

Actually attested language

Here we take up again matters that have been mentioned earlier in the book. In Chapter 1, the point was made that communication involves the production of text to indicate intended meaning. In Chapter 2, reference was made to Hymes' account of communicative competence, in which he suggests that when confronted with a text in a particular language we, as competent users of that language, are capable of making four kinds of judgement about it. We can say how far it is possible, that is to say how far it conforms to what is lexically and grammatically encoded in the language. In this case, we treat the text simply as a collection of linguistic elements. But texts are normally only produced if there is an intention to communicate a message of some kind and the other three kinds of judgement have to do with this normal communicative functioning of text. Thus we can say how far a particular text is feasible, that is to say, how easy it is to process, and this, as we have seen, will depend on how it keys in with shared knowledge. How far it is appropriate (the third kind of judgement) will also, crucially, depend on how the text can be related to context to bring about reference, force, and effect.

The fourth kind of judgement has to do with whether and to what extent a particular text is attested as actually occurring usage. Here we are concerned with conformity not to linguistic rule but conventions of usage. As has already been pointed out, communication involves the selective use of the encoded resources of lexis and grammar, and we have been looking at what motivation there might be for the selection of one possibility rather than another. We have already looked at ways in which the

feasibility and appropriateness factors may affect selection. But we also need to consider which selections are generally preferred by convention. Though all lexical and grammatical elements are equally possible in the linguistic code, they are obviously not all equally common in actual usage. How then does this fourth factor of the actually performed relate to the other factors in the process of communication?

Norms of usage

With the advent of the computer, we are now in a position to provide exact descriptions of actual usage. Vast quantities of text can be collected into a **corpus** and electronically analysed. As a result, it is now possible to establish the relative frequency of occurrence of words and structures either in particular domains of use or more generally across domains to provide profiles of frequency in the language as a whole. We therefore have a **norm of usage** against which we can establish the relative (ab)normality of the occurrence of a word or structure in a particular text. If it is **abnormally rare**, we might then infer that its selection goes against the **manner maxim** and so creates an implicature. So, to take a simple example, if somebody in a conversation were to talk about *circumventing a hindrance* rather than *getting round a problem*, you would suppose that they were trying to impress or striving for a comic effect. An extreme example of abnormal usage is to be found in Henry Green's novel *Living*, with its very infrequent use of one of the most frequently occurring words in the English language, the definite article *the*:

They had taken bus. They had gone Saturday afternoon to Mr Jones' uncle and aunt that were lodge-keepers at gate of big house one mile out from bus terminus.

This avoidance of the definite article results in impossible sentences (in the Hymes sense) and, together with other deliberate abnormalities, is intended by the author, on his own testimony, to represent the basic simplicities of everyday working life.

Corpus analysis can tell us not only about the overall frequency of words, but also about their **range**, that is to say their distribution in different domains of use. The words *circumvent*

and *hindrance*, for example, uncommon though they are, might well be found quite often in certain kinds of legal genre. They would then constitute a local norm of appropriate usage. The same point can be made even about the definite article. Though this word is generally recorded as having a higher frequency and wider range than almost any other in the English language, it is quite normal for it to be omitted in newspaper headlines.

So relative frequency can be taken as having a **schematic significance** in that certain words mark particular genres or discourse domains. Thus the words *customer*, *consumer*, *product*, *marketing*, *retailer*, *sales* will be shown to have a much higher concentration in business domains than in any other genre. And again, if they were to show up in comparable numbers elsewhere this would be noticed as abnormal and the question would naturally arise as to what the motivation for such non-conformity might be. If these terms occurred in the discourse domain of clinical medicine or education, for example, one might infer that the writer's attitude to health care or education is that it is just a kind of commodity to be produced and retailed in response to market forces like any other.

Patterns of collocation

It is not, however, only the simple frequency and range of single items that is revealed in the corpus analysis of text but also, more interestingly and significantly, the **frequency and range of their patterns of co-occurrence with other items**. In Chapter 5 we looked at the kind of co-textual links that are established on-line to establish cohesion whereby the pro-form *he*, for example, makes an anaphoric connection with the preceding noun phrase *The Prime Minister*. But there are also other connections across linguistic items, certain **patterns of co-occurrence**, which **recur quite regularly as a property of texts in general**. Thus **certain items tend to keep company**, or **collocate**, with others: there is, at it were, a kind of **mutual attraction that draws them together**. So the word *unforeseen* will attract the word *circumstances*, *foregone* will attract *conclusion*, *crying* will attract *shame* and *pious* will attract *hope*, and so on. But **collocation goes beyond the relationship between two lexical items in a noun phrase to include**

