



**HAL**  
open science

## Learning outcomes from corpus consultation.

Alex Boulton

► **To cite this version:**

Alex Boulton. Learning outcomes from corpus consultation.. M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez. Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching., Equinox, 2010. hal-00502629v1

**HAL Id: hal-00502629**

**<https://hal.science/hal-00502629v1>**

Submitted on 15 Jul 2010 (v1), last revised 9 Jan 2015 (v2)

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

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

Alex Boulton. Forthcoming. Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (Eds.), *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox. [pre-publication version]

## Learning outcomes from corpus consultation

Alex Boulton, CRAPEL–ATILF/CNRS, Nancy-Université, France

### 1. 0. Introduction

Large corpora have been hailed as a veritable “revolution” in kuhnian terms for the field of language teaching (e.g. McCarthy, 2008: 564), certainly insofar as they have influenced syllabus design, testing, and even materials. However, the effects may remain largely invisible downstream to teachers and learners, especially if they remain passive consumers of upstream work. Certainly, as Carter (1998: 64) pointed out a decade ago, there is no necessary reason why corpus description (*what* to teach) should translate into classroom practice (*how* to teach). However, it is possible for learners to explore corpora themselves, either directly via a concordancer, or mediated by other software or materials, in what has come to be known as data-driven learning or DDL (Johns, 1990). DDL represents a fairly “radical” approach (Johns, 1988: 20-22), and remains generally confined to the university research environment despite efforts by many researchers to undertake a “corpus mission” (Römer, 2009: 92) to introduce DDL to a wider teaching community.

One frequently-cited problem could be the type of information available: in particular, there is very little in the way of ‘off-the-peg’ materials, or indeed any texts geared towards end-users, teachers and learners alike (see, for example, Boulton, in press a). Furthermore, there may be literally hundreds of academic papers discussing applications of corpora in language learning and teaching, but, it is routinely alleged, very few that attempt any empirical evaluation of the approach. It is commonplace for articles in many disciplines to call for more empirical research, but this seems to be particularly the case for DDL. Among many others, one could cite Johansson (2009: 42) who has recently insisted that:

Corpora [...] should not be used in language teaching just because we now have this wonderful tool and would like to apply it in language teaching as well. Their use is vindicated to the extent that it agrees with what we know about language and language acquisition, *and can be shown to be an effective learning tool.* [emphasis added]

Yet he goes on to lament the lack of “systematic studies testing the benefits of the approach” and calls for more “controlled experiments” (Johansson, 2009: 41). Corpus linguistics is an inherently empirical discipline, and a lack of empirical research in applying it to language learning would at best be ironic – and at worst might suggest that researchers have deliberately avoided the issue, or abstained from publishing negative findings. A lack of a solid empirical support base would thus undermine the arguments for DDL, leaving us with little more than rhetoric and *ex cathedra* pronouncements along the lines of ‘it seems...’ or ‘it is obvious...’.

A survey of empirical research in DDL published to date therefore seems appropriate. This is not a light undertaking, even as regards collecting relevant papers, as DDL itself lacks a single, watertight definition, leading to multiple interpretations (Boulton, *in press b*). For present purposes, a broad interpretation is chosen to ensure maximum coverage, and DDL is therefore taken to refer to any use of overt corpus data for foreign or second language (L2) learning or teaching, whether the researchers claim it as DDL or not. As the focus is on empirical research, papers are included only insofar as they subject some aspect of DDL to observation or experimentation with some kind of externally validated evaluation other than the researchers’ own intuition. Given these broad criteria, a search of the literature has so far brought to light at least 70 separate studies published to date in English alone.<sup>1</sup>

Perusing these papers, the variety of research questions becomes apparent, but they fall into a number of broad categories (some with multiple objectives into two or more at once). Firstly, there are those that attempt to test whether learners are capable of corpus investigation, and look at their behaviour using DDL. The evaluations tend to be qualitative in the main, analysing learners’ productions (especially written or oral project reports) and their representations, and use instruments ranging from teacher observation to class discussion, interviews, logbooks, diary grids, and other forms of self-report protocols. The results of such studies are hugely valuable, showing the multitude of different ways corpora may be introduced successfully to the language classroom.

Secondly, a large number of papers evaluate learners’ affective reactions to corpus use. Many of the same self-report instruments are used, though the most common here is the questionnaire, asking learners

---

<sup>1</sup> The full current list and summary information can be found on the website accompanying this book and on the author’s homepage (<http://arce.univ-nancy2.fr/course/view.php?id=967>). There are no doubt more than 70: I would be grateful for any leads to other empirical DDL papers.

whether they found the activities easy, useful, enjoyable; whether they think they learned anything from them; and whether they intend to use corpora again in the future. The results are overwhelmingly favourable, the vast majority of studies finding learners (and teachers) responding positively to the approach, with very rare exceptions (e.g. Whistle, 1999). The participants' reactions are certainly very important and have to be taken into account, though of course their subjective appreciations of their own learning may not be reflected in actual learning.

A third group of studies looks at specific outcomes of corpus use, though they further divide into two distinct categories. The first examines outcomes of using corpora as a reference tool, usually for writing, error-correction, or translation. These studies do provide objective, quantitative data from a variety of tests and evaluations of written performance, with highly encouraging results – indeed, it may turn out that the main advantage of corpora for language learners is as a reference resource. However, the main focus of this paper is on learning outcomes, a question not addressed directly in the majority of these studies: while some learning probably does result from using a corpus as a reference (just as one may learn something from consulting a dictionary), this is a separate issue requiring separate analysis.

Chambers (2007) has already provided a survey of empirical studies in DDL, but the present paper adds to this in two ways. Firstly, the 12 papers in her survey included evaluations of learners' behaviour and representations, as well as use of corpora as a reference resource. These studies provide rich insights, but are included here only insofar as they also report on learning outcomes, thus specifically addressing the key questions of whether DDL works and how effective it is. Secondly, Chambers noted that the majority were small-scale, qualitative studies, and that it is “worth asking why there are not more large-scale quantitative studies” (2007: 5). Again, without denigrating the relevance of small-scale, qualitative studies, focusing on measurable outcomes should allow more specific insight into the effectiveness of DDL in different contexts.

