure to naturally occurring language in use".

It remains, however, that even the authentic language used in this experiment led to encouraging results, and the lack of tried-and-tested means to simplify the learning process does not mean that lower level learners should be denied access to corpora. Similarly, training in the use of concordances would presumably lead to substantially greater benefit, but lack of opportunities for this does not mean DDL should be abandoned. As with dictionary use, explicit training would no doubt be of use to many learners, but the absence of such training does not mean the tool should be abandoned altogether. Cobb (1999a; 2003) provides further evidence for this.

After years of interest within the research community, DDL has yet to make significant inroads into mainstream teaching environments. As Cresswell (2007) points out, the lack of a substantial body of concrete evidence is no doubt a contributing factor. More empirical studies are needed to indicate different conditions for use of DDL – for what types of learners, what minimum resources, what language points, how it can be integrated with other techniques, and so on. The conclusions from a single experiment can of course only be tentative, but this paper shows that simple experiments can yield useful insights, and complement existing longitudinal or qualitative studies.

References

Aarts, J. and Granger, S. (1998) Tag sequences in learner corpora: a key to interlanguage grammar and discourse. In: Granger, S. (ed.), *Learner English on computer*. London: Longman, 132–142.

Adolphs, S. (2006) *Introducing electronic text analysis: a practical guide for language and literary studies*. London: Routledge.

Aijmer, K. (2003) Discourse particles in contrast: the case of 'in fact' and 'actually'. In: Wilson, A., Rayson, P. and McEnery, T. (eds.), *Corpus linguistics by the lune: a festschrift for Geoffrey Leech*. Frankfurt: Peter Lang, 23–35.

Allan, R. (2006) Data-driven learning and vocabulary: investigating the use of concordances with advanced learners of English. *Centre for language and communication studies, occasional paper*, 66. Dublin: Trinity College Dublin.

Allan, R. (2008) Can a graded reader corpus provide 'authentic' input? *ELT Journal* (advance access).

Altenberg, B. and Tapper, B. (1998) The use of adverbial connectors in advanced Swedish learners' written English. In: Granger, S. (ed.), *Learner English on computer*. London: Longman, 80–93.

Bernardini, S. (2001) 'Spoilt for choice.' A learner explores general language corpora. In: Aston, G. (ed.), *Learning with corpora*. Houston: Athelstan, 220–249.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Pearson.

Boulton, A. (2007a) But where's the proof? The need for empirical evidence for data-driven learning. *BAAL 40: technology, ideology and practice in applied linguistics*. University of Edinburgh, September.

Boulton, A. (2007b) DDL is in the details… and in the big themes. In: Rayson, P. (ed.), *Proceedings of 4th Corpus Linguistics conference*. Birmingham: University of Birmingham Centre for Corpus Research.

Boulton, A. (2008) Looking for empirical evidence of data-driven learning at lower levels. In: Lewandowska-Tomaszczyk, B. (ed.), *Corpus linguistics, computer tools, and applications – state of the art*. Frankfurt: Peter Lang, 581–598.

Boulton, A. and Wilhelm, S. (2006) Habeant corpus – they should have the body: tools learners have the right to use. *ASp*, **49–50**: 155–170.

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: Braun, S., Kohn, K. and Mukherjee, J. (eds.), *Corpus technology and language pedagogy: new resources, new tools, new methods*. (*English corpus linguistics, 3*). Frankfurt: Peter Lang, 157–176.

Brown, D. (2007) Language learner motivation and the role of choice in ESP listening engagement. *ASp*, **51–52**: 159–177.

Chambers, A. (2007) Popularising corpus consultation by language learners and teachers. In: Hidalgo, E., Quereda, L. and Santana, J. (eds.), *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 3–16.

Cheng, W. and Warren, M. (2000) The Hong Kong corpus of spoken English: language learning through language description. In: Burnard, L. and McEnery, T. (eds.), *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang, 133–144.

Chujo, K., Utiyama, M. and Nishigaki, C. (2007) Towards building a usable corpus collection for the ELT classroom. In: Hidalgo, E., Quereda, L. and Santana, J. (eds.), *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 47–69.

