The uses of reality: a reply to Ronald Carter

Guy Cook

Introduction

Computerized language corpora have inspired some of the most important insights in recent linguistics. They have shown us, for example, that actual language use is less a matter of combining abstract grammar rules with individual lexical items, and more a matter of collocation; that there are grammatically possible utterances which do not occur, and others which occur with disproportionate frequency; that in systematic descriptions of occurrences, grammar and lexis cannot be as easily separated as they have been traditionally, either in pedagogy or in linguistics. Ronald Carter is right to find such insights 'exciting', and his own work with Michael McCarthy on the CANCODE corpus, has added to them. As his article illustrates very well, the grammatical constructions we find in actual conversations are not always accounted for in traditional grammars.

Clearly all these findings are important, and they do have implications for language teaching. The problem is, however, that some corpus linguists (e.g. Sinclair 1991, Stubbs 1996) overreach themselves. They talk as though the entire study of language can be replaced by the study of their collections, and as though all important insights will emerge only from automatic searches of their data and nowhere else. Clearly such solutions to the study of complex human phenomena exert a good deal of seductive power. If the traditional concern of linguistics—language in all its cultural and psychological complexity—could be replaced by a neat computer bank of data, life would be much simpler.

Yet the leap from linguistics to pedagogy is—as Carter realizes—far from straightforward. He is not one of the extremists, and his paper is, for that reason, a worthwhile and interesting contribution to language teaching. He proceeds cautiously, providing some interesting 'real' data, and pointing out significant differences between actual and textbook English. He does not say one should replace the other. In his view, materials should be influenced by, but not slaves to, corpus findings. (In this he seems to agree with the view of Summers and Rundell (1995) that pedagogic materials should be 'corpus based not corpus bound', and to disagree with the COBUILD slogan that they should be 'corpus driven' (Stubbs 1997).) This is eminently reasonable, though for that very reason not particularly radical. My problem with what Carter says is that he seems a little hesitant—or perhaps unwilling—to say where he stands. Does he reject the fundamentalist views of those linguists and language teaching theorists for whom corpus findings are the only source of truth?
My first aim in this reply is to pursue some of the shortcomings of corpus-driven approaches which I think Carter avoids confronting. I shall also consider some of the more extreme applications of corpus findings to language teaching. My argument is that there is an important difference between the hard and soft line approaches, that the former, by appearing to offer yet another easy 'scientific' solution, can do immense damage, and that we all, including Carter, would do well to consider more precisely whether we think corpus findings merely add a new dimension to earlier approaches, or replace them.

**Uses and abuses of corpora**

A number of false conclusions can be reached about corpora. It is often assumed, for example, that as a description of language behaviour, they are the only valid source of facts about language; the same as a description of language in the mind; provide a goal and a route for language learning. There is much in computerized corpus analysis to make us reconsider received ideas about the learning, representation, and use of language. But where pedagogy is concerned, corpus statistics say nothing about immeasurable but crucial factors such as students’ and teachers’ attitudes and expectations, the personal relationships between them, their own wishes, or the diversity of traditions from which they come. Consequently computer corpora—while impressive and interesting records of certain aspects of language use—can never be more than a contribution to our understanding of effective language teaching.

**Corpus as fact**

Even as a record of ‘facts’ computer corpora are incomplete. They contain information about production but not about reception. They say nothing about how many people have read or heard a text or utterance, or how many times. Thus a memo hastily skimmed by one person and consigned to the wastepaper basket counts equally with a tabloid headline read by millions, or with a text, such as a prayer or poem, which is not only often repeated but also deeply valued. Occurrence, distribution, and importance, in other words, are not the same. This applies to whole texts, but also to shorter units. Some phrases pass unnoticed precisely because of their frequency, others strike and stay in the mind, though they may occur only once. And because different individuals notice different things, such saliency can never be included in a corpus. The same is true of a whole host of aspects of language use: metaphors, speech acts such as apologies or compliments, interactive events such as interruption or awkward uses, levels of formality. They are not ‘facts’ but matters of varying perception. It is a truism to observe that there is no straightforward correlation between the words people use, the intentions they had in them, and the interpretations which other people put upon them. If this were not so, there would be no disputes over the meaning of what people say.

