26

Classroom applications of corpus analysis

Thomas Cobb and Alex Boulton

1 Introduction

Corpus linguistics is almost by definition *applied* linguistics, as was tacitly acknowledged when the American Association of Applied Corpus Linguistics (AAACL) dropped its third A in 2008. Its methodologies can be applied far beyond the discipline itself (see McEnery *et al.* 2006: 8), not least in language teaching and learning, where its influence has been of three main types. The first lies in improved descriptions of language varieties and features which can inform aspects of the language to be taught; the second makes corpora and tools for analyzing them available to the teacher; the third puts them directly into the learner's hands. We begin this chapter with an overview of all three types before concentrating mainly on the third type in the final sections, since other chapters in this volume deal in more detail with corpora and vocabulary, lexicography and phraseology, pedagogical materials, and translation.

1.1 Upstream use

Early instantiations of the first approach predate modern electronic corpora, with famous examples including Thorndike and Lorge's *Teacher's Wordbook of 30,000 Words* (1944) or West's *General Service List* (1953) for English, and Gougenheim and colleagues' *Dictionnaire fondamental de la langue française* (1958) for French. Work on frequency lists continues to this day derived from ever larger, electronic corpora, such as the British National Corpus (BNC: Oxford, 1995) and the Corpus of Contemporary American English (COCA: Davies, 2009), and has spread to other languages, as seen in recent series of lists from Routledge based on corpora of Spanish, German, Portuguese, Chinese, Czech, Arabic, French, and Japanese. Of

course, frequency applies not only to words, but also to larger units like phrases and chunks, as in Martinez and Schmitt's BNC-based phrasal expressions list (2012). While by no means the only criterion, the basic idea is that frequency of form and meaning is the most reliable predictor of what can be most usefully taught at different points in the learning process, as argued by Cobb (2007) for the early stages, or Schmitt and Schmitt (2014) for later stages. This type of work can thus inform syllabus design and testing, as the choice and sequence of forms and meanings to teach and test becomes more empirically based, for example in the design of TOEFL tests (Biber *et al.* 2004) and frequency-based vocabulary tests (Nation and Beglar 2007). Frequency analysis of learner corpora can also help to determine what learners of different backgrounds typically can and cannot do at different levels, again feeding into syllabus design more effectively than previous attempts at contrastive analysis based on qualitative structural differences, as argued by Granger (e.g. 2009). The *English Profile* project from Cambridge University is a major example of this type of work informed by both native-speaker and learner corpora.

Corpus research has not only informed syllabus and testing but has also been the driving force behind many other tools in language description, one of the most influential being the COBUILD project at Birmingham University (see Sinclair 1987). This large monitor corpus was specifically designed with pedagogical aims in mind, including a radically new type of dictionary with the entries chosen and organized according to frequency, and uncompromisingly authentic examples taken from the corpus. All the large publishing houses have followed this lead, and today it is inconceivable to produce a dictionary in a major language without substantial corpus input. The influence does not stop at lexis but can also be exploited in the production of usage manuals and grammar books, such as the *Longman Grammar of Spoken and Written English* (*LGSWE*: Biber *et al.* 1999). Corpora have also been used in the construction of teaching materials, though in many cases (e.g. *Touchstone*; McCarthy *et al.* 2006) the activities are indistinguishable from those in traditional books; the innovation is that the language taught is based on "real" usage and frequency data rather than depending on the authors' (often fallible) intuitions or fortuitous occurrences in the language inputs selected for learners' attention.

But it is possible to go further still and make direct use of corpus material with learners. Reppen (2010b: ch. 2) and Bennett (2010: ch. 3) discuss activities that make explicit use of the corpus information featured in grammar books such as the *LGSWE*, sensitizing learners to issues of frequency, morphology, chunking, collocations, register, and so on. A small quantity of published materials include corpus data too, from grammar books (e.g. Thornbury 2004) to supplementary materials (e.g. Thurstun and Candlin 1997) and even full courses (e.g. Mohamed and Acklam 1995). In books like these, concordance lines and other corpus data are turned into activities that students can use to explore the

language, either deductively (e.g. to test a rule or categorize different uses), or inductively (i.e. to formulate their own hypotheses about usage).

1.2 Teacher use

This brings us to the second major use of corpora in the language classroom, when teachers consult corpus data directly rather than relying on decision-makers upstream. First, corpus tools can be applied to individual texts, in helping decide whether a text is appropriate and what elements to focus on. Free software such as VocabProfile online (www.lexutor.ca/vp) or AntWordProfiler offline (www.antlab.sci.waseda.ac.jp/) allows a teacher to input a text which is then returned with the lexis color-coded according to the frequency of each word in the BNC or COCA corpus. Such information can help with decisions about which items to teach in a given text, for example, ignoring or glossing over less frequent items while using the highly visible multiple occurrences of others as an aid to teaching in context (Cobb 2007).

From the teacher's perspective, corpora can help in deciding what to teach. Often the corpora used for this purpose are not large modern corpora like the BNC or COCA but rather smallish corpora like the Brown (Kučera and Francis, 1979), or else purpose-built and sometimes level-appropriate text collections not necessarily meant to be representative of a language in its entirety. Such corpora can be particularly useful in teaching languages for specific purposes where published materials are difficult to come by. Frequency of occurrence and typical usage can be a useful guide, though of course these need to be tempered by pedagogical considerations. Corpora can also provide a useful source of authentic language, as the teacher can select typical language *samples* to complement or replace the invented language *examples* often found in teaching materials (Gavioli 2005: 7). This applies not just to teaching, but also to testing: Stevens (1991) found the use of multiple authentic concordance lines especially beneficial in gap-fill tests, effectively allowing English for specific purposes (ESP) tests to be constructed from authentic rather than made-up language.