Ciezielska-Ciupek, M. (2001) Teaching with the internet and corpus materials: preparation of ELT materials using the internet and corpus resources. In: Lewandowska-Tomaszczyk, B. (ed.), *PALC 2001: practical applications in language corpora*. (*Lodz studies in language, 7*). Frankfurt: Peter Lang, 521–531.

Cobb, T. (1999a) Breadth and depth of lexical acquisition with hands-on concordancing. *CALL*, **12**(4): 345–360.

Cobb, T. (1999b) Giving learners something to do with concordance output. *IMELT 99*. Hong Kong Polytechnic University, November. http://elc.polyu.edu.hk/conference/papers/Cobb%20-%20Giving%20learners%20something%20to%20do%20with%20concordance%20output.htm

Cobb, T. (2003) Do corpus-based electronic dictionaries replace concordancers? In: Morrison, B., Green, G. and Motteram, G. (eds.), *Directions in CALL: experience, experiments, evaluation*. Hong Kong: Polytechnic University, 179–206.

Conrad, S. (1999) The importance of corpus-based research for language teachers. *System*, **27**(1): 1–18.

Cresswell, A. (2007) Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In: Hidalgo, E., Quereda, L. and Santana, J. (eds.), *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 267–287.

Crewe, W. J. (1990) The illogic of logical connectors. *ELT Journal*, **44**(4): 316–325.

De Haan, P. and Van Esch, K. (2007) Assessing the development of foreign language writing skills: syntactic and lexical features. In: Fitzpatrick, E. (ed.), *Corpus linguistics beyond the word: corpus research from phrase to discourse*. Amsterdam: Rodopi, 185–202.

Flowerdew, J. (2001) Concordancing as a tool in course design. In: Ghadessy, M., Henry, A. and Roseberry, R. (eds.), *Small corpus studies and ELT: theory and practice*. Amsterdam: John Benjamins, 71–92.

Flowerdew, L. (1998) Integrating expert and interlanguage computer corpora findings on causality: discoveries for teachers and students. *ESP journal*, **17**(4): 329–345.

Frankenberg-Garcia, A. (2005) Pedagogical uses of monolingual and parallel concordances. *ELT Journal*, **59**(3): 189–198.

Garton, J. (1996) Interactive concordancing with a specialist corpus. *ON-CALL*, **10**(1): 8–14. http://www.cltr.uq.edu.au/oncall/garton101.html

Gaskell, D. and 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.

Granger, S. and Rayson, P. (1998) Automatic lexical profiling of learner texts. In: Granger, S. (ed.), *Learner English on computer*. London: Addison Wesley Longman, 119–131.

Granger, S. and Tribble, C. (1998) Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In: Granger, S. (ed.), *Learner English on computer*. London: Longman, 199–209.

Granger, S. and Tyson, S. (1996) Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, **15**(1): 17–27.

Hadley, G. (2002) An introduction to data-driven learning. *RELC journal*, **33**(2): 99–124.

Holec, H. (1990) Des documents authentiques, pour quoi faire? *Mélanges CRAPEL*, 65–74.

Horst, M. and Cobb, T. (2001) Growing academic vocabulary with a collaborative on-line data-base. In: Morrison, B., Gardner, D., Keobke, K. and Spratt, M. (eds.), *ELT perspectives on IT and multimedia: selected papers from the ITMELT conference 2001*. Hong Kong: Polytechnic University, 189–225. http://elc.polyu.edu.hk/conference/papers2001/cobb.htm

Hyland, K. and Milton, J. (1997) Qualification and certainty in L1 and L2 students' writing. Reprinted. In: Sampson, G. and McCarthy, D. (eds.), *Corpus linguistics: readings in a widening discipline*. London: Continuum, 371–386.

Johns, T. (1986) Micro-Concord: a language learner's research tool. *System*, **14**(2): 151–162.

Johns, T. (1991) Should you be persuaded: two examples of data-driven learning. In: Johns, T. and King, P. (eds.), *Classroom concordancing*. (*English language research journal, 4*), 1–16.