**Corpus as record**

Corpora are records of language behaviour. The patterns which emerge in that behaviour do not necessarily and directly tell us how people organize and classify language in their own minds and for their own use, or how language is best systematized for teaching. Linguists’ analyses of
these data are not necessarily users' analyses, or those which are most useful to teachers and learners. They are just one kind of fact. The ways in which grammarians and pedagogues have organized their material—in grammars, syllabuses, and dictionaries—are also facts about language. So are people's emotional beliefs that one type of language use is better than another. We should not promote some kinds of facts at the expense of others.

Corpora are only partial authorities. The cumulative language experience of an individual, though less amenable to systematic access, remains far larger and richer. Even a three hundred million word corpus is equivalent to only around three thousand books, or perhaps the language experience of a teenager. This is why our intuition (in effect our random and incomplete access to our total experience of the language) can still tell us facts about the language which can not be evidenced by a corpus (Widdowson 1990). For example, the canonical forms of sayings and proverbs occur very rarely in corpora, though they are obviously well known by people (Aston 1995). Such omissions, however, are not merely a quantitative issue; they cannot be remedied simply by making corpora larger and larger. They are inevitable in an approach which accepts only one of the three sources of fact about language: observation; and ignores or villainizes two others: introspection and elicitation. For there are aspects of language which are known but not used. Corpus linguists are fond of observing that the commonest uses of words are not the same as their standard definitions. 'I bet' for example, is more rarely used in the sense of 'wager', and most often in the sense of 'suppose' (Sinclair 1987: xvi). But this unsurprising observation does not at all invalidate the view that 'wager' is a central prototypical meaning for many speakers to which more colloquial uses are attached. (And indeed, the 'wager' meaning is still given as the first meaning of 'bet' in the Cobuild dictionary.)

**Description and prescription**

But let us assume for the sake of argument that corpora are accurate records of language behaviour, that they do catalogue and reveal all the important 'facts' about the language. The question then arises as to whose language behaviour is accurately recorded—and the question takes on a particularly sinister significance when the corpora in question start being used not as data for descriptive linguistics, but as sources of prescription for TESOL. For the answer to the question is (as Carter seems painfully aware) that corpora are primarily records of native speakers' language behaviour. 'Real' language in effect means native-speaker English, and the only language excluded from this category (apart from the invented examples of linguists and textbook writers) is that used to and by language learners. To his credit, Carter confronts this issue, and intends to remedy it. But the proposed addition of 'a wider variety of international Englishes' (see Note 2, page 54) will not solve the problem. This will only add other standard Englishes as spoken by their own native speakers.

*The uses of reality: A reply to Ronald Carter*
And then a second question arises to which Carter explicitly refers, but does not answer. Why should the attested language use of a native-speaker community be a model for learners of English as an international language? If a certain collocation occurs frequently among British or American English speakers, must it also be used by the Japanese or the Mexicans? This is where we encounter an easy slippage from description to prescription, in effect making the former into the latter. The English which is used by one or more native-speaker communities, it is implied, ought to be the English learned for international communication.

The ready-made lexical phrases which corpora reveal to be so frequent in native speaker use are moreover—as Carter readily recognizes—very often culturally specific and loaded. In deploying such units, the foreign speaker is very likely to produce corpus-attested but contextually inappropriate language. (This is why attempts to teach set phrases are likely to be as tragically disastrous in lexical syllabuses as they were in functional ones.) Carter's own example of repeated tails ('nice it was', 'nice run that') may well be frequent in data, but successful deployment by a foreign learner could easily go wrong. In the terms of Hymes's (1972) four parameters of communicative competence, corpus-driven language teaching always risks stressing what is actually done at the expense of what is appropriate in a particular context.

