



HAL
open science

DDL: Reaching the parts other teaching can't reach?

Alex Boulton

► **To cite this version:**

Alex Boulton. DDL: Reaching the parts other teaching can't reach?. Teaching and Language Corpora 8., Jul 2008, Lisbon, Portugal. pp.38-44. hal-00326706

HAL Id: hal-00326706

<https://hal.science/hal-00326706>

Submitted on 5 Oct 2008

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

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

DDL: Reaching the parts other teaching can't reach?

Alex Boulton

CRAPEL-ATILF/CNRS, Nancy Université - France

The potential applications of electronic corpora in language teaching and learning have received considerable attention in recent years, but their direct application to the classroom has not become part of mainstream practice. Indeed, the majority of published research in data-driven learning (DDL) focuses only on advanced learners using sophisticated equipment for complex language points. But there seems to be nothing essential about including all these ingredients from the start. In particular, DDL in early stages can eliminate the computer from the equation by using prepared materials on paper – considerably easier for the novice learner to deal with.

This paper reports on a simple experiment to see how lower-level learners cope with such paper-based corpus materials and a DDL approach compared to more traditional teaching materials and practices. Pre- and post-tests show both are effective compared to control items, with the DDL items showing the biggest improvement. A questionnaire shows a favourable reaction to the activities.

It is hoped that a larger body of empirical evidence will help DDL break out of its current research confines and into more mainstream teaching practices, especially if it can be shown that DDL has benefits for lower-level learners without expensive resources or extensive training. Simple materials of this sort, if they are seen to be effective, might counter a number of frequent objections to DDL, and contribute to greater awareness of its potential as they require little training, are easily shared, and can be incorporated into published materials.

data-driven learning, worksheets, concordance print-outs, lower levels

Background

Electronic corpora have made their mark in many areas connected with language teaching and learning. They can even be used directly by teachers and learners in what Johns (1991a) calls data-driven learning or DDL; but it is notable that such uses have not crossed over into mainstream practice or been taken up by major publishers (Boulton 2008b). There are certainly many barriers to the implementation of DDL, and it may be that work to date has been insufficiently convincing, showing frustratingly small effects (Boulton & Tyne 2008) and concentrating on a minority audience: a survey of fifty empirical DDL studies (Boulton 2008a) found only four conducted outside higher education, and only four with beginning or low level learners. It may even be that current research encourages the belief that DDL is *only* useful for advanced

learners in a computer laboratory, and with experts (i.e. the researchers) devoting considerable time to developing corpora and training learners in small groups.

One particular problem is that researchers have to keep one eye on publishable output by pushing things ever further, so tend to focus on an ambitious hands-on approach to corpus manipulation. It is unsurprising that learners find it difficult to get to grips with new material (the corpora), new technology (the software) and a new approach (DDL) all at once – especially at lower levels of language ability. The methodology itself is “revolutionary” enough (McCarthy 2004: 16) to warrant keeping other things simple, and one way to do this is to take the computers out of the equation at the start: much research has found the technological aspects to be a substantial source of frustration (e.g. Farr 2008), and students may even be “technophobic” (Bernardini 2002).

The use of prepared, paper-based materials has other benefits too: no need for a computer laboratory, computers that go wrong, websites that crash, unexpected findings, and so on. Such materials ought to be useful in themselves, and provide a gentle lead-in to direct interaction with the corpus (e.g. Estling Vannestål & Lindquist 2007; Turnbull & Burston 1998). Although this was part of Johns’ (1991b) original vision of DDL, there have been only occasional attempts to promote prepared concordance print-outs from publicly available resources (e.g. Chambers 2007); Boulton’s (2008a) survey found only eight examples.

Although there is comparatively little research to date with lower level learners and using paper-based materials, what does exist is encouraging (see Boulton 2008a for a summary). This paper reports an empirical study combining the two.

Method

The learners in the study were second-year architecture students in France with no prior experience of DDL. The questionnaire results are based on 71 sets of data and the experimental results on 62, some students being absent at crucial stages. The median age at the time was 19½; most were women (38/62); all but six had French as a mother tongue. Most had been studying English for eight years, though levels are not high: in a start-of-year levels test based on the TOEIC, the average score was only 52.85% overall, corresponding to approximately 450 on the official TOEIC scale, towards the lower end of the “intermediate” band (405-600).¹