The present paper thus represents a survey of 27 empirical studies of DDL which focus on L2 learning outcomes, a surprising amount given the repeated lament of a lack of such research – and there are no doubt more, especially published in languages other than English which are not included in the present survey. The disadvantage of such a large number is that the discussion of individual studies will inevitably be fairly succinct. To help, *Table 1* provides some basic information on the studies: the paper(s) reporting each study; the country where it was conducted, along with the mother tongue (L1) of the majority of participants; the target language; the

type of institution where it was conducted, as well as the major field of study of most participants; their level, from low to advanced, passing through (lower or upper) intermediate where the information can be derived; the number of learners involved (including any control group, indicated separately); the duration of the study in hours, weeks or semesters; whether the participants used a hands-on concordancer, software including corpus data, or paper-based materials; the main research tools (including use of control items or populations); and whether some kind of statistical analysis is provided, or merely raw figures and percentages.

Section 2.0 describes the background context to the studies as a whole, along with research questions and design; section 3.0 briefly summarises the learning outcomes of individual studies, and examines issues of statistical significance. Finally, section 4.0 attempts to synthesise the findings to date and put them into some kind of perspective; the conclusion outlines areas for future empirical research.

study	country (L1)	L2	context	level	learners	time	interface	test design/tools	quant
Gan et al., 1996	Malaysia (Malay)	English	teacher-training (L2)	adv	48	10 wks (20h)	hands-on	pre+post test, control items	stat
Cobb, 1997a, 1997b, 1999	Oman (Arabic)	English	university (business)	low	11	1 year	program	pre+post test + delayed, control pop	stat
Ciesielska-Ciupek, 2001	Poland (Polish)	English	high school	low	33	?	paper	post test + delayed	raw n°s
Cobb, Horst, 2001; Horst et al., 2001	Canada (mixed)	English	university (mixed)	int	33	12 wks	program	post test, control items	stat
Cobb et al., 2001	Canada (English)	French	? (?)	int	1	14h	program	pre+post test	raw n°s
Curado Fuentes, 2003, 2002	Spain (Spanish)	English	university (business)	int	20 (c=10)	2 wks	hands-on	post test, control pop	raw n°s
Lee, Liou, 2003	Taiwan (Chinese)	English	high school	int ?	46	10 wks (8h)	hands-on	pre+post test	stat
Sun, Wang, 2003	Taiwan (Chinese)	English	high school	int ?	81 (c=40)	1h40	hands-on	pre+post test, control pop	stat
Chan, Liou, 2005	Taiwan (Chinese)	English	university (mixed)	int ?	32	5 wks (5h)	hands-on	post test + delayed, control items	stat
Kaur, Hegelheimer,	USA (mixed)	English	university (?)	int	18 (c=9?)	4 wks?	hands-on	post test, control pop	stat

study	country (L1)	L2	context	level	learners	time	interface	test design/tools	quant
2005									
Tian, 2005a, 2005b	Taiwan (Chinese)	English	university (mixed)	int +	98 (n=48)	5 wks (10h)	paper	pre+post test, control pop	stat
Allan, 2006	Ireland (mixed)	English	language centre (mixed)	adv	18 (n=5)	12 wks	paper	post test, control pop	stat
Koosha, Jafarpour, 2006	Iran (Farsi?)	English	university (L2)	adv	200 (n=100)	1 sem	paper	pre+post test, control pop	stat
Liou et al., 2006	Taiwan (Chinese)	English	university (?)	adv ?	varied	variable	program	post test + delayed	stat
Braun, 2007	Germany (German)	English	high school	int ?	25 (n=13)	4 wks (16h)	program	post test, control pop	stat
Cresswell, 2007	Italy (Italian)	English	university (translat)	adv	126 (n=65)	1 sem?	hands-on	post test, control pop	stat
Curado Fuentes, 2007	Spain (Spanish)	English	university (tourism)	int +	20 (n=10)	5h	hands-on	post test, control pop	stat
Estling Vannestål, Lindquist, 2007	Sweden (Swedish)	English	university (teachers; admin)	adv	a) 37 (n=23); b) 35	1 sem each	hands-on	pre+post test, control pop	raw n°s
Huang, Liou, 2007	Taiwan (Chinese)	English	university (L2?)	int	38	12 wks (out of class)	program	pre+post test	stat
Yeh et al., 2007	Taiwan (Chinese)	English	university (L2)	int +	19	4 wks (1h20)	hands-on	post test + delayed	stat
Belz, Vyatkina, 2008, 2005a, 2005b	USA (English)	German	university (?)	adv	2	8 wks (4h)	hands-on, paper	pre+post test	raw n°s
Boulton, 2008, 2010	France (French)	English	university (architect)	int -	62	1h	paper	pre+post test, control items	stat
Johns et al., 2008	Taiwan (Chinese)	English	high school	int ?	22 (n=11)	16 wks (48h)	program, paper, hands-on	post test, control pop	stat
Lin, 2008	Taiwan (Chinese)	English	university (L2)	adv	25	8 wks (13h20)	program	pre+post test + delayed	stat
Smith et al., 2008	Taiwan (?)	Chinese	internet volunteers (?)	int ?	25 (2 post tests)	6 wks	hands-on	pre+post test	raw n°s
Boulton, 2009a	France (French)	English	university (engineer)	int -	132 (n=64)	30 mins	paper	post test, control pop	stat
Boulton, 2009b	France (French)	English	university (architect)	int -	59 (n=25)	12 wks (3h)	hands-on	post test + delayed, control pop	stat

Table 1: Overview of studies

## 2. 0. Background

This section introduces the background to the studies – firstly the general context (countries, languages, learners, corpora and software), followed by the context specific to the research itself (participants, duration, language focus, study design and research instruments).