**Pedagogical issues**

In an extensively quoted, and in itself excellent, essay by Pawley and Syder (1983) on native-like selection and fluency, corpus-based language teaching finds a source of inspiration, providing a potential link between the facts of language behaviour and a theory of how language is acquired and processed in the mind. Here is the claim that mature native speakers (for this is whom the essay is explicitly about) have 'hundreds of thousands' of institutionalized lexicalized or semi-lexicalized units in memory. Though many of these units can be analysed grammatically, the likelihood is, so the argument goes, that they are often produced and understood holistically. Native speakers acquire, represent, and process language in lexicalized chunks as well as grammar rules and single words.

Yet it by no means follows that foreign learners must do the same. They may not want to study language in this way; they may live within culturally diverse pedagogic traditions not compatible with this approach; they may not aspire to or need native-like English; they may not have as much time available as native-speaker children; above all, as adults with conscious learning strategies available to them, they can choose. And why should they not choose to continue viewing the language as grammar structures and slot-filling words? This may not lead to native-like English, but it may lead to communicative and expressive English. It may be learnt more quickly. And it will avoid the tedious rote learning of mundane phrases, or the bewildering refusal to teach grammar, which are the inevitable consequences of an over-emphasis on 'lexical chunks'.

Guy Cook
Yet even if appearing native-like were accepted as the goal of language learning, it would still not follow that frequency and desirability are the same. There is a hidden irony in the dogma that frequent native-like collocations are the best model to imitate. It is that even within the native-speaker community it is often the infrequent word or expression which is most powerful and most communicatively effective, and therefore most sought after. This is also why foreigners’ speech is often expressive and striking. Both for native and non-native speakers there is an alternative goal to seeking the most usual, the most frequent or, in short, the most clichéd expression. It is the goal of rich, varied, and original language. Among native speakers it is unusual language which is valued. Should non-native speakers be treated differently?

This leads to the important point that not all types of language are equally valued, either by native speakers or foreign learners. Something is not a good model simply because it occurs frequently. A good deal of actual language use is inarticulate, impoverished, and inexpressive. Inevitably, because one cannot teach everything, part of the job of teachers and course designers is to select the language use which they wish their students to emulate. Many foreign language students have strong feelings about this too. They do not want to learn just any English because it occurs in a corpus, and it is patronising to overrule them. In advocating selection and modelling of corpus data, in the use of literary rather than transcribed dialogues, and in his recognition that one of the topics in his authentic data ('straggly hair') may have a limited topic life in many classrooms, Carter seems to agree.

To be corpus driven, in short, deprives everyone (native and non-native speaker alike) of the opportunity for choice and to make their own impact on the language. Corpora are inevitably records of what has happened rather than what is happening. They present us with a fait accompli, a fixed product rather than an open process.

**Means and ends**

So corpora do not necessarily provide a goal for language learners. Yet even if they did, it would not follow that the best route to this goal is to present real language use, and to try to persuade them to emulate it straight away. Here there is a certain oddity in the corpus argument. Of course expert-speaker use of the language, and the rules which generate it, is usually more complex than that of language learners. If it were not, there would be nothing to learn. Hardly surprisingly, the description of English which emerges from corpus analysis (taking into account as it does the way in which linguistic items and structures vary across genres, social groups, and linguistic contexts) is dauntingly complex and particular. But this description cannot be presented to students all at once. The issue still remains how to simplify and stage the language presented to learners, and to simplify the rules used to explain it, in a way which will enable them to come gradually closer to native speaker use (if that is their goal). Surely the point of grammars and textbooks is that they select, idealize, and simplify the language to make it more accessible? Indeed, this seems to be Carter’s view too.

