Consultant advice **Professor Mike McCarthy**

Corpora and grammar

For several centuries, English grammar reference books and grammar text books for the classroom were dominated by the idea that Latin was the ideal model for grammar: good grammar should be modelled on the great writers of classical Latin. This influence of (a) Latin and (b) written texts on how we thought of grammar persisted right through till the end of the 20th century and is still occasionally heard from politicians and traditional grammarians in the UK and elsewhere.

The advent of corpora changed everything. Thanks to the power of the computer to analyse huge numbers of written and spoken texts, we can see how English grammar has evolved and has its own, independent grammar. We should not try to force English grammar into a Latin mould. Moreover, spoken corpora enable us to see objectively that some grammatical forms are specially designed for interactive speaking and that some of these spoken forms have been neglected in grammar reference books and, as a result, neglected in language teaching. Even some of the most popular grammar books used in EFL/ESL classes are still subtly dominated by written norms.

There is a good deal of overlap between the grammar of speaking and the grammar of writing, and there are no grammar forms that are forbidden in one mode or the other. What is important is that the grammar repertoire is distributed differently in spoken and written language (Leech 2000). For instance, some conjunctions and linking expressions that are very common in formal writing are rare in conversational speaking, and vice-versa. Table 1 shows some differences in frequency of such expressions, based on the Cambridge International Corpus.

Table 1: Frequency of conjunctions and linking expressions (per million words)

	Formal writing	Conversation
therefore	56	26
thus	60	0
cos	9	3122
in the end	57	96
as a result	36	3

For more examples, see Carter and McCarthy (2006: pp. 256–262).

The grammar of writing is primarily concerned with creating coherent texts which can exist independently and which are usually read at a different time and place from when and where they were written. The grammar of speaking, on the other hand, is crucially concerned with creating and maintaining relationships with the listener(s). A good example of this is situational ellipsis, which involves not saying things because they are visible or obvious in the context. This often means that things which we normally consider to be obligatory in the grammar are simply, not present. The example below is an extract from a conversation in the corpus where people are looking at photographs on a computer. In square brackets and greyed out are words which we would normally think of as obligatory in English grammar, but which the speakers did not in fact say.

Speaker 1: 'Oh. I'm like my father there in aren't I?'

Speaker 2: 'Yes.'

Speaker 1: 'Mm. [I have the] Same eyes, look.'

Speaker 2: 'Mm hm mm.'

Speaker 1: '[They are the] Same shape.'

Speaker 2: '[You've] Seen that one of Jim, haven't you?'

Speaker 1: 'Yeah. It's a good one, that.'

This kind of ellipsis is common in conversation, but very rare in writing, apart from in texts such as advertisements which try to imitate informal speaking, and in electronic communications such as emails, text messages and online chat. Another notable feature of this example is the last line, 'Yeah. It's a good one, **that**'. The final *that* repeats the subject *it* (referring to the photograph) as a sort of *tail*. Tails are common in conversations, especially where people are evaluating something or someone, but tails are not found in formal writing (McCarthy and Carter, 1997). Tails are one of those features of spoken grammar that are highly interactive. We often neglect such features in our grammar teaching because we feel they are not correct or are too informal, but to ignore them is, in my opinion, a mistake. All languages have such interactive features in their grammar; why should we not want to teach them, especially if our learners need or want to speak fluently and naturally?

Corpora show us that there is a grammar of speaking, in which the items and structures are geared towards the relationship with the listener(s). It is important, therefore, to include the grammar of speaking in our teaching of grammar. If we do not do this, our students are likely to sound like books when they speak – too stiff, too formal, unnatural – and they will probably project an identity which is not the real, living, caring human being that they would wish to project. Nobody wants to sound like a robot. However, there remains a good deal of prejudice about the grammar of speaking, among teachers and among the general public. The objective evidence of corpora shows us that educated native users and expert non-native users exploit the grammar of speaking and use the forms we have exemplified above; there is no reason to believe that spoken grammar is any less valid or correct than written grammar. In the past, grammarians had no solid evidence of how people actually spoke. Nowadays, with easy access to audio recording techniques and the growing number of spoken corpora that show us how people really use language in their everyday lives, we have no excuse to ignore spoken grammar. If you want to read further and look at more examples of spoken grammar, see O'Keeffe et al (2007: chapters 5 and 6) and Carter et al (2011).

References

Carter, R A and McCarthy, M J (2006) *Cambridge Grammar of English*, Cambridge: Cambridge University Press.

Carter, R A McCarthy, M J, Mark G and O'Keeffe, A (2011) *English Grammar Today*, Cambridge: Cambridge University Press.

Leech, G (2000) Grammars of spoken English: new outcomes of corpus-oriented research, *Language Learning* 50 (4), 675–724.

McCarthy, M J and Carter, R A (1997) Grammar, tails and affect: Constructing expressive choices in discourse, *Text*, 17 (3), 405–429.

O'Keeffe, A, McCarthy, M J and Carter, R A (2007) *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.