### 2. 1. General context

Though DDL first appeared as a term in this sense in 1990 in a paper by Tim Johns, he had discussed applications of corpora in language teaching and learning as far back as 1984 in a book co-authored with John Higgins; Sandra McKay had independently published a paper on such uses of corpora as early as 1980; and McEnery and Wilson (1997: 12) attribute the first applications of corpora to language teaching to Peter Roe at Aston University in 1969. It is perhaps surprising then that the first empirical study of learning outcomes did not appear until 1996. Since then, at least 27 studies have appeared in various sources, over two thirds of them (19) in the last five years. Some have been the subject of more than one paper, totalling 35 separate publications. Of these, 23 have been in journals (notably six in *Computer Assisted Language Learning* and four in *ReCALL*), and nine in publications arising from conferences, whether proceedings or selected papers. Although Cobb (1997a) devoted his doctoral thesis to DDL research, and a number of books on corpus use in L2 involve some discussion of DDL (e.g. Hunston, 2002; O’Keeffe et al., 2007), there has as yet been no book-length treatment of DDL itself, much less with detailed empirical analysis.

Asia seems to be a particularly fertile ground for empirical DDL research, with 10 of the 27 studies conducted in Taiwan alone (plus one each in Malaysia, Iran and Oman); a further 10 studies were produced in European countries (three in France, two in Spain, one each in Germany, Ireland, Italy, Poland and Sweden); the remainder were in North America (two each in Canada and the USA). This suggests that learners of many different linguistic and cultural backgrounds can make use of corpora, and also that there is substantial interest in DDL in Asia and Europe in particular, although there are of course other studies exploring various uses of corpora in language teaching and learning in other parts of the world. In 23 cases, the majority of the learners spoke the local language as their L1; the other four involved speakers from mixed L1 backgrounds. It is likely that this reflects the reality of much language teaching, with learners mainly

in their home environment tackling foreign rather than second languages. The target language was also fairly uniform, with 24 studies focusing on English; the remainder concerned Chinese, French and German. The dominance of English comes as no surprise due to its perceived influence today, though other factors may also be at work: corpora are more readily available in English than for most other languages; interfaces are frequently in English; awareness is likely to be higher given the number of publications in English; and it may even be that English, with its relatively reduced morphological variation, is more suitable for a corpus approach than some other languages.

Over two thirds of the studies (at least 19) involve learners in higher education, in seven cases majoring in the target language or translation studies; in eight they may be classified as advanced, in 17 as intermediate. Although it is frequently claimed that DDL may be most appropriate (or even exclusively so) for advanced, sophisticated learners, five of the studies take place in high school, and five (including four at university) explicitly claim the learners have lower intermediate levels of proficiency at best. The comparative rarity of studies in other contexts or with lower levels may be due to more prosaic factors. Universities represent the researchers' home environment where their students represent a captive audience; relatively sophisticated resources are also likely to be more readily available, not least of which is the time available for teachers and researchers, as well as greater flexibility in deciding their own programmes; and learners at university will already in most cases have spent several years studying the L2, whether they are majoring in it or need it for academic or special purposes. It would certainly seem worthwhile exploring further with younger learners in secondary education, with adults in other contexts (only a single study takes place in a language centre, one on-line), and with learners at lower levels.

It is more difficult to provide a completely accurate account of the corpora and tools used, not least because the information is not always explicit in the original articles; many also use a variety of corpora and tools in the same study. Seven use very large, often publicly available corpora, including the *British National Corpus* in four cases. However, the majority of studies draw on in-house corpora, often created by the researchers specifically for their students. The smallest is only 2000 words, but there appears to be no upper limit: nine use the web as corpus with a variety of tools such as *WebCorp* and *SketchEngine*. Most of the corpora consist of 'general' language, though six are compiled from news texts, and one or two each from literary texts, course books, business documents or advertisements. Two used graded texts from the corpus; only five of them involved parallel corpora (*TANGO*, *TOTALrecall* and *TeleKorp*), and one a



multimodal corpus (*ELISA*); these aspects thus reflect scope for future work if accessible software allows it.

In 14 of the studies, the learners had direct access to the corpus via a concordancer; eight use some kind of CALL program or other software which incorporates a concordancer; only eight make significant use of paper-based materials (three of the studies involving a combination). Again, the software used can be difficult to pin down, sometimes integrated to the particular corpus, sometimes written specially for the purpose or designed at the researchers' university. Other publicly available packages include *WordSmith Tools*, *LexTutor*, and Internet search tools. Most of the time, the learners accessed the data during the class or in special sessions in the computer room; occasionally this was supplemented with work outside class time or even at home. Only in Smith et al. (2008) was the whole experiment conducted elsewhere, in this case with volunteers recruited via Internet and working entirely on their own.

## 2. 2. Research contexts

On average, each study involved 47 learners, including control groups in 14 cases; four compared experimental and control items among the same population; 12 used a pre-test / post-test format, including seven also with a delayed post-test for longer-term retention. At the smaller end of the scale, two case studies featured only one or two learners; at the other end, three studies involved over 100 participants. These figures are sufficiently high overall to suggest that researchers have not ignored the importance of quantitative evidence for DDL. Indeed, all of the studies here provide some form of quantitative evaluation, and 21 subject their results to statistical analysis of some kind.

Different articles report the duration of the study in different ways, making comparison difficult. For those that give the minutes or hours, the average exposure to DDL is just over 10 hours. This would correspond to nearly an hour a week over a semester; indeed, those that describe the exposure in longer periods also report nearly 10 weeks of DDL. The shortest are clearly experimental in nature, with perhaps a single contact with corpus data (one study lasting only 30 minutes); however, 10 studies run for one or two semesters. The difference is apparent in a number of ways in the articles themselves: the studies based on shorter time-scales tend to begin with a research question and create an experimental situation to answer it; the longer ones typically begin with a course that is in place, and introduce research questions to assess some aspect of it.

While some of the corpora and tools used in the older papers may appear rather dated now, the methodologies employed have tended to remain fairly stable. Most studies used more than one research tool, including classroom observation, discussion, interviews, tracking, and most frequently questionnaires, although these were frequently for additional questions concerning learners' attitudes. For issues explicitly related to learning outcomes, a large variety of different tools were used for testing purposes, usually several in each study. Some are quite open and lead to a qualitative assessment (of writing, oral presentations, creating dictionary entries, etc.), but most require learners to complete fairly closed tasks (such as cloze, matching, sentence completion, substitution, error-correction, translation, or reading comprehension). Occasionally external tests were used, such as the *Vocabulary Knowledge Scale* and the *Vocabulary Levels Test*, as well as general exams not designed explicitly to test the items covered by the experiment – though most that did so acknowledge a number of problems inherent in this.