Native and non-native teachers can also turn to corpora when they have a language question, as intuition is notoriously unreliable in many cases (even textbook rules are at times quite inadequate descriptions of actual language use; e.g. Carter *et al.* 1998). This can be helpful in correcting work outside the class, but can also serve as an in-class "informant" when responding to unforeseen language points. Where no explanation comes readily to mind, it gives the teacher a way to test intuitions, and an alternative to inventing a spurious rule or simply replying "because" (see Johns 1990). Finally, teachers can use corpus data in similar ways to the manuals outlined above, selecting corpus data (concordance lines, distributions, collocates, clusters, and so on) to create focused activities.

1.3 Learner use

Here we come to the third and final major use of corpora by language learners themselves. Corpus-based learning tasks and activities can be designed along a wide spectrum from "hard" to "soft" (see Gabrielatos 2005), beginning with totally controlled exercises as in the examples above: the teacher can decide the question, query a relevant corpus, and choose the appropriate information, which is then modeled into an activity with focused instructions and closed answers leading to predetermined outcomes. With time, any or all of these decisions and stages can, however, be taken over by learners themselves. The learner querying of corpora involves techniques that are essentially akin to the activities of corpus linguists: "Like a researcher, the learner has to form preliminary hypotheses on the basis of intuition or scanty evidence; those hypotheses then have to be tested and rejected or refined against further evidence, and finally integrated within an overall model" (Johns 1988: 14). Corpus consultation in this manner may focus on learning *per se*, or it may use a corpus as a reference tool alongside dictionaries and other resources in both comprehension and production, especially of written language. In reading, learners can quickly check specific patterns that may not be frequent enough to warrant a mention in dictionaries, or they can access all the occurrences of unknown words or uses in a given text, thus providing more relevant and focused contexts than may be found in a dictionary (Cobb *et al.* 2001). In drafting or revising texts or translations, learners can also check their tentative work against "normal" use in large or specialized corpora (e.g. O'Sullivan and Chambers 2006; Gaskell and Cobb, 2004).

Clearly in its most open-ended form, such activity can be quite demanding on the learner, who is likely to need intensive training or, perhaps preferably, scaffolding during extensive practice over a period of time in order to reap the full benefits of corpus consultation. We therefore need sound theoretical reasons to introduce work of this type, to be clear we are not doing so for contrived reasons (Chambers *et al.* 2004). The basic idea is that massive but controlled exposure to authentic input is of major importance, as learners gradually respond to and reproduce the underlying lexical, grammatical, pragmatic, and other patterns implicit in the languages they encounter. This can be through unconscious habit-formation from a behaviorist/emergentist perspective – see Hoey's (2005) theory of priming, or Taylor's (2012) account of implicit accumulated memories in Mental Corpus theory – or through some element of conscious noticing from a language awareness perspective. Other proposed benefits include the motivation inherent in use of ICT for individualized, relevant purposes where the learners build their knowledge based on their own needs and interests; learner corpus work is thus a generally constructivist and

inductive approach to language learning, the discovery and problem-solving procedures favoring cognitive and metacognitive development, critical thinking and noticing skills, language awareness and sensitivity in dealing with authentic text, as well as autonomy and life-long learning (see e.g. Römer 2006: 26; O'Sullivan 2007: 277–278).

All of these would appear to be desirable elements in current applied linguistic thinking. The question of course is whether corpus work really lives up to expectations, with benefits sufficient to justify the investment. For this, we need to look at research to date, which is the purpose of the rest of this chapter. The following section takes an overview of the research field as a whole, then focuses in on a number of studies we have conducted. The subsequent section takes the form of a preliminary meta-analysis in order to assess more broadly the benefits derived (or costs incurred) from the direct use of corpora by learners.

2 Empirical research in L2 corpus use

Getting learners to explore language is nothing new: they are frequently asked to compare example sentences on the blackboard, or identify features of written or spoken texts (Boulton and Tyne 2014). Using corpora merely moves it up a level, increasing the quantity of data available for examination, systematizing the querying procedures and output language, and potentially allowing learners a greater role in the process. According to McEnery and Wilson (1997: 12), the first such uses of corpora go back to the late 1960s at Aston University in Birmingham; other beginnings can be found in ESP courses at the University of Nottingham in the early 1970s (Butler 1974). The first published paper to our knowledge is by McKay (1980) at San Francisco University, describing learner use of printed corpus-based materials; the first description of hands-on concordancing can be found in Ahmad *et al.* (1985) at the University of Surrey. But the approach is largely associated with Tim Johns at the University of Birmingham, where he and other colleagues allowed their students access to COBUILD and other corpora and software in the 1980s for pedagogical purposes (see Johns and King 1991). Since then, there have been tremendous advances: many large corpora are available free on the Web (e.g. bncweb.lancs.ac.uk or corpus.byu.edu), as is software to aid rapid compilation from internet sources in just a few minutes (e.g. bootcat.sslmit.unibo.it), not to mention simple, stable, fast, and free tools with user-friendly interfaces (e.g. www.antlab.sci.waseda.ac.jp), often accompanied by video tutorials and online help.

