**Computer-Assisted Methods of Analysing
Textual and Intertextual Competence**

Michael Stubbs

Individual texts are interpreted against a background of norms of usage in the language as a whole and/or in specific genres. Texts seem "natural" and "idiomatic" if they fit our expectations of frequent phraseology. This is illustrated by using computer-assisted methods to compare a fragment of a non-fiction text about the environment with norms of usage as sampled in a large computer-readable corpus. There is no sharp distinction between mechanisms of cohesion in non-fiction and fiction texts, but literary texts do reveal other ways of exploiting intertextuality. This is illustrated from a well-known short story by James Joyce.

When we read or hear a text, we may find the language familiar or not, and correspondingly easy or difficult to follow. Difficulties in understanding a written or spoken text – such as a set of instructions, a textbook, a lecture, a story in a conversation or a novel – can have many causes. However, by and large, we find a text easy to understand if it consists of familiar topics being talked about in ways which are familiar from other texts. In a word, understanding depends on both our textual and our intertextual competence. If everything is totally familiar, of course, the text will strike us as boring or full of clichés. But there are limits to the rate at which we can take in new information, and we can understand connected text only if we are able to relate it to what we have heard in the past, and to predict, at least partly, what is likely to be said. Conversely, we find a text difficult to understand if it is lexically and semantically dense: that is, if there is too little repetition of vocabulary, and if too many of the words are unfamiliar or are used in unusual combinations.

All texts are interpreted against an intertextual background of norms of language use, which are expressed largely in recurring multi-word combinations. These norms can be identified by the computer-assisted analysis of large corpora, and we can then compare what occurs in individual texts with what frequently occurs in large numbers of texts of different kinds. For texts of all kinds, our linguistic competence therefore depends partly on our knowledge of norms of general language use. For some texts – especially literary texts – our understanding also depends on our knowledge of how to read texts both literally and metaphorically, and this requires different aspects of intertextual competence.

I therefore give examples from two very different texts, which illustrate these two aspects of linguistic and literary competence: a short fragment from a factual text, which I analyse in some detail, and a fragment from a famous short story, which I comment on more briefly. These examples also illustrate what can – and cannot – be successfully analysed with replicable computer-assisted methods.