The majority of studies tend to have a fairly specific language focus, although some of the longitudinal data covers a range of features. Unsurprisingly, these mostly revolve around lexical aspects, as they are comparatively easy to search for in a corpus. The vocabulary may be of a particular type, such as connectors or items from Coxhead's (2000) *Academic Word List*, each featuring in two studies. The concordancing often includes induction and retention of meaning, but generally with a focus on larger units including compounds, clusters and collocations; this correlates with Johns' insight that DDL is most effective on the "collocational border' between syntax and lexis" (2002: 109). Although four studies do tackle wider themes of syntax and grammar directly, most are closer to the dividing line of 'usage', i.e. how words behave in different contexts, genres or text types. Some studies avoid a specific linguistic focus, including three that concentrate on reading skills, and one on noticing.

### 3.0. Learning outcomes

Although most of the studies surveyed here also feature some analysis of attitudes and behaviour, the present discussion will be limited to learning outcomes alone. As the majority focus on advanced adult learners using hands-on concordancing, we will first concentrate on the minority that look at younger or less advanced learners, or use of paper-based materials. Design problems are indicated at times, but space does not permit detailed

discussion of individual studies, and the reader is referred to the original papers.

### 3. 1. General outcomes

Firstly, five of the studies feature teenage learners in secondary schools. Ciesielska-Ciupek (2001) supplemented course books with Internet materials and concordance print-outs. Tests on language items covered were positive, and maintained four weeks later with no further revision; the testing procedures are not entirely transparent, however, and no statistical analysis is provided. The experimental group in Braun's (2007) study used a unit of her ELISA video corpus, scoring higher on the follow-up computer-based tasks, although there was no overall difference with the control group in the final test on the unit as a whole. Learners with an inductive preference scored higher in Lee and Liou's (2003) study, and all levels improved following the DDL sessions, with the differences between them being reduced; although these differences are not statistically significant, they are taken to show that concordancing is of particular benefit to lower-level learners. In a study on collocations, Sun and Wang (2003) report that the inductive group showed significantly greater improvement than the control group, especially for the apparently easier items tested. Finally, Johns et al. (2008) find that the experimental group performed significantly better than the control group in the post-test for reading comprehension, and had double the reading speed; they also performed significantly better on the end-of-term exam, suggesting improvement extending beyond the specific tasks covered.

Another four papers claim to implement DDL with lower-level learners in higher education. In his doctoral thesis and two other papers, Cobb (1997a, 1997b, 1999) reports learners in the experimental treatment scoring significantly higher on vocabulary learning overall, in terms of both meaning and use in new contexts; the advantage was maintained in delayed tests for longer-term retention. The other three studies were all conducted in France. In the first (Boulton, 2009a), the experimental groups performed significantly better in an immediate post-test, especially from KWIC concordances as opposed to complete sentences; however, no difference was found between groups in the delayed post-test, suggesting that corpus data may be most useful as a reference resource for lower levels of language ability. In the second (Boulton, 2008, 2010), learners improved most on items in the experimental treatment, although the difference with the control treatment is not significant; the lower-level students narrowed the

gap using DDL, while the more advanced ones maintained their advantage using the traditional approach. In the final study (Boulton, 2009b), learners who had experienced hands-on corpus exploration were later found to perform better in a test of noticing skills than the group undergoing traditional teaching, although the difference was not statistically significant.

Two other papers discuss implications of level, although neither claims to be dealing exclusively with lower levels. Chan and Liou (2005) find no significant correlation between level and post-treatment scores. DDL led to significantly greater improvement in both the immediate and delayed post-tests; the pattern of results is argued to show that an inductive DDL approach takes time to produce its maximum effect. Similarly, Tian (2005a, 2005b) reports no correlation with proficiency, and although both DDL and traditional teaching are successful, DDL is found to be significantly more useful for work on grammar and text type, but not significantly so for the usage points tested.

In total, eight studies make substantial use of paper-based materials. In three cases this is in combination with hands-on concordancing or other software, including the paper by Johns et al. (2008) discussed above. In Belz and Vyatkina (2008, 2005a, 2005b), following paper-based and hands-on treatment from native-speaker and peer-produced corpora, learners' on-line productions showed the target items being used more frequently and appropriately than before, although no statistical analysis is available as much of the study focuses on qualitative analysis of two learners only. Estling Vannestål and Lindquist (2007) report trials of experimental materials for grammar learning, although learning outcomes were only evaluated in one semester. No difference was detected between the experimental and control groups, perhaps because the post-test seems to have covered a wider range of items than those dealt with explicitly in the course.

Other papers involving printed materials include Ciesielska-Ciupek (2001), Boulton (2009a) and Boulton (2008, 2010) reported above. In addition to these, Koosha and Jafarpour (2006) compared treatments for prepositional collocations from paper-based materials only, the statistical analysis finding that the experimental group scored significantly higher in the use of the target language. A detailed paper by Allan (2006) shows the experimental group making significantly greater gains than the control group. Intriguingly, the advantage remained even for items not covered in the study, which is taken as evidence that the benefits of concordancing include strategies that can carry over to other language items.

Like Johns et al. (2008) mentioned above, the paper by Cobb et al. (2001) makes use of a single novel as a corpus, although here in a case

study of only one learner, ruling out any statistical analysis. Nonetheless, the immediate post-test shows substantial gains being made in vocabulary, and retained in a delayed post-test. Smith et al. (2008) also report a small-scale study – inadvertently so, as they started out with 25 volunteers recruited via the Internet. Only 2 completed the post-test, one of whom had achieved the maximum score on the pre-test; however, the other did double his score following the treatment. Lin (2008) encounters a similar problem deriving from very high scores on the pre-test, but the *Vocabulary Knowledge Scale* showed depth of knowledge improving significantly following the experiment. Additionally, analysis of the participants' written productions revealed substantial increase in productive use of the target items, which declined only slightly in the delayed post-test.