Most of the early academic publications emanating from all this activity were descriptive and argumentative in nature; the first empirical evaluation comes from Baten *et al.* (1989). A much-lamented paradox of data-driven learning (another term for corpus-based learning) has been the slow

appearance of very much research data investigating whether learners do actually benefit from corpus consultation as a part of their language learning (e.g. Flowerdew 2012: 206). There are some reasons for this apparent lack of empirical support (e.g. the long-term nature of some of its goals are hard to operationalize, such as fostering autonomy, noticing, pattern induction, and language awareness). Nonetheless, our consultation of various databases, intensive trawls of individual journals, and serendipitous findings brings together a total of 132 papers which seek to empirically evaluate some aspect of corpus use in foreign or second language (L2) learning and teaching;[1] eliminating duplicates reduces the number to 116 (i.e. where the same study was presented in more than one paper). This is a not inconsiderable body of work.

Of these 116 publications, 76 were published in 36 different journals, 53 of them ranked on the 2011 *European Reference Index for the Humanities* (*ERIH*) lists; 35 were book chapters, some from major publishers, often resulting from thematic conferences (11 include the word *proceedings* in the title); the remainder are "fugitive" literature in the form of unpublished PhDs and working papers. Though they spread from 1989 to 2012, the increasing interest can be seen in that nearly half the papers were published in the last five years. Virtually all the publications are in English; though this might be due in part to search bias, we have only found five in French, which suggests that publications in other languages are likely to be comparatively rare too. About half of the total were conducted within the European Union, and half of the rest in Asia; most were in a foreign language environment, but about a third comprised mixed L1 classes in a second language context. English was the target language in 95 cases, though some feature learners of French (eight studies) or another European language, and in one case Chinese.

Over 100 of the studies are from higher education settings, though only about half seem to feature students majoring in languages (such basic meta-data are often frustratingly missing). There are at present only nine studies in secondary education, and a handful of other contexts such as language schools. Unsurprisingly, perhaps, many of the participants have quite substantial language proficiency: advanced or upper-intermediate in just over half, but lower levels in at least fifty studies. The language objectives generally tend towards the level of vocabulary or lexico-grammar (including clusters and collocations, i.e. word usage in context), but there are attempts to use corpora in learning grammar and syntax, and even occasionally in phonetics or semantics. A recent development is an increase in studies at the level of text, including discourse and critical analysis, genres, sensitivity to text type or sociolinguistic variation. Some go further still, using corpora in courses on literature or cultural studies

[1] References to studies in most of these categories can be found in the meta-analysis that follows. A complete and evolving list can be found in the supplement to Boulton (2010a) on the author's homepage.

with non-native students who thus combine linguistic and non-linguistic uses of corpora.

The Web is used as a corpus in ten studies, whether through a general-purpose search engine (e.g. Google) or a dedicated concordancer (e.g. www.webcorp.org.uk). Large corpora such as the BNC or COCA feature in about a third of all studies, but about half use locally built corpora, especially where the students have specific disciplinary or language needs such as writing research articles. These are sometimes created by the learners themselves, and can comprise as few as 2,000 words. It is worth noting that only 26 studies use corpora that are available free online, which means that many students would not be able to continue their explorations after the end of their courses. Mostly these corpora are explored on computer, only 24 using exclusively or in part printed activities derived from a corpus. WordSmith Tools is used in 18 studies despite its relatively advanced features and interface; AntConc and LexTutor are also popular, and a small number use purpose-built concordancing software.

The study duration varies from just a few minutes in some experimental contexts to a semester or more in five cases; the majority involve part of a course that lasts several hours over a few weeks. There is an average of 40 participants (including control or comparison groups), ranging from case studies with just one participant to quite large-scale studies with 100 or more. This gives rise to considerable methodological heterogeneity, with statistical analysis of quantitative results in 49 studies, raw figures and percentages in 41 more, and the remaining 26 favoring a purely qualitative approach.

This factual description of the work to date can do little more than scratch the surface. The rest of this section presents a small selection of our own empirical studies featuring a variety of research designs and objectives. They provide a flavor of research in this area and prepare the reader for a synthesis of some of the more general outcomes in the section that follows.

2.1 Learning with corpora (Cobb 1997b, 1999b)

This sequence of studies gathers together several of the themes introduced above: it uses an in-house corpus of learners' existing materials and purpose-built concordancing software; it responds to a specific learning need within an ESP context (English for commerce); it involves a mainly semantic analysis of concordance lines over a reasonably longitudinal exposure; and it measures its outcomes in both within- and between-subjects comparisons. It is also one of the earliest confirmations of "measurable learning from hands-on concordancing."

Cobb's (1997b) work with a corpus as a vocabulary learning tool took place in the context of a new university in a developing country (Sultan

Qaboos University, in Oman) that wished to use English as the medium of instruction but whose students were seriously underprepared for such a venture. This was particularly true with regard to the vocabulary needed for academic reading. The students' average vocabulary size was under 1,000 word families, while 3,000 families is typically reckoned a bare minimum (Cobb 2007). The goal of this project was to use corpus and concordance as a way for these students to meet and learn a relatively large number of words, for use in reading comprehension, in a relatively short time. The rationale for using a corpus was that the presence of metalanguage could make a purely definitional approach unsuitable, while the shortness of time available would not allow sufficient encounters with new words in context for natural word learning to occur. The corpus was a digitization of all the ESL materials that the students were using to prepare them for forthcoming English-medium study.