*The uses of reality: A reply to Ronald Carter*
For language teachers the issue remains as to what the principles for selection, idealization, and simplification should be. Here there is already a wealth of long-standing ideas (dating back at least to the work of Palmer (1921) and West (1926)) concerning the relationship between the frequency with which an item occurs and the point at which it should be taught—ideas of which many corpus linguists, in their haste to advertise themselves as promulgating a totally new approach to language, seem unaware. For example, an item may be frequent but limited in range, or infrequent but useful in a wide range of contexts. Or it may be infrequent but very useful, or appropriate for some pedagogic reason. These are factors beyond mere description. Unlike many corpus linguists, Carter does show himself aware of such considerations in his conclusions. But that leaves me wondering whether his approach is such a break from tradition as he suggests.

**The hard line**

This brings me from Carter’s views—moderate, sensible, and informed—to the more extreme, but unfortunately associated, views of language teaching based on corpus linguistics. Here is the belief that what is perceived as a linguistic revolution necessarily constitutes a pedagogic one. Very often writers are carried away by a single insight into language, taking it illogically to be sufficient to change language teaching. Thus Willis (1990) elevates frequency counts to the guiding principle for his lexical syllabus. Lewis (1993) considers the high occurrence of lexical chunks as a cue to decree (in a diatribe characterized by bombastic assertion rather than reasoned argument) that language teaching has changed forever, to be replaced by ‘the way forward’ (p. 196), with an ominously authoritarian definite article: his own lexical approach. ‘Abstract, absolute knowledge of a system has had its day’, and people who think otherwise ‘are wrong’ (p. 74); ‘woolly mindedness in this matter leads to bad practice which has negative long term effects’ (p. 167).

Such approaches are firmly in the tradition of using linguistics theory to dictate to language teaching practice. Their gross over-generalization and over-confidence are potentially damaging to good teaching practice. They invoke corpus linguistics as an unassailable authority, side-step all serious engagement in debate, and cannot take on board the kind of reservations expressed by Carter. Such corpus-driven pedagogy is a vain attempt to resuscitate a patriarchal attitude to ELT, invoking the latest linguistics theory to intimidate teachers into believing that all previous practice, all their own and their students’ intuitions, all the culturally various pedagogic traditions in which they work and study, are, as Lewis would put it, ‘wrong’.

**Conclusion**

I have contrasted throughout this reply what I see as the soft and the hard line views of the relevance of corpus findings to language teachings. In the one, we have the voice of moderation urging a limited application—‘modelling’ as Carter calls it—which by virtue of its very reasonableness does not amount to anything very radical. In the other,
we have the stronger view: evangelical, authoritarian, and dismissive of tradition, assuming that a little of the latest linguistics theory is all that is needed to change the course of language teaching. I believe that if Carter were to follow his arguments through to their conclusion, he too would explicitly reject, as I do, the more extreme versions both of corpus linguistics and of corpus-driven language teaching. But it is by no means clear whether he does so.

Received June 1997

Notes
1 This point has been made by corpus linguists themselves (Francis 1979, Stubbs 1996:11) but the point is not adequately taken on board, either in corpus construction or analysis.
2 Stubbs (1996: 21) tells us that the 'deep patterning' revealed by corpus analysis is 'beyond human observation and memory'.
3 This issue is clouded by snobbish and chauvinistic claims that a particular national or sociolect is better than another. But this is not a necessary component of the notion that certain usages—literary, written, or simply eloquent and elegant ones—are more desirable models than others.

References

The author
Guy Cook is a Reader and Head of Languages in Education at the London University Institute of Education. He has worked as a school teacher in Egypt, Italy, and the UK, and as a lecturer at the University of Moscow and the University of Leeds. His publications include The Discourse of Advertising (Routledge 1992), Discourse and Literature (Oxford University Press 1994), and (with B. Seidlhofer) Principle and Practice in Applied Linguistics (Oxford University Press 1995) He is currently working on a new book entitled Language Play, Language Learning (Oxford University Press, forthcoming).