Future teachers showed significant improvement on experimental items in Gan et al. (1996), despite extremely limited computer skills to start with. Cobb and Horst (2001; also Horst et al. 2001) find DDL helping with definition writing. Concordancing was the most important factor behind significant gains in vocabulary overall, although these were small, which the researchers attribute to the general-purpose measurement tools used. The learners in the study by Liou et al. (2006) showed significant improvement on the various items covered, as well as in inductive learning, although it is clear that the various elements in this paper are intended mainly as pilot studies of the tools rather than rigorous experimental analyses of DDL.

The learners in Cresswell's (2007) paper used concordances either inductively or deductively depending on learning style preference; both groups were generally successful, with some caveats – in particular, the overt knowledge of connectors derived from corpus consultation was not found to translate well into use in essays, as the DDL group performed only very slightly better than the control group. On the other hand, Yeh et al. (2007) report that use of the target items did improve in learners' written productions; there was also significant improvement in immediate and delayed post-tests for collocations of over-used adjectives. Similarly, in Kaur and Hegelheimer (2005), learners used the target items significantly more frequently and more accurately in the final written assignment. The post-test also showed the experimental group performing better, though not significantly so, with no apparent correlation between concordance use and results. An analysis of videoed data in Curado Fuentes (2003, 2002) showed the experimental group making more errors, but also considerably more effective use of the target points in their oral presentations. In a separate study, Curado Fuentes (2007) reports the experimental group performing significantly better than the control group in reading related to their corpus work with tourist advertisements. Huang and Liou (2007) used

graded texts with lexical items colour-coded according to the number of presentations, with limited results: although learning improved as the words were met more frequently, some words encountered up to 15 times were still not learned.

The studies which use a pre-test / post-test design suggest that learning does take place after corpus consultation; in other words, the approach can be *effective*. However, given exposure of some kind, one might be forgiven for supposing that some learning will take place whatever the approach. It is therefore crucial to evaluate whether or not DDL is *efficient*. This is explored in the studies that compare experimental and control groups (or experimental and traditional treatment for language items), which on the whole give the advantage to DDL. Such a general statement of course needs to be supported by statistical analysis, the topic of the following section.

### **3. 2. Statistical significance**

The sheer diversity of research questions and designs makes a formal meta-analysis of the results above difficult, and probably impossible. However, a certain number of general observations do seem present themselves. Firstly, the overall body of empirical research in DDL provides overwhelmingly favourable reactions to DDL on the part of the learners; the very few exceptions include Estling Vannestål and Lindquist (2007) discussed above. As regards learning outcomes, the majority of the studies surveyed here are similarly encouraging, although detailed analysis suggests a slightly more mitigated picture, partly because even experienced researchers in applied linguistics may not be at ease with quantitative methods (Rasinger, 2008), leading to problems in design or analysis. In particular, six of the studies limit themselves to raw figures and do not (or, given the design, cannot) present serious statistical analysis of the main learning outcomes targeted (Belz, Vyatkina, 2008; Ciesielska-Ciupek, 2001; Cobb et al., 2001; Curado Fuentes, 2003; Estling Vannestål, Lindquist, 2007; Smith et al., 2008). Tellingly, of the 21 that do, none report DDL to be less effective than traditional teaching practices. Of course, it might be that negative results are less likely to be written up as they stand, though they may lead to modified research which does produce the desired results and which is then published. Certainly, it seems probable that the majority of researchers behind the papers in this survey are enthusiastic about the potential of DDL, and are therefore unlikely to be seeking actively to disprove its merits, however (un)scientific this may be. On the other hand,

five do report findings that fall short of the usually accepted levels of significance ( $p < 0.05$ ), although they may at times come close, and the results should therefore not be disregarded altogether (Boulton, 2009b, 2008; Cresswell, 2007; Kaur, Hegelheimer, 2005; Lee, Liou, 2003).

Some of the reportedly significant results also need qualification in the light of the experiment design. The paper by Liou et al. (2006) is evidently a series of pilot studies of work in progress and does not provide full description of the research design (e.g. the number of learners involved) or the results (e.g. the entire set of results from the delayed post-test). The results in Tian (2005a, 2005b) may be compromised by the different question formats for each type of language item in the pre- and post-tests, making direct comparison problematic. A similar concern may arise with the early study by Gan et al. (1996), as learners were apparently allowed to choose the items to study; this is not discussed further, and it is not clear how the common post-test catered for this. Allan (2006) attributes the carry-over of positive effects to untreated items as a sign that DDL promotes various strategies; an alternative may lie in the design of the experiment itself – in particular, there is substantial variability between the learners regarding the work completed, and the small control group of only five learners scored substantially higher in the pre-test than the experimental group, and thus had less room for improvement. Similarly, the learners in the study by Lin (2008) began with high scores on the vocabulary test, so only depth of lexical knowledge could be tested (though with significant results).

A number of other studies also report positive findings only for some of the research questions. Sun and Wang (2003) find a significant advantage for the experimental treatment only for the two easier collocation patterns tested and not the harder ones, while Boulton (2009a) and Braun (2007) each produce outcomes significantly in favour of corpus use in the immediate post-tests, but not in the delayed tests. The results reported in Huang and Liou (2007) and Cobb and Horst (2001) are significant, but the researchers report that they are nonetheless disappointingly small.

This leaves a total of six studies with unambiguously positive findings that meet the normal requirements of statistical significance, all published in peer-reviewed journals or books (Chan, Liou, 2005; Cobb, 1997a, 1997b, 1999; Curado Fuentes, 2007; Johns et al., 2008; Koosha, Jafarpour, 2006; Yeh et al., 2007). That is not to say that these researchers are blinded by enthusiasm: they typically qualify their findings with a variety of hedging devices ('overall', 'on the whole', 'in general', 'by and large', 'on average', etc.); are careful not to overgeneralise their conclusions to wider

populations in other circumstances for different language points; point out limitations of their study, especially in regard to uncontrolled variables; and, without exception, call for further research to validate their findings.

