Unit 16 Language teaching and learning

16.1 Introduction

While the usefulness of corpora in language pedagogy is an area of ongoing debate (see unit 12.4), we noted in unit 10.8 that there has been an increase in interest in recent years in the corpus-based approach to language teaching and learning. This unit uses three excerpts from published papers to discuss the theoretical and practical issues related to using corpora in language education, focusing on English as a second or foreign language.

16.2 Gavioli and Aston (2001)

We noted in unit 12.4 that the use of corpora in language pedagogy is a topic causing ongoing debate. Among those who support using corpus data in language teaching there has been considerable discussion of how far teaching syllabuses and materials should be 'corpus-driven'. Gavioli and Aston (2001) summarize this debate and argue that corpora should not only be viewed as resources which help teachers to decide what to teach, they should also be viewed as resources from which learners may learn directly. This excerpt is cited the first four sections from their paper.

Gavioli, L. and Aston, G. 2001. 'Enriching reality: language corpora in language pedagogy'. *ELT* Journal 55/3: 238-246.

Introduction

Ever since the *Cobuild* project started producing corpus-based dictionaries, grammars, and materials for ELT, applied linguists have been divided between those who have seen the findings and methods of corpus linguistics as providing new ways forward in language teaching, and those who have warned against over-enthusiasm. In a debate in *ELT Journal* 52/1 (1998), Ron Carter and Guy Cook focused on two of the main terms of this argument. First, they asked how far the analysis of corpora provides descriptions of the workings of 'real English'; second, they asked whether such English is what foreign learners need. In this paper, we begin by summarizing our own position on these issues, and then go on to argue that the terms of this debate should be redefined in relation to learners' needs to experience language as 'real' for themselves.

Can corpora capture reality?

The largest corpora of English are still smaller than the average adult user's experience of the language, and very different in their composition (most notably in the ratio of speech to writing). Nonetheless, they provide evidence about linguistic performance which can undoubtedly be helpful in deciding what we should teach.

First, they can be used to test claims based purely on intuition. For instance, Carter (1998: 43) proposes that the word 'real' 'invariably carries positive associations', as in 'Real ale', 'Get real!', 'Real English', 'that real country taste', etc. But if we look up 'real' in the 100-million word British National Corpus (BNC), the picture is rather different. We find that the most frequent lexical items to collocate with 'real' are 'world(s)', life/lives', 'term(s)', and 'problem(s)'. Can we really say that 'the real world', 'real life', 'in real terms', or, most strikingly, 'a real problem', have positive associations? The corpus evidence makes it clear that the linguistic 'fact' Carter proposes is over-generalized, and suggests that we might want to reformulate it somewhat for teaching purposes. Second, corpora can help clarify our motives for teaching particular features. Carter discusses a number of spoken formulae which carry 'cultural' content, including expressions referring to other nations: 'Dutch courage', 'to

go Dutch', 'double-Dutch', 'Dutch cap', 'Dutch auction', 'then I'm a Dutchman', etc. He goes on to claim:

> Here we learn several useful and widely used phrases, but we can also learn something about British insularity, and that distrust of foreigners to the point where the British can be interpreted as believing almost all of them to be either unintelligible, untrustworthy, or 'unreal'. (Carter 1998:49)

Judging from publicly-available corpora of speech, however, these expressions are far from 'widely used'. In the spoken component of the BNC (10 million words), 'go Dutch' and 'double-Dutch' each occur twice, 'I'm a Dutchman' once, while 'Dutch cap', 'Dutch courage', and 'Dutch auction' are not found at all. A similarly-sized British speech component of the Bank of English presents an even more desolate picture: there are four instances of 'double-Dutch', and that is all. Given these very low frequencies, these items would only seem worth teaching if we have other good reasons for doing so. In the passage cited, Carter actually suggests two such reasons: 'double-Dutch, 'go Dutch', and indeed, 'Dutch cap', could all be useful expressions for a learner wishing to avoid social embarrassment in Britain; and the study of British insularity, as revealed through linguistic references to foreign nationals and nations, could constitute a stimulating activity which could increase learners' awareness of cultural issues.

The inclusion in syllabuses of language which is very rare in large corpora thus calls for justification, and the same is equally true for the exclusion of language which is common. As we saw with 'real', corpora can remind us of frequent uses which might otherwise tend to be ignored. Thus McCarthy and Carter (1995) notice the frequency in speech of the semi-modal 'tend to' (it occurs almost as often as 'ought' in the BNC spoken component). Although this verb has traditionally received little attention in teaching, it arguably provides learners with a valid alternative to frequency adverbs such as 'usually' and 'often'. A more problematic case noted by the same authors is that of structures with 'tails' (as in 'That's enough, *don't you think?*'), which are rarely found among the prototypical patterns presented by textbooks. While their frequency in conversation suggests they should be included in syllabuses, other considerations may argue against this, at any rate from the perspective of spoken production. Their use being highly context-dependent, they seem difficult to teach and harder to master than other markers of affect with similar functions. The point is that while corpora do not tell us what we should teach, they can help us make better-informed decisions, and oblige us to motivate those decisions more carefully.