Prior to the experiment, fifteen common problems were selected from students’ own written productions for greater perceived relevance (cf. Seidlhofer 2000). The focus was on grammar/usage as this tends to lend itself to a corpus approach rather than, say, purely grammar (though see Boulton 2007) or formal problems as DDL is generally more suited for depth of knowledge rather than for learning new items (Cobb 1999). The experimental session was conducted during normal class time with the regular teachers. Students were first given a five-minute introduction to corpora and their potential applications based on a specially-prepared booklet. This print document then

¹ The students’ objective is an official score of 700 points by the end of their third year as a precondition to obtaining their architecture diploma.

presented ten of the items (the remaining ones being used as a control): five using corpus data and DDL techniques, five using dictionary entries and traditional teaching methods – dictionaries providing an obvious point of comparison (Yoon & Hirvela 2004). Different items were given different treatment in each group: this was considered the most reliable control as the same students are involved, thus eliminating a number of variables (e.g. Stevens 1991).

The sources used are all available free on line, and easy to use for regular teachers and learners. The corpus was the BYU interface to the British National Corpus (Davies n.d.); the materials mainly comprised selected but unedited KWIC concordances, usually of between five and thirty lines, as well as some information on register, frequency and collocation. The dictionary entries were taken from the monolingual Collins COBUILD English Dictionary for Advanced Learners (2003) and Collins English-French Electronic Dictionary (2005), both available via the Reverso website; the entries were presented in the same layout as the original. Every attempt was made to produce equivalent materials: one page for each language item, the information being interspersed with questions to focus attention on particular points.

Prior to the experiment, teachers had a one-hour training session on DDL, at the end of which not all were convinced it would work. For the DDL treatment, they were encouraged to stick to the format by allowing students to discuss the questions and data in small groups to reach their own conclusions on each item before class feedback. For the traditional presentation, the teachers were allowed to intervene as they saw fit on a traditional “knowledge transmission” model. The experimental session itself lasted one hour of a 90-minute class. It took between five and ten minutes to go through each item, although all teachers were surprised that the DDL treatment did not take substantially longer than the traditional treatment.

Knowledge of the ten language items, along with five others as a control, was assessed the week before the experimental session and three weeks later to test for recall. The test of thirty questions (two for each item) was in a familiar format based on the TOEIC part V: gap-fill sentences with four possible choices each, the contexts mainly derived from dictionaries and other teaching materials. Additionally, to assess reactions to the materials and methods used, students were asked at the end of the experimental session to complete a short questionnaire in French combining closed questions on a five-point Likert scale, and open questions to be completed in their own words.

Results and discussion

Test results

The overall scores are fairly low (only 14.3/30 in Test 1) even though the test questions were based on clear examples and offered only four possible answers. This suggests the items in question are, as intended, problematic for these learners. The highest scores were 25 out of 30 in Test 1 and 27 in Test 2; the lowest were six and eight respectively. The average scores increased from 14.6 in Test 1 to 17.4 in Test 2 – an improvement of 2.8 points, or 19.4% (Table 1).

	DDL /10	traditional /10	control /10	total /30
TEST 1	4.9	4.6	5.1	14.6
TEST 2	6.4	5.7	5.3	17.4
difference	+1.5	+1.0	+0.3	+2.8
change	+31.6%	+22.2%	+5.1%	+19.4%

Table 1. Average scores, by treatment.

There are two main ways to compare the data: in Table 2, the horizontal arrows show changes between tests; the vertical arrows compare different treatments. Taking the first of these, a two-tailed paired t-test shows there to be a significant improvement overall between tests ($p < 0.0001$). One possibility is that there may simply have been a “test effect”, with students scoring higher the second time simply as they become more used to the test design and what was required. There was indeed a small improvement in the control items of 5.1%, but this is not significant ($p > 0.5$). This means that the significant improvement must derive from the other items: a 22.2% increase in score for the traditional items ($p < 0.01$), 31.6% for the DDL ones ($p < 0.0001$). The first conclusions therefore are that the test effect is minimal, while both kinds of presentation do have a significant effect.

	Test 1	Test 2
Corpus items	●	● ● ● ●
Dictionary items	●	● ● ● ●
Control items	●	● ● ● ●

Table 2. Points of comparison.

The key question however lies in the vertical comparisons of Table 2, i.e. between the different types of presentation. Again using t-tests, there is no significant difference between the control items and the dictionary items ($p > 0.01$), nor between the dictionary items and the corpus items ($p = 0.15$); but there is a significant difference between the corpus and control items ($p < 0.001$). Although the DDL treatment was more effective than the traditional treatment, there is a 15% likelihood that this could be due to chance alone.