A target set of 240 word families was chosen as a 12-week test of a corpus-based approach to word learning. In a within-groups design, 11 learners met 20 new words per week via game-like computer activities that used either concordances or short definitions as an information source, on alternate weeks. A post-test of the 240 new words showed that 75.9 percent of the words met through concordances were retained, but only 63.9 percent of those met via definitions, an advantage for concordancing of more than one standard deviation.

Following this indication that corpus work could help these learners expand their lexicons, a scaled-up version of the project was prepared using two levels of learner, both experimental and control groups, two outcome measures corresponding to experimental and control conditions, and a learning target of 200 new word families per week for twelve weeks (or 2,400 words, roughly the number these learners would need to have a chance of reading for content in English). Experimental subjects used concordances to work with their new words exclusively, inferring meanings from multiple concordance lines and only using a dictionary to confirm their inferences, while controls used the same software but with a bilingual dictionary as the information source.

Weekly and pre-post tests recorded word knowledge on both definitional and novel-text gap-fill measures. It was hypothesized that learning words via concordances would facilitate the gap-fill task. The results showed that both experimental and control groups made significant and substantial pre-post gains on the definitional measures (4 to 8 percent), but only concordancers made significant gains on the novel text/gap-fill measure. This was true for both lower (13 percent gain) and upper intermediate concordancers (16 percent gain), gains of just under and just over one standard deviation, respectively. Further, a delayed post-test showed that even definitional knowledge was quick to decay for definitional learners, but the opposite was true for concordance learners (reported from different perspectives in Cobb 1999a, b). The advantages for concordance-based

vocabulary expansion seem clear, at least in these circumstances, though the generalizability of this finding remains to be determined.

2.2 Types of learning, types of learner (Boulton 2009, 2010b, 2011, 2012b)

As with Cobb, most of Boulton's work with corpora involves students who are not majoring in languages; this particular series of experiments involves first-year architecture students in France. For learners such as these, English classes are compulsory, but are not a major interest nor a priority within their overall degree; consequently, many have relatively low levels of English proficiency and lack inherent motivation for studying the language. Their overt objective for the end of their three-year degree is to attain at least an intermediate level (B1 on the Common European Framework of Reference for Languages); without independent certification for this they cannot graduate – also a source of some resentment. They are, however, intelligent, creative, and autonomous students; the question then was whether a discovery approach might help not only with their level of English but also in their motivation for the language, empowering them in their learning. A major difference with the work by Cobb is that here we were concerned to provide only publicly accessible corpora and tools, namely the BNC and COCA, 100 million and 400 million words of British and American English respectively.

In the first study in this series (Boulton 2010b), 62 learners were given a five-minute introduction to concordancing, then spent thirty minutes working in groups on printed corpus-based materials for five language items (inductively in pairs, feeding back to the whole group), and five others using standard dictionary entries, the instructions being as close as possible between the two groups. These problematic grammar/usage points had been collected from their own written productions earlier in the year, and featured in an earlier pre-test as well as a post-test the following week. The post-test showed significant improvement from both treatments (unlike for five untreated items). Although the improvement was greatest for the experimental treatments, the difference between the two was not significant. The students with lower levels did relatively better using corpora, while the more advanced ones maintained their advantage using the traditional approach. A final questionnaire showed very positive reactions to the experimental treatment. Overall, this study was taken to show that this student body could achieve results at least as good working with concordance lines as with other methods, without substantial training, and more importantly were open to the discovery approach in corpus use, especially those who had been less successful with traditional teaching methods in the past.

A subsequent question was whether such learners could cope with online corpus work. This allows greater learner responsibility and less

programmatic input, but also greater room for problems. In this longer-term study (Boulton 2012b), 40 students were again briefly introduced to corpus work, then experienced a variety of corpus activities on problem lexicogrammar points for a few minutes over ten weekly classes, alternating between paper-based and computer-based concordancing activities similar to the within-subjects design in Cobb (1997b) above. A test in the final session gave a small but not significant advantage to paper-based activities, though questionnaires showed the students had a slight preference for computer-based activities. They were generally receptive to hands-on corpus work as a whole, but surprisingly this did not seem to correlate with learning outcomes. A link was found between proficiency and outcomes from the paper-based treatment, but this can be interpreted as meaning computer-based data-driven learning is open to all levels even among these lower-intermediate learners.

In both these studies, the learners were generally receptive to working with corpora, but it was noted that there were quite substantial individual differences, suggesting that corpus work might not be equally appropriate for all learners. In the next study (Boulton 2009), 34 learners experienced hands-on concordancing as part of their class over twelve weeks, and then completed the Index of Learning Styles questionnaire adapted for French (Soloman and Felder 1996). This widely used psychometric instrument assesses respondents on four dimensions: Active–Reflective, Sensing–Intuitive, Visual–Verbal, and Sequential–Global. The objective here was to see if any of these proclivities correlated significantly with receptivity to using corpora, as rated by the participants themselves in a separate questionnaire. Of the learners who had the strongest feelings towards corpus use (positive or negative preference), the only significant correlation was that the most receptive were more likely to have a strong Visual learning style. This is consistent with a smaller pilot study (Boulton 2010c), though that suggested that liking corpus work and doing it well are not necessarily connected: those with an Active learning style achieved better outcomes. Though significant, these correlations are not very large, and the general conclusion is that learners with different learning styles can work successfully and enjoyably with corpora.