Can corpora provide valid models for learners?

Most existing corpora are collections of spoken and/or written texts produced by native speakers. Both Cook (1998) and Carter (1998) ask whether learners in fact need to imitate native-speaker behaviour, and whether, in consequence, corpus data are relevant to them as models — a doubt, incidentally, which relates not only to corpora, but to 'authentic' materials in general. There is, however, no reason to assume that the materials we present to learners should constitute models for imitation (were this the case, it would be difficult to imagine a role for literature, advertising, or other 'creative' genres in the language classroom), and it would be wrong to expect corpus data to do so either. When linguists abstract generalized patterns from corpora, and interpret the data as exemplifying them, these patterns are rarely immediately apparent. Sinclair, who attempted to include only actual corpus instances as examples in the *Cobuild* dictionary (Sinclair 1995), reports how difficult it was to find instances which reflected 'typical' usage in every respect (Sinclair and Kirby 1990: 114–15). The *Cobuild* team was, moreover, only looking for single sentences exemplifying a limited range of features. The chances of finding a complete corpus text which consistently shows typical usage is minimal, so if we want to propose a model of conversation at the hairdresser's, we will almost certainly do better to use an invented dialogue than a corpus extract — though we may want to compare it with corpus extracts before proposing it to students.

It is precisely because they do not simply offer models to imitate, however, that corpus data seems valuable for learners. As Leech and Candlin (1986: xiv–xvi) observed well over a decade ago, data from corpora has to be interpreted subjectively. Their reality, from this point

of view, is a characteristic not just of the data, but above all of the interpretative process. For learners, the reality of corpus data would seem principally to lie in the extent to which they can interpret them to create models of their own.

From real texts to real discourse

In one of his most widely-quoted distinctions, Widdowson (1978) contrasts *genuineness* (a quality of texts) and *authenticity* (a quality of discourse interpretation). Viewed in these terms, corpora of naturally-occurring texts provide samples of genuine language, since they are produced by speakers and writers with real communicative goals. The reproduction of such samples in pedagogic contexts does not, however, guarantee them authenticity as discourse, which depends on their context of reception. In their discussion, both Carter (1998) and Cook (1998) seem to treat reality as an inherent characteristic of materials, i.e. as a matter of genuineness of the text. But if — as communicative language teaching has traditionally held — learning is primarily a product of discourse authenticity, the question is not whether corpora represent reality, but rather whether their use can create conditions that will enable learners to engage in real discourse, authenticating it on their terms — and whether this engagement can lead to language learning.

Widdowson (1998) claims that learners will often be unable to authenticate real texts, since they do not belong to the community for which those texts are designed, and are therefore unqualified to participate in the discourse process. This, however, overlooks the fact that there is an alternative way of authenticating discourse, by adopting the role of an observer (Aston 1988). While the participant interacts with the text as an intended recipient, the observer views this interaction from the outside, adopting a critical, analytic perspective. Observer as well as participant roles can allow learning: observation allows strategies of interaction to be noticed, while participation allows such strategies to be tested.

Corpora clearly allow many opportunities to authenticate discourse through observation. As already noted, unlike the examples provided by textbooks and dictionaries, the samples of language provided by corpus data do not immediately illustrate particular linguistic patterns. A concordance does not make sense in itself: sense has to be attributed to it by the reader, who must infer patterns which will as far as possible account for the data. In other words, a concordance can be viewed as a text that provokes 'a pragmatic reaction' in the observer (Widdowson 1998: 713). As we shall see in the following examples, this pragmatic reaction can also constitute a focus for discourse participation, thereby allowing learners to alternate and integrate these two roles.

16.3 Thurstun and Candlin (1998)

Thurstun and Candlin (1998) explore ways of using concordancing in teaching the vocabulary of academic English. In this excerpt, the authors discuss the rationale for their decision to focus in detail on a restricted set of vocabulary items.

Thurstun, J. and Candlin, C. 1998. 'Concordancing and the teaching of the vocabulary of academic English'. *English for Specific Purposes* 17: 267-280.

Introduction

Since concordancing programs have become available to teachers and students, their possibilities have been seen as offering new and exciting directions for developing teaching materials, enabling students themselves to make direct discoveries about language (Johns 1991a; Tribble & Jones 1990) and as an aid to course design (Flowerdew 1993). This particular project has used the concordancing program, *Microconcord*, and the *Microconcord Corpus of Academic Texts* (1993) to develop teaching materials for independent study of the vocabulary of academic English.