Another point of comparison can be made between students’ level, as measured by the start-of-year TOEIC scores, and the test results. Pearson’s product-moment coefficient shows a strong positive correlation with both: 0.82 with Test 1 and 0.77 with Test 2. It is also possible to compare levels against the performance on the three types of items in Test 2. Unsurprisingly, the correlation was strongest (0.76) for the control items: as these were not explicitly covered in class, the students could only draw on their previous knowledge of the language. The coefficient is lower but still substantial (0.54)

for the traditional items; in other words, it can be inferred that more advanced students gained greater benefit from using dictionaries and traditional teaching. On the other hand, the correlation is virtually non-existent for the DDL items (-0.13), suggesting that all levels benefited as much as each other from this type of information and approach.

Questionnaire results

The first four items on the questionnaire were closed questions, asking the students to compare the two approaches they had just experienced on a five-point Likert scale. Looking only at the positive results (*agree* or *strongly agree*), 30 of the 71 students found the dictionary work easy compared to 54 for the corpus work; 31 found the dictionary work useful, compared to 59 for the corpus work. 37 thought the dictionary work would help them avoid certain errors in the future (suggesting they felt they had learned something from the work), rising to 58 for the corpus work. Clearly the traditional treatment was less positively received overall: this is reflected in the final pair of questions, as only 28 students would like to do more dictionary activities in the future, while 51 would like to pursue the DDL work.

To see how favourably DDL was received at different levels, the learners were divided into three bands according to their start-of-year TOEIC test: the upper level corresponds to a TOEIC average of 68%; the middle level to 51%; the lower level to 39%. The responses are largely very positive for all categories; the highest band is perhaps slightly more receptive than the others, but the patterns of differences are not significant.

Two open questions allowed the students to say what they felt were the respective advantages of dictionaries and corpora. Dictionaries were considered most useful for new or unknown words (26) and for meanings or definitions (26), while 19 simply wanted translations. 20 were interested in usage information; for some this was best presented in the form of “rules”, while others preferred looking at the examples – although one particularly wanted meanings “independent of any context”.

Corpora, on the other hand, were felt to be most useful for the contexts and “concrete examples” which highlight usage and grammar (58), and to represent “practical English”, “frequent usage”, the “language of today”. Only six mentioned “formulae” or “idiomatic expressions” as such, though allusion to context and, more specifically, “words that go together” reveals a certain sensitivity to this. Most responses seem to refer to corpus use for productive purposes, although some explicit reference was also made to comprehension (13). Some were extremely enthusiastic, including the following:

- Very interesting, an experience to repeat several times with other usage difficulties.
- I'd never heard of corpora. Thank you!
- It's the first time I'd done this type of exercise – but none too soon! Thank you! I'll assimilate things better this time! (Now go and kick out the teachers in high school!!!)

The final closed question asked the students if they would prefer to explore corpora on their own on computer rather than via the intermediary of paper-based materials. Although the students had no experience of hands-on computer-based DDL, they showed comparatively little interest for this: only 20 of the 71 students (i.e. less than 30%) *agreed* or *strongly agreed*. It is worth noting that the highest of the three ability

levels was least keen on such an approach: 17%, compared to 38% of the middle group and 35% of the lowest. 55 of the students took the opportunity to explain why: nearly half (25) believed the prepared exercises would get straight to the point and avoid time-wasting, and teacher guidance would be essential to avoid drawing wrong conclusions from the mass of data. As two students pointed out, they would need to try hands-on DDL first, but two others simply found the possibility “unattractive”. Two felt that talking about things was a useful part of the activity rather than just sitting in front of a computer, while eight thought that “doing it themselves” would be more relevant, motivating and lead to more effective learning. More generally, many stressed the importance of context and felt that the numerous samples would help to “visualise” or get a “feel” for the items under study, whether via prepared materials or on their own.

Conclusions

The experiment reported here with learners at lower levels of language ability found that, with no prior training, they managed to gain significant benefit from prepared, paper-based DDL materials. In particular, they performed better with this approach than they did using dictionary entries and traditional teaching methods.

On the whole, these results seem to contradict the received wisdom that DDL is best reserved for more advanced, sophisticated learners, despite a number of counter-examples (e.g. Boulton in press; Yoon & Hirvela 2004). It certainly needs further exploration, though it does corroborate some of the findings of our earlier research (e.g. Boulton forthcoming). The interpretation offered then was that more advanced learners have reached their current level by traditional means – in other words, they are comparatively good with the system currently in place. Learners who have been through the same system but who come out with lower levels are, by definition, not as good at learning through traditional methods.