Johns, T. (1997) Kibbitzing one-to-ones (Web version). *BALEAP: academic writing*. University of Reading, 29 November. http://www.eisu.bham.ac.uk/johnstf/pimnotes.htm

Kennedy, C. and Miceli, T. (2002) The CWIC project: developing and using a corpus for intermediate Italian students. In: Kettemann, B. and Marko G. (eds.), *Teaching and learning by doing corpus analysis: proceedings of the fourth international conference on teaching and language corpora*. Amsterdam: Rodopi, 183–192.

Kilgarriff, A. (2001) Web as corpus. Reprinted in: Sampson, G. and McCarthy, D. (eds.) (2004) *Corpus linguistics: readings in a widening discipline*. London: Continuum, 471–473.

Koosha, M. and Jafarpour, A. (2006) Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL adult learners. *Asian EFL journal quarterly*, **8**(4): 192–209. http://www.asian-efl-journal.com/December_2006_EBook.pdf

Lake, J. (2004) Using 'on the contrary': the conceptual problems for EAP students. *ELT journal*, **58**(2): 137–144.

Lamy, M.-N. and Klarskov Mortensen, J. (2007) Using concordance programs in the modern foreign languages classroom. Module 2.4. In: Davies, G. (ed.), *Information and communications technology for language teachers (ICT4LT)*. Slough: Thames Valley University. http://www.ict4lt.org/en/en_mod2-4.htm

Levy, M. (1990) Concordances and their integration into a word-processing environment for language learners. *System*, **8**(2): 177–188.

O'Sullivan, I. and Chambers, A. (2006) Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of second language writing*, **15**(1): 49–68.

Renouf, A., Kehoe, A. and Banerjee, J. (2007) WebCorp: an integrated system for web text search. In: Hundt, M., Nesselhauf, N. and Biewer, C. (eds.), *Corpus linguistics and the web*. Amsterdam: Rodopi, 47–67.

Roby, W. (2005) The internet, autonomy, and lexicography: a convergence? In: Debaisieux, J-M. and Boulton, A. (eds.), *TIC et autonomie dans l'apprentissage des langues*. (*Mélanges CRAPEL*, 28), 47–66.

Schiffrin, D. (1987) *Discourse markers*. Cambridge: Cambridge University Press.

Scott, M. and Tribble, C. (2006) *Textual patterns: key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Sealey, A. and Thompson, P. (2007) Corpus, concordance, classification: young learners in the L1 classroom. *Language awareness*, **16**(3): 208–216.

Sinclair, J. (2003) *Reading concordances: an introduction*. Harlow: Pearson Longman.

Sinclair, J. (2004) New evidence, new priorities, new attitudes. In: Sinclair, J. M. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 271–299.

Stevens, V. (1991) Concordance-based vocabulary exercises: a viable alternative to gap-filling. In: Johns, T. and King, P. (eds.), *Classroom concordancing*. (*English language research journal, 4*), 47–61.

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. and Wang, L-Y. (2003) Concordancers in the EFL classroom: cognitive approaches and collocation difficulty. *CALL*, **16**(1): 83–94.

Swan, M. (2005) *Practical English usage*. Oxford: Oxford University Press, 3rd edition.

Test of English for International Communication (TOEIC) http://www.uk.toeic.eu/index.php?id=2760

Thurstun, J. and Candlin, C. (1997) *Exploring academic English: a workbook for student essay writing*. Sydney: CELTR.

Thurstun, J. and Candlin, C. (1998) Concordancing and the teaching of the vocabulary of academic English. *English for specific purposes*, **17**(3): 267–280.

Todd, R. W. (2001) Induction from self-selected concordances and self-correction. *System*, **29**(1): 91–102.

Widdowson, H. G. (1998) Context, community, and authentic language. *TESOL quarterly*, **32**(4): 705–716.

Wilson, E. (1997) The automatic generation of CALL exercises from general corpora. In: Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.), *Teaching and language corpora*. Harlow: Addison Wesley Longman, 116–130.

Yoon, H. and Hirvela, A. (2004) ESL student attitudes toward corpus use in L2. *Journal of second language writing*, **13**(4): 257–283.