#### **4. 0. Discussion**

The results as presented here may appear somewhat pessimistic – often with mitigated outcomes, small or not statistically significant results, problems of research design, and so on. But there are grounds for cautious optimism. A comparison can be made with corpus linguistics as a whole, where it is a basic principle that a single piece of data (such as an individual concordance line) may be interesting, but needs to be interpreted with caution: only with a large number of cases can one begin to have confidence in the underlying patterns that emerge. The same is broadly true of empirical studies: individually the results are often promising, though inconclusive; but taken together, they can be highly encouraging. Statistically, a small amount of data is likely to produce results that are not significant, but pooling the results increases their value tremendously. This is usually the domain of meta-analyses, but as mentioned above, such an undertaking would currently seem to be unrealistic given the fragmented nature of the studies in this survey with their disparate research questions, designs and data reporting. However, were anyone to have access to the full data sets and the necessary tools and skills to combine them into a formal meta-analysis, it is difficult to imagine that effect size would not take the overall significance of the studies well beyond the usual levels of acceptability.

Furthermore, we should perhaps not expect absolutely clear-cut results from individual studies in any case – indeed, there might be reason for suspicion if this were so. The number of factors to take into account makes it virtually impossible to isolate a single variable absolutely, especially over longer periods. This underlines a problem inherent in most quantitative analysis, as the tendency is to base the findings on average outcomes, as if the learners constituted a “monolithic group rather than [...] idiosyncratic individuals” (Yoon, 2008: 32); or as Estling Vannestål and Lindquist (2007: 336) put it:

If a new idea is tested and it turns out not to be successful for everybody in the experimental group all the time, it is easy to draw the conclusion that the new methodology is not successful at all, even if it is perhaps successful for some people sometimes, which may in itself be a positive outcome.



Work on learning styles may be able to shed some light on this (Boulton, forthcoming), but there is no reason to suppose this is a problem unique to DDL. The overarching modern paradigm in language teaching and learning for the past three decades has been the communicative approach, which has encountered similar problems and has not necessarily found it any easier to produce convincing empirical research into learning outcomes (cf. Rasinger, 2008: chap. 1). Yet the communicative approach certainly did not wait for the evidence to be in before it began to spread. If it has come to prevail, it is rather because it has a broad theoretical basis supported by promising (rather than conclusive) individual studies that break the larger issues down into manageable research questions to reveal overall tendencies.

The same seems to be true of DDL, which has substantial theoretical support and, as seen here, empirical evidence that it can work in a variety of contexts. However, simply showing that it is effective may be playing to the sceptics, who are quick to point out a number of objections, especially in terms of logistical barriers – for example, that DDL can be tedious, mechanical and time-consuming, with learners drowning in unnecessarily complex data; that it depends on advanced ICT skills and entails extensive training for both learners and teachers; that it requires computer rooms full of expensive and complex technology which, even if available, is prone to breaking down; that only the most advanced, sophisticated and motivated learners can make sense of the complex data and truncated KWIC concordances; and so on (see Boulton, 2009c, for a fuller discussion). The studies here go to great pains to show that learners can gain benefits from corpus consultation and overcome these barriers; and while most of them may indeed focus on advanced, sophisticated adult learners with training in hands-on concordancing and access to sophisticated equipment, there is also an increasing number showing that simple incarnations of DDL (e.g. with paper-based materials or more controlled activities) can lead to immediate benefits even for lower-level learners with negligible training and limited resources.

But these are largely side-issues given the main advantages attributed to DDL – amongst other things, that it promotes a range of cognitive skills, and increases sensitisation and ability to deal with authentic language; that the interactive, discovery-based approach fits with the current constructivist view of language learning; that induction of patterns and regularities is a more ‘natural’ approach than the intellectually rigorous rule-based approach characteristic of much traditional teaching; that it can increase motivation where learners are allowed to pursue their own queries, leading to greater autonomisation and life-long learning; and so on.

Countless articles discuss such theoretical considerations and promote these arguments as the real advantages of DDL. However, few of the papers in this survey address such considerations, which tend to be ignored or glossed over in empirical studies – in part, no doubt, as all such long-term, general skills are extremely difficult to assess. The exceptions in the present study include Boulton (2009b), which finds some evidence that corpus training enhances noticing skills; and Johns et al. (2008) and Allan (2006), who both find that DDL learners score higher than controls on items not covered explicitly in the course, suggesting incidental benefits and increased learning ability. Five of the seven studies that use delayed tests also find that DDL is more effective than traditional treatment for longer-term retention.

## **5.0. Conclusion**

This survey has covered a number of studies that aim to evaluate learning outcomes from corpus consultation. Although it falls short of conclusive proof of the effectiveness of DDL, it does provide grounds for optimism. Firstly, there is substantially more empirical research into the learning outcomes of corpus consultation than is frequently alleged: 27 separate studies to date. If there are repeated complaints about the alleged lack of research, it may be partly because corpus linguists are more demanding of empirical studies than researchers in the field of language teaching and learning as a whole. Secondly, the overwhelming majority of studies produce encouraging results, even if they are not always statistically significant on all research questions. The evidence may not be totally foolproof, but this is true of many widely-held tenets in applied linguistics; crucially, what there is points to the usefulness of corpus consultation for language learning in the short term, and possibly enhancement of language sensitivity and learning ability. The studies here show that DDL can be usefully employed for learners of many different language backgrounds and in different situations when appropriately adapted, whether using sophisticated equipment or the simplest of materials, in pursuing individual language interests or in tightly controlled activities, for high and low levels alike.

Inevitably, research needs to continue, and a number of areas are in urgent need of further investigation, not least multimodal, spoken or parallel corpora. Other issues comparatively understudied include, as mentioned above, lower-level learners using simple tools and techniques for basic language questions – in other words, “ordinary teachers and

learners in ordinary classrooms” (Mauranen, 2004: 208). These of course include paper-based materials (Boulton, in press a) which require little if any training to use, and which might be suitable for bringing DDL to a wider audience, beyond the university environment to younger learners in state schools, as well as to adults in language centres and continued education, and indeed outside the formal educational context altogether. As Chambers (2007: 13) puts it:

If corpus consultation [...] is to become a common activity for learners across the broad spectrum of language studies (general language learning, literary studies, languages for specific purposes, translation, etc.), it would seem necessary for developments to take place in a broader context than that which has been examined here, namely the classrooms of researchers with expertise in corpora and concordancing. [...] It is perhaps outside the classroom that the next important step in research in this area will take place.