In fairness, the objections to DDL at lower levels are usually within the context of hands-on exploitation of corpora, and do not necessarily extend to the use of paper-based DDL materials. This is a potentially crucial point: if the findings here are confirmed, it suggests not only that DDL can be successful using paper-based materials, but that these can be used with a wide range of levels, and may thus serve as a stepping-stone to hands-on corpus exploration (Johns 1997: 113).

Most current research publications seem to throw learners in at the deep end, requiring them to master the concept of corpora, the software and DDL techniques all at once. However, it is still DDL even if learners initially work only with paper-based materials and not directly on a computer (Breyer 2006; Frankenberg-Garcia 2005); this makes the learners’ task considerably easier as it reduces a number of methodological obstacles by, among other things, reducing the amount of data and limiting the range of possible answers (Thompson 2006) – not to mention technical, logistical and financial obstacles for the teacher. A proven track record for such paper-based materials might also help to convince a wider public that surprisingly little investment is required for rapid and substantial returns.

Paper-based materials, we have seen here, can bring benefits in themselves (Chambers 2005: 121). They can also serve as an introduction before moving on to more

autonomous corpus exploration: clearly learners need to understand the nature of corpus data and analysis before they can explore on their own, and need guidance in their use – autonomy does not come automatically to all (O’Sullivan 2007). As Sun (2003: 609) points out, “the learning curve... is arduously steep, in that students tend to get confused easily about the concordancer outputs; thus, they need either a stronger degree of teacher involvement, or to learn in a more structured environment”; using paper print-outs may reduce some difficulties in early stages. Furthermore, learners such as ours may initially feel paper-based resources are more relevant or efficient and, as Whistle (1999: 77) puts it, simply have difficulty seeing “why the concordances could not be prepared in advance and handed out in class”.

DDL materials such as those used here are extremely time-consuming to produce: each of the items here required half a day’s work – compare to Johns’ (1991a) eight hours for a single handout. Clearly such investment cannot be expected in normal teaching contexts, and yet there are virtually no published materials available (Boulton 2008b): of the eight empirical studies using paper-based materials reported in Boulton (2008a), all but one had to create these materials themselves. Even downloadable worksheets remain scarce, are not necessarily transferable to new contexts, and are dependent on researchers’ goodwill. Greater research interest producing positive results might inspire publishers to produce materials in the area in the form of books or paying websites; integrating DDL activities into more general works; or including corpora and interactive tools on websites or DVD-ROMs which accompany their publications.

References

- Bernardini, S.** 2002. “Exploring new directions for discovery learning.” In *Teaching and Learning by Doing Corpus Analysis*, B. Kettemann & G. Marko (eds.). Amsterdam: Rodopi: 165-182.
- Boulton, A.** 2007. “DDL is in the details... and in the big themes.” In *Proceedings of Corpus Linguistics 2007*, M. Davies, P. Rayson, S. Hunston, & P. Danielsson (eds.). <http://www.corpus.bham.ac.uk/corplingproceedings07/> [Access date 05/2008]
- Boulton, A.** 2008a. “Evaluating corpus use in language learning: state of play and future directions.” Paper presented at the *American Association of Corpus Linguistics*. Provo, March 13-15 2008.
- Boulton, A.** 2008b. “‘Off-the-peg’ materials for data-driven learning.” Paper presented at the *New Trends in Corpus Linguistics for Language Teaching and Translation Studies: in honour of John Sinclair*, Granada, September 22-24 2008.
- Boulton, A.** forthcoming. “Testing the limits of data-driven learning: language proficiency and training.” *ReCALL*.
- Boulton, A.** in press. “Looking for empirical evidence of data-driven learning at lower levels.” In *Practical Applications of Language and Computers: PALC07*, B. Lewandowska-Tomaszczyk (ed.). Frankfurt: Peter Lang.

Boulton, A. and Tyne, H. “Learning with corpora: changing learning practices.” Paper presented at the 4th *Inter-Varietal Applied Corpus Studies (IVACS) group Conference: Applying Corpus Linguistics*. Limerick, June 13-14 2008.

Breyer, Y. 2006. “My Concordancer: tailor-made software for language learners and teachers.” In *Corpus Technology and Language Pedagogy: new resources, new tools, new methods*, S. Braun, K. Kohn & J. Mukherjee (eds.). Frankfurt: Peter Lang, 157-176.