The final study was inspired by Allan (2006) and Johns *et al.* (2008), who independently found that corpus work seemed to lead to improved performance not only on targeted language items, but also in other areas. As neither study specifically focused on this, Boulton (2011) focused on noticing ability following corpus work in the same context as the previous studies with both paper-based and hands-on corpus work. At the end of the year, both experimental and control groups were given a short text to read for five minutes, then tested on whether they had noticed a number of language points (focus on form and on meaning) entirely unrelated to any work conducted during the year. The results show the experimental group performing better in noticing than the control group, though the

difference did not quite reach statistical significance but suggest it could do in a further better-targeted study.

The questions at the end of this discussion of some reasonably encouraging studies of learning from corpora are: how typical are these research studies? How typical are the results? Do enough of the larger cull of studies have the design criteria and data to support any sort of generalization about outcomes, and if so, what is the generalization? To answer these questions we assemble as much of the learner concordancing research as possible into a preliminary meta-analysis of findings.

3 A meta-analysis of corpus results

This chapter has so far surveyed various uses of corpora for language teaching/learning purposes. This type of "literature review" is common in the introductory sections of research articles, and the effects of corpus use have been the object of several extensive narrative syntheses (e.g. Chambers 2007; Boulton 2010a). This involves selecting the papers to review, deciding on their relative importance, interpreting the results, and putting everything back together to arrive at general conclusions, thus inevitably concealing a substantial degree of subjectivity. It is, however, possible to conduct a more rigorous survey in the form of a meta-analysis, which entails a near-exhaustive collection of studies in a given area (see Norris and Ortega 2006). The quantitative results are combined to provide a statistically meaningful picture over the many different situations covered, which clearly has advantages over the traditional narrative review in that it attempts to systematically reduce the bias inherent in subjective evaluation (Jeon and Kaya 2006), providing a way to "accumulate the results of the studies, the empirical findings, in as objective and data-driven a fashion as is possible" (Ellis 2006: 303). As with corpus linguistics itself, the adage "there's no data like more data" applies, and several non-significant results may, when combined, nevertheless contribute to substantial and significant findings. This methodology allows us to iron out many of the minor flaws in individual studies (assuming that the flaws in each are different); the counterpoint of course is that important differences can be lost, and great care is needed to avoid the trap of identifying the overarching research question with a single figure as a measure of its value.

Rather than providing new experimental data, this part of the chapter provides a preliminary meta-analysis of research in the field so far. For present purposes, the research questions are kept as simple as possible:

- Is corpus use *effective* for L2 learners – i.e. does it have a demonstrable effect?
- Is corpus use *efficient* for L2 learners – i.e. compared to other forms of learning?

While this may appear reductionist to an extent, it does respond to a clear desire on the part of researchers and practitioners to have simple answers to complex questions, and allows us to make some kind of sense of a highly heterogeneous collection of studies as objectively as possible.

3.1 Methodology

The procedures and criteria of meta-analysis in second language acquisition (SLA) are now well established, and the present consideration of the empirical work on integrating corpora in language teaching and learning will follow those of Norris and Ortega (2000) and Spada and Tomita (2010) as much as possible, although in less detail for this preliminary survey of research. The procedure will be to amass the greatest number of research studies with descriptive statistics (and ideally a control group) to calculate their standardized mean differences on the common scale of standard deviation units, or effect size, as measured by Cohen's *d*. This measure of effect size is, simply stated, the difference between two means (whether of the same group pre- and post-treatment, or experimental and control groups after treatment) divided by the combined standard deviation.

When an effect size has been calculated for each study (where this is possible), then a provisional average effect size and standard deviation can be calculated and the overall effect assessed within acceptable confidence limits. Upper and lower confidence intervals can be determined for the range within which the mean should statistically occur 95 percent of the time; if this range does not include zero, then the results can be deemed reliable. All things being equal, the larger the effect size, the more confident we can be that the focus variable is indeed statistically dependable. Traditionally, effect sizes up to $d = .2$ are considered small, $d = .5$ medium, and over $d = .8$ strong (Cohen 1988), though Oswald and Plonsky (2010: 99) suggest revising these up to $d = .4$, $d = .7$, and $d = 1.0$, respectively, to cater for the specificities of research in language teaching/learning.

The data considered here are drawn from the corpus of 116 individual studies described in the previous section. These date from 1989 to 2012, and include journal papers and book chapters, but also PhDs and conference proceedings (published as text and not just slides or oral presentations). Some meta-analyses avoid such "fugitive literature" (Norris and Ortega 2000: 431), but given the likelihood of a smallish number of eligible studies in the present meta-analysis, such studies are included here. However, the aim is not to pass judgment on the quality of individual studies, and all are weighted equally in the meta-analysis itself.