More is also needed on variables between different learners, including attitudes, motivations and learning styles (cf. Boulton, forthcoming). This might help to counter the problems of treating all learners equally in quantitative studies, where low average results may conceal a variety of different outcomes for individual learners. At the same time, the findings would need to be off-set against evaluation of the usefulness of generic materials, as most of the papers here create their own activities from scratch, any reuse being confined to other learners in the same institution. Johns himself (1990: 36) proposed there should be available a bank of reusable “ready-made DDL materials”; so far Hadley (2002) is the only one to have reported on their use (specifically, the *COBUILD Samplers*: Goodale, 1995), but his paper concentrates on learner behaviour and attitudes rather than learning outcomes. Developments such as these would help to take DDL out of the hands of expert teacher-researchers and make it accessible to ordinary learners.

The current state of empirical research into learning outcomes from DDL is, as this survey has shown, more extensive than frequently claimed. Sceptics are likely to seize on the inconclusive results from many individual studies, though it has been argued that little more could reasonably be expected. On the contrary, the overall weight of evidence is encouraging, implying that teachers should not hesitate to introduce DDL to their learners in a variety of contexts, though empirical research should of course continue.

## References

- Allan, Rachel (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.
- Belz, Julie A., Vyatkina, Nina (2005a). 'Learner corpus analysis and the development of L2 pragmatic competence in networked intercultural language study: the case of German modal particles'. *Canadian Modern Language Review* 62, 1: 17-48.
- Belz, Julie A., Vyatkina, Nina (2005b). 'Computer-mediated learner corpus research and the data-driven teaching of L2 pragmatic competence: the case of German modal particles'. *CALPER Working Papers* 4, April.
- Belz, Julie A., Vyatkina, Nina (2008). 'The pedagogical mediation of a developmental learner corpus for classroom-based language instruction'. *Language Learning & Technology* 12, 3: 35-52.
- Boulton, Alex (2008). 'DDL: reaching the parts other teaching can't reach?', in Ana Frankenberg-Garcia (ed.), *Proceedings of the 8<sup>th</sup> Teaching and Language Corpora Conference*. Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa, pp. 38-44.
- Boulton, Alex (2009a). 'Testing the limits of data-driven learning: language proficiency and training'. *ReCALL* 21, 11: 37-51.
- Boulton, Alex (2009b). 'Language awareness and medium-term benefits of corpus consultation'. *Proceedings of EuroCALL 2009: New trends in CALL – working together*. Valencia: Universidad Politécnica de Valencia.
- Boulton, Alex (2009c). 'Data-driven learning: reasonable fears and rational reassurance.' *CALL in Second Language Acquisition: New approaches for teaching and testing*. *Indian Journal of Applied Linguistics* 35, 1: 81-106.
- Boulton, Alex (2010). 'Data-driven learning: taking the computer out of the equation'. *Language Learning* 60, 3.
- Boulton, Alex (in press a). 'Data-driven learning: on paper, in practice', in Tony Harris and Maria Moreno Jaén (eds), *Corpus Linguistics in Language Teaching*. Bern: Peter Lang.
- Boulton, Alex (in press b). 'Data-driven learning: the perpetual enigma', in Stanislaw Roszkowski and Barbara Lewandowska-Tomaszczyk (eds), *PALC 2009*. Frankfurt: Peter Lang.
- Boulton, Alex (forthcoming). 'Corpora for all? Learning styles and data-driven learning'. *Proceedings of the 5<sup>th</sup> Corpus Linguistics Conference*.
- Braun, Sabine (2007). 'Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora'. *ReCALL* 19, 3: 307-28.
- Carter, Ronald (1998). 'Orders of reality: CANCODE, communication, and culture'; and 'A reply to Guy Cook'. *ELT Journal* 52, 1: 43-56; 64.
- Chambers, Angela (2007). 'Popularising corpus consultation by language learners and teachers', in Encarnación Hidalgo et al. (eds), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, pp. 3-16.

- Chan, Tun-pei, Liou, Hsien-Chin (2005). 'Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations'. *Computer Assisted Language Learning* 18, 3: 231-51.
- Ciesielska-Ciupek, Maria (2001). 'Teaching with the internet and corpus materials: preparation of the ELT materials using the internet and corpus resources', in Barbara Lewandowska-Tomaszczyk (ed.), *PALC 2001: Practical applications in language corpora*. Frankfurt: Peter Lang, pp. 521-31.
- Cobb, Tom (1997a). *From Concord to Lexicon: Development and test of a corpus-based lexical tutor*. Montreal: Concordia University.
- Cobb, Tom (1997b). 'Is there any measurable learning from hands-on concordancing?' *System* 25, 3: 301-15.
- Cobb, Tom (1999). 'Breadth and depth of lexical acquisition with hands-on concordancing'. *Computer Assisted Language Learning* 12, 4: 345-60.
- Cobb, Tom, et al. (2001). 'Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources'. Translation of 'Peut-on augmenter le rythme d'acquisition lexicale par la lecture? Une expérience de lecture en français appuyée sur une série de ressources en ligne', in Patricia M. Raymond and Claudette Cornaire (eds), *Regards sur la Didactique des Langues Secondes*. Montreal: Editions Logique, pp. 133-53.
- Cobb, Tom, Horst, Marlise (2001). 'Growing academic vocabulary with a collaborative on-line data-base', in Bruce Morrison et al. (eds), *ELT Perspectives on IT and Multimedia: Selected papers from the ITMELT conference 2001*. Hong Kong: Hong Kong Polytechnic University, pp. 189-225.
- Coxhead, Averil (2000). 'A new academic word list'. *TESOL Quarterly* 34, 2: 213-38.
- Cresswell, Andy (2007). 'Getting to "know" connectors? Evaluating data-driven learning in a writing skills course', in Encarnación Hidalgo et al. (eds), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, pp. 267-87.
- Curado Fuentes, Alejandro (2002). 'Exploitation and assessment of a business English corpus through language learning tasks'. *ICAME Journal* 26: 5-32.
- Curado Fuentes, Alejandro (2003). 'The use of corpora and IT in a comparative evaluation approach for oral business English'. *ReCALL* 15, 2: 189-201.
- Curado Fuentes, Alejandro (2007). 'A corpus-based assessment of reading comprehension in English', in Encarnación Hidalgo et al. (eds), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, pp. 309-26.
- Estling Vannestål, Maria, Lindquist, Hans (2007). 'Learning English grammar with a corpus: experimenting with concordancing in a university grammar course'. *ReCALL* 19, 3: 329-50.
- Gan, Siowck-Lee, et al. (1996). 'Modeling teaching with a computer-based concordancer in a TESL preservice teacher education program'. *Journal of Computing in Teacher Education* 12, 4: 28-32.
- Goodale, Malcolm (1995). *COBUILD Concordance Samplers 2: Phrasal verbs*. London: HarperCollins.
- Hadley, Gregory (2002). 'Sensing the winds of change: an introduction to data-driven learning'. *RELC Journal* 33, 2: 99-124.