It was decided to develop corpus-based learning materials that would be of assistance to students from any discipline, focusing on lexical items shared across various disciplines. Li & Pemberton (1994: 184) point out that tertiary students do not necessarily find discipline-specific technical vocabulary difficult:

Rather, it is the vocabulary with a middle frequency of occurrence across texts of various disciplines that students find most problematic.

Nation (1990) refers to this range of vocabulary items as "academic vocabulary".

Rationale

The features of this project which require preliminary explanation are firstly, the decision to focus in detail on a restricted set of vocabulary items, and secondly the use of concordancing techniques to provide the student with intensive exposure to the use of these items.

In choosing the vocabulary items to be dealt with in this project, we began with the extensive University Word List cited in Nation (1990). Using this list, we developed categories of vocabulary items according to the various rhetorical purposes they can serve in academic writing. We then selected items according to frequency of use cited by Nation and our own perception (based on having marked many hundreds of student essays and in consultation with teachers of English for Academic Purposes) of the extent to which their investigation would be helpful to students. In this way, we created a list of about 150 examples (see Appendix A), attempting to group them according to purpose. We will refer to these purposes as rhetorical functions. Further selection was based on frequency counts provided by *Microconcord*, using the *Microconcord Corpus of Academic Texts*. The main rhetorical functions we identified, and the key words for each function on which the project finally focused were:

- Stating the topic of your writing
factor
issue
concept
- Referring to the research literature
evidence
research
source
- Reporting the research of others
according to
suggest
claim
- Expressing opinions tentatively
may
possible
unlikely
probably
- Explaining procedures undertaken in a study
identification
analysis
criteria
- Drawing conclusions
conclude
summarize
it is clear
thus

Such a focus on particular rhetorical functions provides a purposeful basis for learning and potentially some structuring for the teaching of academic writing skills. The intensive focus on a limited number of vocabulary items is characteristic of concordance-based materials and is supported by the experience of Tim Johns (University of Birmingham, UK, personal communication). As with Li & Pemberton (1994), students, in his view, do not necessarily need to master a wide range of academic terms in order to write acceptable academic essays. They do, however, need to be competent users of a restricted set of "semi technical"

vocabulary items. Pickard (1994: 218) suggested that, when preparing concordance-based material, there is

> a potential tension between the aims of wanting to expose students to a variety of vocabulary as a means of encouraging variety in their writing, and using a concordancer to search for a key word. Through selection of key, frequent words how does one encourage variety?

Variety, however, is not necessarily an end to be pursued for its own sake in the teaching of academic writing, and Pickard goes on to point out that, by drawing students' attention to collocates of the key word, concordance-based study has considerable potential for expanding student vocabulary while dealing in detail with selected items.

Academics from four Australian universities responded in a recent study (Bush et al. 1996) to questions about their expectations of student writing. Results indicated that accurate and appropriate use of academic vocabulary is considered to be extremely important, but there is much more concern that students convey their ideas clearly than that they attempt to rely on jargon.

These comments and studies supported our view that the most useful approach to helping students unfamiliar with the vocabulary required in academic writing would be to focus on intensive work on a few of the most useful lexical items selected as typically realising each of the rhetorical functions listed.

These three or four selected words for each rhetorical function are all frequently used, most appearing more than once every 6000 words in the corpus. Words such as *unlikely* and forms associated with *summary* are not used quite so frequently in the corpus of professional, published work, (appearing with a frequency of 1/12600 and 1/13254 respectively) but were nonetheless included given their usefulness for student writers, particularly for dealing with modality and the creation of final statements. A range of grammatical forms of each word in question was included where appropriate.

Concordance-based materials offer the learner a rich experience of language (in this case, the language of academic English). Concordancing has been used in this project to present students with the opportunity to condense and intensify the process of learning through exposure to multiple examples of the same vocabulary item in context, and to promote awareness of collocational relationships. According to Nattinger (1988: 63),

> guessing vocabulary in context is the most frequent way we discover the meaning of new words.

Johns (1991b) has argued that the central justification for using concordance-based materials is that they can help to develop this ability to guess the meaning and use of unknown words from context. In our materials, exposure to concordances for the purpose of discovering meaning focuses learner attention on the central importance of collocational relationships in connection with the key words. The broad objective of the materials is to develop the writing competence of students by promoting discovery of meaning and by making students aware of representative patterns of language use and of selected grammatical structures. This awareness is accompanied by guided opportunities for research, practice and improvisation.