For this preliminary meta-analysis we retained only studies that focused on some kind of broadly defined "outcome" in terms of learning or of performance, in order to include, for example, using concordances as an aid to translation or in retrieving lexical items, which are not strictly speaking learning outcomes. In other words, this meta-analysis

investigates whether corpus use can have an effect over a wide range of variables, including vocabulary and grammar learning, error correction, lexical retrieval, and translation success.

Further exclusion criteria are needed for the purposes of a meta-analysis of this type; in particular, only experimental or quasi-experimental studies with a pre/post-test or a treatment/control group design, or both, can provide appropriate comparative data. It should also be noted that few studies assign students randomly to treatment groups, though the intact groups they use may themselves be randomly assigned; and the distinction between control and comparison groups is blurred.

It is precisely this type of quantitative reporting that is likely to be consistent over many studies, thus lending itself to comparison and synthesis. However, application of the exclusion criteria unfortunately means that many valuable qualitative studies cannot be represented – especially regarding such un- or under-operationalized variables like awareness, noticing, and autonomy which, as already mentioned, are difficult or impossible to quantify (Boulton 2012a). Even among the studies reporting quantitative data, essential information is often missing, from group sizes to means, or more frequently standard deviations, which in most cases cannot be calculated from the results. Following application of the exclusion criteria, the final number of papers included in this preliminary meta-analysis is thus reduced to just 21. This proportion of 18.1 percent (21 out of 116) is just over half of Norris and Ortega's 30.8 percent (77 out of 250) and Spada and Tomita's 33.0 percent (34 out of 103), both drawing on the more established research area of mainstream SLA. Where a single study reports several data sets, only the one representing the most relevant or concrete language learning or performance objective is included.

The pre/post-test and experimental/control studies were kept separate for the purposes of analysis, for the reasons outlined below. However, no other variables will be considered at this stage of the meta-analysis, such as participant meta-data (e.g. age, L1, L2, level of proficiency), instructional design (e.g. duration, hands-on or mediated interaction with various corpus types) or experiment design (e.g. immediate or delayed post-test). Many of these outcome types and conditions could be coded and investigated separately as moderating variables in a fuller meta-analysis, but that is beyond the scope of the present chapter. To conclude: while our meta-analysis will depart from the standard model on several points, the basic idea of the meta-analysis model is preserved.

Furthermore, this model is particularly suited to help us understand the state of research in this area, even in its nascent state. That is because studies are particularly vulnerable to the problems inherent in the significance-testing type of research, where the credibility of experiments depends so much on their *n*-sizes (see Norris and Ortega 2000; Ellis 2006), which in this area are often bound by the number of posts in a computer room.

3.2 Results

The 21 studies are summarized in Tables 26.1 and 26.2. These separate within-subject studies (comparing pre– and post-tests) and between-subjects studies (i.e. comparing treatment and control groups), as the different designs tend to produce rather different results. The former show whether the treatment is *effective* (whether or not there is a difference before and after), while the latter show whether the treatment is *efficient* (whether or not there is a difference compared to the comparison group). Since almost any form of instruction is likely to lead to some effect (the main conclusion from Hattie's 2009 meta-analysis of meta-analyses), it is to be expected that the results of a within-groups analysis will be markedly higher than a between-groups analysis. This is indeed precisely what Oswald and Plonsky (2010) found in their survey of 27 meta-analyses in second language acquisition.

The answers to our two main research questions are drawn from the information in Tables 26.1 and 26.2, which show the authors and year of publication in the first column, followed by the essential research focus in simplified form, and then the basic data necessary to calculate the effect size (number of participants, means, standard deviations and pooled standard deviations) for the 21 studies. At the bottom is the combined effect size along with its standard deviation, and the 95 percent confidence intervals.

The mean gain effect size as shown in Table 26.1 is 1.68 standard deviation units (with its own standard deviation (or SD *d*) of .84, and a reasonably narrow 95 percent confidence interval of 1.36–2.00 (note too that this does not contain 0). This is extremely high even by Oswald and Plonsky's (2010) more exacting limits (strong ≥ 1.0), showing that corpora can be effective in the sense that the results are significantly higher following treatment (see Research Question 1). For Table 26.2, the effect size is predictably somewhat lower at 1.04 (SD *d* = .73). However, it is still well within the confidence limits (.83–1.25) and can be characterized as "very strong" by conventional estimates, showing that corpus-based learning is more *efficient* than traditional treatments (see Research Question 2).

These effect sizes of 1.69 and 1.04 compare favorably with Norris and Ortega's (2000) average effect size of .96 (SD *d* = .87; CI = .78–1.14) for focused or explicit L2 instruction, over unfocused or minimally focused instruction. They also compare favorably to Spada and Tomita's (2010) effect sizes of .86 (SE = .14) for the effect of explicit instruction on complex grammatical constructions, and .63 (SD *d* = .11) for simpler constructions. And they compare particularly favorably with Grgurović *et al.*'s (2013) average effect size for the efficiency of CALL (computer-assisted language learning) over non-CALL of .35 within groups and .24 between groups. In other words, research evidence is stronger for using corpora in language teaching and learning than it is for explicit instruction or for use of computers in language learning.