- Higgins, John, Johns, Tim (1984). *Computers in Language Learning*. London: Collins.
- Horst, Marlise, et al. (2001). 'Expanding academic vocabulary with an interactive on-line database'. *Language Learning & Technology* 9, 2: 90-110.
- Huang, Hung-Tzu, Liou, Hsien-Chin (2007). 'Vocabulary learning in an automated graded reading program'. *Language Learning & Technology* 11, 3: 64-82.
- Hunston, Susan (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johansson, Stig (2009). 'Some thoughts on corpora and second-language acquisition', in Karin Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 33-44.
- Johns, Tim (1988). 'Whence and whither classroom concordancing?', in Theo Bongaerts et al. (eds), *Computer Applications in Language Learning*. Dordrecht: Foris, pp. 9-27.
- Johns, Tim (1990). 'From printout to handout: grammar and vocabulary teaching in the context of data-driven learning'. *CALL Austria* 10: 14-34. Reprinted in Tim Johns and Philip King (eds), *Classroom Concordancing*. *English Language Research Journal* 4: 27-45.
- Johns, Tim (2002). 'Data-driven learning: the perpetual challenge', in Bernhard Kettemann and Georg Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, pp. 107-17.
- Johns, Tim, et al. (2008). 'Integrating corpus-based CALL programs and teaching English through children's literature'. *Computer Assisted Language Learning* 21, 5: 483-506.
- Kaur, Jagdish, Hegelheimer, Volker (2005). 'ESL students' use of concordance in the transfer of academic word knowledge: an exploratory study'. *Computer Assisted Language Learning* 18, 4: 287-310.
- Koosha, Mansour, Jafarpour, Ali Akbar (2006). 'Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL adult learners'. *Asian EFL Journal Quarterly* 8, 4: 192-209.
- Lee, Chuen-Yi, Liou, Hsien-Chin (2003). 'A study of using web concordancing for English vocabulary learning in a Taiwanese high school context'. *English Teaching and Learning* 27, 3: 35-56.
- Lin, Ming-Chia (2008). 'Building a lexical syllabus on Moodle with web concordancers for EFL productive academic vocabulary'. *Proceedings of WorldCALL 2008*. Fukuoka: Fukuoka University.
- Liou, Hsien-Chin, et al. (2006). 'Corpora processing and computational scaffolding for an innovative web-based English learning environment: the CANDLE project'. *CALICO Journal* 24, 1: 77-95.
- Mauranen, Anna (2004). 'Speech corpora in the classroom', in Guy Aston et al. (eds), *Corpora and Language Learners*. Amsterdam: John Benjamins, pp. 195-211.
- McCarthy, Michael (2008). 'Accessing and interpreting corpus information in the teacher education context'. *Language Teaching* 41, 4: 563-74.
- McEnery, Tony, Wilson, Andrew (1997). 'Teaching and language corpora'. *ReCALL* 9, 1: 5-14.

- McKay, Sandra (1980). 'Teaching the syntactic, semantic and pragmatic dimensions of verbs'. *TESOL Quarterly* 14, 1: 17-26.
- O'Keeffe, Anne, et al. (2007). *From Corpus to Classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Rasinger, Sebastian M. (2008), *Quantitative Research in Linguistics*. London: Continuum.
- Römer, U. (2009). 'Corpus research and practice: what help do teachers need and what can we offer?', in Karin Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 83-98.
- Smith, Simon, et al. (2008). 'A corpus query tool for SLA: learning Mandarin with the help of SketchEngine', in Barbara Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications: State of the art*. Frankfurt: Peter Lang, p. 673-86.
- Sun, Yu-Chih, Wang, Li-Yuch (2003). 'Concordancers in the EFL classroom: cognitive approaches and collocation difficulty'. *Computer Assisted Language Learning* 16, 1: 83-94.
- Tian, Shiauping (2005a). 'Data-driven learning: do learning tasks and proficiency make a difference?' *Proceedings of the 9<sup>th</sup> Conference of the Pan-Pacific Association of Applied Linguistics*. Tokyo: Waseda University Media Mix Corp, pp. 360-71.
- Tian, Shiauping (2005b). 'The impact of learning tasks and learner proficiency on the effectiveness of data-driven learning'. *Journal of Pan-Pacific Association of Applied Linguistics* 9, 2: 263-75.
- Whistle, Jeremy (1999). 'Concordancing with students using an "off-the-web" corpus'. *ReCALL* 11, 2: 74-80.
- Yeh, Yuli, et al. (2007). 'Online synonym materials and concordancing for EFL college writing'. *Computer Assisted Language Learning* 20, 2: 131-52.
- Yoon, Hyunsook (2008). 'More than a linguistic reference: the influence of corpus technology on L2 academic writing'. *Language Learning & Technology* 12, 2: 31-49.