At various stages in the preparation of this material, as with Flowerdew (1993: 240) who found "areas where concordancing has revealed a discrepancy between published materials and the specialist corpus," we also encountered language use which questioned the standard patterns usually recommended to EAP students by teachers and grammar books. For example, it was found that the plural form, *researches*, occurred in the corpus on ten occasions, indicating that it is, in fact, accepted practice in published texts though usually not accepted by markers of student essays. Although we did not draw student attention to this specifically, we decided to modify questions about the use of this form so that students were not obliged to understand that it is never used in its plural form. Likewise, the standard advice to students to use the past tense of the reporting verb for author-prominent statements (see Weissberg & Buker 1990: 45, who advise students that "in these citations the *simple past tense* is used in the verb of report") was not supported by the corpus, these reporting verbs being shown to be used more frequently in the present tense.

16.4 Conrad (1999)

Conrad (1999) presents a corpus-based study of linking adverbials (e.g. *therefore* and *in other words*), on the basis of which she suggests that it is important that a language teacher do more than using classroom concordancing and lexical or lexico-grammatical analyses if language teaching is to take full advantage of the corpus-based approach. This excerpt is taken from section 1 of her paper.

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

Interest in the use of language corpora and computer analysis tools for language education has grown tremendously in the past decade. Understandably, articles written for language teachers have emphasized the use of corpora and computers in the classroom. The greatest attention has been paid to the use of concordancers — software programs whose primary purpose is to display words or simple grammatical items with their surrounding context. Writers emphasize the usefulness of concordancing for vocabulary and grammar development because it facilitates the use of authentic language, makes students more active and independent analyzers of language, and provides empirical evidence about language use (Johns, 1986, 1994; Taylor, 1991; Hanson-Smith, 1993; Aston, 1995; Stevens, 1995; Qiao and Sussex, 1996; Cobb, 1997). In addition, articles about corpora and computer-assisted analyses have also addressed the development of cloze tests (Coniam, 1997) and of test construction in general (Alderson, 1996). Textbooks also are beginning to integrate the use of corpora; e.g. a recent textbook for spoken English (Carter and McCarthy, 1997) is designed around extracts from a spoken language corpus.

There is no denying that these publications provide important information for language teachers. However, teachers may get the impression that these articles represent the only ways that language corpora can be useful to them. On the contrary, the growing field of corpus linguistics offers much more for teachers who want to understand language use and design effective materials for their students. In fact, even if teachers do not have computer expertise or computer facilities in their schools, corpus-based studies can be valuable resources for them.

The previous corpus-based work addressed to teachers has commonly been constrained in two ways. First, most of the articles for teachers stick to small-scale analyses. Usually, a small collection of texts — often compiled by convenience rather than following a principled design — is used, and analyses typically consist of looking at all the occurrences of a certain word or reading a transcript from the corpus. These analyses do provide interesting information, but many of the most useful aspects of corpus-based research are lost. For example, studies with larger, more diverse corpora can make comparisons in the characteristics across varieties of language (e.g. spoken conversation vs academic articles, or research articles in one discipline vs other disciplines), and thus better meet the needs of students or teachers in special purpose situations. Studies that include statistical analyses or even frequency data can identify strong patterns in language use that we do not recognize intuitively — patterns that may be very helpful to discuss with our students.

In addition, previous studies aimed at teachers rarely go beyond lexical or lexicogrammatical analyses; i.e. they concentrate on studies of words alone or words in connection with a grammatical feature (e.g. verbs occurring before *that* complement clauses) and do not undertake more complex grammatical analyses. Even one of the best-known publications covering analysis of a large corpus (Sinclair, 1991) focuses on lexical and lexico-grammatical analyses. In general, therefore, the fact that corpus-based studies can also provide insight into complex grammatical and discourse features remains unknown to many teachers. Even less well known is that corpus-based studies can address the interactions of many aspects of a grammatical feature: e.g. the frequency, syntactic forms of the feature, typical lexical items realizing the feature, and differences in use across language varieties can

all be analyzed together — and can all be tied to the communicative functions fulfilled by the feature.

In sum, the limited types of studies that have been addressed to teachers have run the risk of restricting teachers' appreciation of corpus linguistics. However, corpus linguistics is a means of studying and describing language use which offers a great deal beyond classroom concordancing and lexical/lexico-grammatical analyses. Practising teachers and teachers-in-training can learn a great deal from corpus-based studies and, in fact, owe it to their students to share the insights into language use that corpus linguistics provides.

16.5 Unit summary and looking ahead

In this final unit of Section B, we discussed the theoretical and practical issues of the corpus-based approach to language teaching and learning, using three excerpts from published material. The discussion not only showed that corpora are a valuable resource for language education, it also warned readers of potential problems of using corpora in language pedagogy. We will return the learning issue in case study 3 in Section C, where readers will be engaged in an interlanguage analysis on the basis of the Longman Learners' Corpus.

Section C, the final Section of this book, will explore some of the areas in language studies which we have introduced in Section A and further discussed in this Section B.