Table 26.1 *Within-groups effect size (k = 8), sorted by effect size*

			Pre-test		Post-test			
Study	Research question	n	Mean	SD	Mean	SD	Pooled SD	Cohen's d
Chang & Sun 2009	Does scaffolded corpus work improve proof-reading performance?	13	56.15	16.35	91.54	8.26	12.95	2.73
Chan & Liou 2005	Can a bilingual concordancer assist learning of verb–noun collocations?	32	10.59	3.26	19.53	3.95	3.62	2.47
Lin 2008	Does corpus work increase the accuracy rate of academic vocabulary in writing?	25	68.88	3.57	80.64	5.72	4.77	2.47
Moreno Jaén 2010	Do corpus materials improve learners' collocational knowledge?	21	48.02	11.18	67.97	11.31	11.25	1.77
Chang 2012	Does corpus work help improve use of stance and move in writing?	7	8.43	3.79	13.57	2.08	3.06	1.68
Cobb 1997a	Does corpus work lead to vocabulary gains?	11	63.90	14.80	75.90	7.10	11.61	1.03
Huang & Liou 2007	Does corpus work help receptive and productive vocabulary learning?	38	39.00	17.13	49.50	15.41	16.29	0.64
Liou *et al.* 2006	Does corpus work lead to vocabulary gains?	38	39.00	17.13	49.50	15.41	16.29	0.64
					Effect size (mean gain)			**1.68**
					SD *d*			**0.84**
					95% CI lower			**1.36**
					95% CI upper			**2.00**

Table 26.2 *Between-groups effect size (k = 13), sorted by effect size*

		Control group			Experimental group				
Study	Research question	n	Mean	SD	n	Mean	SD	Pooled SD	Cohen's d
Stevens 1991[a]	Do multiple concordance lines aid comprehension?	20	49.10	15.00	20	90.00	18.60	16.90	2.42
Supatranont 2005[b]	Does corpus work lead to vocabulary gains?	50	27.46		50	20.04			2.00
Gordani 2012	Does autonomous corpus work have a positive effect on lexical knowledge?	35	17.77	3.37	35	22.97	3.22	3.30	1.58
Johns *et al.* 2008	Does concordancing with a novel lead to improved reading ability?	11	60.00	16.73	11	83.64	13.62	15.25	1.55
Cobb 1999b	Does corpus work lead to vocabulary gains?	12	70.75	12.35	12	86.83	8.90	10.76	1.49
Gan *et al.* 1996	Does corpus work help vocabulary skills development?	24	9.36	2.55	24	13.04	3.16	2.87	1.28
Sun & Wang 2003	Does corpus work help with learning collocation patterns?	40	48.50	21.25	41	65.00	24.57	22.97	0.72
Tian 2005	Is corpus work an effective approach for grammar?	48	67.39	27.13	50	80.52	12.20	21.03	0.62
Rapti 2010	Does corpus work impact the teaching and learning of grammar?	14	48.29	28.59	14	60.89	21.74	25.40	0.50
Boulton 2011	Does corpus work help with noticing skills?	25	18.84	3.78	34	20.50	3.37	3.58	0.46
Kaur & Hegelheimer 2005	Does corpus work lead to correct use of new vocabulary in writing?	9	44.22	12.94	9	49.00	12.12	12.54	0.38
Boulton 2010a[a]	Can learners use corpus worksheets to help with problem items?	62	5.68	1.70	62	6.39	2.10	1.91	0.37
Sripicharn 2003[c]	Does corpus work transfer to new tasks?	22	25.00	5.24	48	25.36	3.80	4.58	0.08
					Effect size (mean difference)				**1.04**
					SD *d*				**0.73**
					95% CI lower				**0.83**
					95% CI upper				**1.25**

[a]Control provided by within-group design.
[b]SDs not given, ES calculated by researchers from original data.
[c]Data combined from two studies.

3.3 Discussion

The overall effect sizes reported here of 1.68 (within subjects) and 1.04 (between subjects) is respectable in educational terms, suggesting not only that corpora can be effective but that they can be efficient compared to other treatments. In other words, the answers to both our research questions (Is corpus use *effective* for L2 learners – i.e. does it have a demonstrable effect? Is corpus use *efficient* for L2 learners – i.e. compared to other forms of learning?) are clearly Yes and Yes, based on the studies available to date. Given the broad sweep of focus in the various primary studies, it seems that corpora can be of benefit to L2 users for a range of purposes: learning and use of language anywhere on the lexico-grammatical continuum (including collocation and idiom) for both receptive and productive purposes, as well as in more extensive reading and writing tasks or in translation. It seems particularly appropriate in the usual problem areas that feature prominently in these studies (i.e. where conventional transmission-based teaching has been found ineffective). It can be useful in both controlled, paper-based work and in more autonomous, hands-on concordancing, and can be suited to both general and specific purposes. The evidence suggests that corpus work is now ready to expand beyond the university ESP class, where it has largely been used to date, into mainstream second and foreign language learning – where, of course, its effects can continue to be investigated and the conditions of its success elaborated.

Yet, inevitably, a note of caution must be added. Attaching a single figure to a meta-analysis helps to make sense of a body of research with limited risk of bias or subjectivity, provides a convenient yardstick by which to gauge individual studies past and future, and may be politically expedient for attracting interest to the area (see Grgurović *et al.* 2013: 2). On the downside, it may lead some to suppose that this is the final word, and that no future research is necessary. However, quite the opposite is the case (Norris and Ortega 2006: 10–11).