Chambers, A. 2005. “Integrating corpus consultation in language studies.” *Language Learning & Technology* 9/2: 111-125. <http://llt.msu.edu/vol9num2/chambers/> [Access date 02/2006]

Chambers, A. 2007. “Language learning as discourse analysis: implications for the LSP learning environment.” *ASp* 51-52: 35-51.

Cobb, T. 1999. “Breadth and depth of lexical acquisition with hands-on concordancing.” *CALL* 12/4: 345-360.
<http://www.er.uqam.ca/nobel/r21270/cv/Breadth.htm> [Access date 03/2006]

Collins Cobuild English Dictionary for Advanced Learners, 4th edition. 2003.
HarperCollins. <http://dictionary.reverso.net/english-cobuild/> [Access date 01/2008]

Collins English French Electronic Dictionary. 2005. HarperCollins.
<http://dictionary.reverso.net/french-english/> and
<http://dictionary.reverso.net/english-french/> [Access date 01/2008]

Davies, M. n.d. *BYU-BNC: British National Corpus*. <http://corpus.byu.edu/bnc/x.asp>
[Access date 01/2008]

Estling Vannestål, M. and Lindquist, H. 2007. “Learning English grammar with a corpus: experimenting with concordancing in a university grammar course.” *ReCALL* 19/3: 329-350.

Farr, F. 2008. “Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users.” *Language Awareness* 17/1: 25-43.

Frankenberg-Garcia, A. 2005. “Pedagogical uses of monolingual and parallel concordances.” *ELT Journal* 59/3, 189-198.

Johns, T. 1991a. “Should you be persuaded: two examples of data-driven learning.” In *Classroom Concordancing*, T. Johns & P. King (eds.). *English Language Research Journal* 4: 41-16.

Johns, T. 1991b. “From printout to handout: grammar and vocabulary teaching in the context of data-driven learning.” In *Classroom Concordancing*, T. Johns & P. King (eds.). *English Language Research Journal* 4: 27-45.

Johns, T. 1997. "Contexts: the background, development and trialling of a concordance-based CALL program." In *Teaching and Language Corpora*, A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.). Harlow: Addison Wesley Longman, 100-115.

McCarthy, M. 2004. *Touchstone: from corpus to coursebook*. Cambridge: Cambridge University Press.
<http://www.cambridge.org/us/esl/Touchstone/teacher/images/pdf/CorpusBookletfinal.pdf> [Access date 04/2008]

O'Sullivan, I. 2007. "Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy." *ReCALL* 19/3: 269-286.

Seidlhofer, B. 2000. "Operationalizing intertextuality: Using learner corpora for learning." In *Rethinking Language Pedagogy from a Corpus Perspective*, L. Burnard & T. McEnery (eds.). Frankfurt: Peter Lang, 207-223.

Stevens, V. 1991. "Concordance-based vocabulary exercises: a viable alternative to gap-filling." In *Classroom Concordancing*, T. Johns & P. King (eds.). *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.

Thompson, P. 2006. "Assessing the contribution of corpora to EAP practice." In *Motivation in Learning Language for Specific and Academic Purposes*, Z. Kantaridou, I. Papadopoulou & I. Mahili (eds.). Macedonia: University of Macedonia.
http://www.reading.ac.uk/internal/appling/thompson_macedonia.pdf [Access date 05/2008]

Turnbull, J. and Burston, J. 1998. "Towards independent concordance work for students: Lessons from a case study." *ON-CALL* 12/2: 10-21.
<http://www.cltr.uq.edu.au/oncall/turnbull122.html> [Access date 04/2006]

Whistle, J. 1999. "Concordancing with students using an 'off-the-web' corpus." *ReCALL* 11: 74-80.

Yoon, H. and Hirvela, A. 2004. "ESL student attitudes toward corpus use in L2." *Journal of Second Language Writing* 13/4: 257-283.

Alex Boulton

I've been a full-time lecturer at Nancy Université since 1999 and a member of the CRAPEL (Centre de Recherches et d'Applications Pédagogiques en Langues) since 2004, now also part of the ATILF-CNRS. My fascination with lexis in language learning goes back to my PhD in 1998, with increasing emphasis on ICT since then. Corpus linguistics and data-driven learning provide an excellent opportunity to combine the two and, I'm convinced, have enormous potential. I'm particularly concerned with

empirical evaluation and experimentation to assess the efficiency of DDL for different learners in different circumstances, and so on. I always try to keep things as simple as possible and look at practical applications for regular teachers and learners – i.e. using free tools and simple techniques.

Homepage: <http://arche.univ-nancy2.fr/course/view.php?id=967>.