Firstly, in the meta-analysis presented here, we have attempted only a preliminary study, and further work would be required to come to more reliable conclusions. In particular, it is essential to note the variation within the studies, which by no means all produce the same results: the details are as important as the major findings (Ellis 2006: 308). A "wish list" for a fully fledged meta-analysis would include a more principled and extensive trawl of papers from databases and other journals, as well as more fugitive literature; better coding for each paper to see more easily what they have (or do not have) in common, and developing this for more rigorous inclusion/exclusion criteria; weighting the studies according to their design; combining effect sizes where more than one is provided in a given study, and allowing more than one effect from the same study where the population samples are different; teasing out more data from studies

which include t-scores or F-scores, for example. Graphic displays should further help to visualize the variation in effect sizes between individual studies, and maybe suggest leads as to what the biggest effect sizes have in common and, conversely, what subvariables are most worth following up.

Finally, and perhaps most importantly, our aim here is to suggest avenues for future work. This includes areas that are underrepresented at the present time. First, in terms of research focus, we would hope the future would bring more discourse-level studies with a focus on text and associated features of genre, stance, etc., to complement the current dominance of studies on lexis and specific grammar points. It will be interesting to see what multimodal or multimedia corpora can bring to the table, and their impact on speaking and listening skills. The ways corpora are used and integrated are also in need of further study: how do controlled, teacher-led corpus tasks compare with the type of more serendipitous, independent hands-on corpus work traditionally associated with Johns' data-driven learning? And how do these relate to learner profiles (such as motivations, styles, or levels of proficiency), i.e. are there some learners for whom corpus work is more or less suitable? Perhaps most strikingly in need of study are the longer-term or secondary effects of regular concordance work on language awareness and sensitivity, autonomy, motivation, noticing, and other cognitive and metacognitive skills, and so on; their virtual absence in the studies covered here is no doubt due in large measure to the difficulty of assessing such features over time.

Secondly, in terms of study design, we would hope for more longitudinal studies with delayed post-tests to balance the short-term focus on very specific target items often found in the work reviewed here. We would strongly encourage the authors of studies to publish their results whatever the outcomes, as experience suggests that many conference presentations in particular are subject to the "file-drawer" problem where they elicit undesired or non-significant results – of all the studies included here, only Boulton (2011) admits to not showing a significant *p*-value. And we would very much hope that empirical research will become steadily more rigorous, with the use of true control or comparison groups, more regular reporting of the essential meta-data (even L1, group size, duration etc. are missing on occasion), descriptive statistics (means and especially standard deviations), and more extensive use of inferential statistics. Indeed, it has become traditional to conclude works of meta-analysis with a scolding about sloppy research and an exhortation to do better in future (e.g. Norris and Ortega 2000: 497–498), at least in research domains that have been long established and should have more to show for a large amount of effort expended and a large number of studies published. In a newish domain such as ours, a nudge for more, more differentiated, and yet also more replicated, and in all cases better-reported studies is probably sufficient.

4 Conclusion

Corpora have found many uses in the field of language teaching and learning in the hands of decision-makers, teachers, and learners. Published research covers classroom applications for a wide variety of learner profiles and for extremely different uses, from highly controlled to entirely autonomous work, from paper-based materials to hands-on concordancing, from reference resource to learning tool. This variety underlines the highly flexible role of corpora – there is no single "right" way to use them. From a research perspective, this may lead to a perceived fragmentation of the field, which a thorough meta-analysis may go some way to resolving.

The meta-analysis as a research form is by definition exploratory rather than confirmatory, starting from questions (to be explored) rather than hypotheses (to be confirmed or denied). Of course, few researchers, meta-analysts or otherwise, would deny hoping that their questions would be answered in a certain way, and take steps to ensure objectivity. In the survey presented here, we were gratified to uncover a measure of confirmation from research to date that corpora have been not only effective in language teaching and learning, but also efficient, insofar as they produce fairly regular advantages of a standard deviation or more over other methods of achieving the same goals. Our meta-analysis is only exploratory; further work will be needed to exploit current research fully, especially in exploring the mediator variables that are likely to be worth investigating.

The synthesis presented in this chapter has shown that there is more research in the area than sometimes claimed, but of highly varying rigor both for qualitative and especially quantitative studies. Further, the questions addressed, though varied, tend towards the short-term and experimental with a focus on specific language items; more longitudinal, ecological, open-ended studies are needed, especially addressing the alleged benefits of corpus work in promoting learning to learn and, consequently, in producing "better learners".

A final word. Traditional corpus consultation is in some ways a relatively marginal activity, to be found in few classrooms around the world. However, it is in many ways analogous with internet searches and use of other technologies for querying the vast stores of data available, which has arguably become the dominant learning mode in our culture. Learners regularly Google up internet-as-corpus data to help with collocations, grammar choices, and many other matters, particularly in their writing (see Boulton, forthcoming). Indeed "Googling" is largely an invention of corpus linguists (Crystal 2012) and the majority of internet users are busy becoming knowledge co-constructors from corpus data. This, of course, is definitely *not* to say that all search-based learning is accurate, permanent or worthwhile – far from it – in language learning or any other area. That is

why research is needed to show us how to take best advantage. How much training is needed? How much ongoing scaffolding? Are certain learning or personal styles favored or disfavored? How is the success of such learning best measured? What is the ideal complementarity between search-based and other forms of instruction? We now see that these questions are central rather than peripheral to language learning; and in our meta-analysis we have seen that ways of answering them are under development.