many other recurring combinations in phrases like: *as a matter of fact, all things considered, when all's said and done, all things being equal, for better or for worse*, and so on. Such formulaic phrases are easy enough for proficient speakers of English to cite: they are aware of them as familiar idiomatic features of their usage, and no corpus analysis is needed to reveal them. But there are innumerable other textual patterns and phrases that emerge from corpus analysis that the language user is not aware of. It is easy enough to demonstrate this by means of a concordance, which displays all the occurrences of a particular word in lines of text so that one can see at a glance where co-textual combinations recur. Concordances for words which are semantically related can be compared and their collocational differences identified. For example, the words *big* and *large*, *little* and *small* (already mentioned in Chapter 2) are shown not only to occur with different overall frequencies, but to have rather different collocational preferences. Similarly, it turns out that the words *amazing*, *astonishing*, and *surprising* not only vary considerably in frequency but also in the way they can combine with other words in lexical phrases or bundles. Thus the last of these often occurs in the phrase: *it is not surprising that ...*, whereas the others do not. *It is not amazing that ...* and *it is not astonishing that ...*, though possible constructions in English (in the Hymes sense), seem hardly to occur in attested usage.

What corpus analysis reveals is that the constituents of texts are not so much separate words and structures as patterns of language, collocations and lexical bundles, of variable flexibility. This suggests that producing a text is, to some extent at least, a matter of assembling it from pre-fabricated parts, making whatever adjustments are necessary—changing a word here, a structure there as co-textually required. This is clearly a more efficient process than composing a text bit by linguistic bit, and knowledge of this process must be part of our communicative competence. So, in reference again to Hymes, what is actually performed relates to what is feasible. The use of recurring patterns of language, stored ready for use in the mind, is in this case motivated by the co-operative principle. To deliberately disregard them would be to violate the manner maxim.

Semantic prosodies

But what motivates the use of particular collocational combinations? In some cases we might suggest that if there is any motivation at all, it is only to signal membership of a particular community by conforming to its customary idiom. In other cases, a collocation may be explained by reference to general semantic principles. Take the example of the phrase mentioned earlier: *it is not surprising that ...* Why is it that it is normal for *surprising* to occur in this combination but not *amazing* or *astonishing*? One answer might be that it is simply a matter of custom, a kind of ingrained habit and that's that. Alternatively we could look at whether there is anything in the semantic meaning of these words which might explain this collocational constraint. Further consultation of the concordance lines would reveal that *surprising* is also often preceded by the intensifier *very*, but the other words are not. The reason for this, we could conclude, is that intensity is already a semantic feature of *amazing* and *astonishing*. We might then propose the general rule that words which are marked for intensity do not normally combine with *very* or appear in the frame *it is not—that*, whereas words unmarked for intensity, like *surprising*, do. Or take the verb *cause*. A glance at a concordance display will show that its normal collocates are words which denote disagreeable things, like *pain, disease, distress, disruption*, and so on. It is therefore said to have a negative semantic prosody. In contrast, the verb *bring about* has a positive semantic prosody in that its typical collocates are words like *improvement, cure, solution, success*. Thus a problem is caused but a solution is brought about.

In these cases, then, we can explain particular collocational combinations in texts by identifying the more general semantic properties of the words concerned. We need to note that these have to do with tendencies and probabilities. The essential point is that these are the default meanings of the words—meanings that they would normally be expected to have—and that any departure from the norm is likely to be noticed, with the likely consequence, yet again, that implicatures will arise. So if we come across a reference to a solution that has been caused, or a problem that has been brought about, and if we have reason to believe that

the selection of words is deliberate, we are likely to interpret them as indicating irony, or some other attitude, on the part of the author.

With reference to the two kinds of knowledge referred to in Chapter 6, these semantic prosodies can be said to be part of systemic knowledge. There are other collocational patterns that, like certain word frequencies as mentioned earlier, can be related to schematic knowledge in that they can be said to represent the way reality is conceptually constructed by a community of language users. If, for example, (to return briefly to the text fragments we looked at earlier) we were to call up a concordance for the words *refugees* or *police* as they occur in a corpus of newspaper texts, the collocations they enter into could be said to indicate how refugees and the police are conceptually represented in the press. A comparison of concordances from the corpora of different newspapers, or of different domains of use altogether, might indicate degrees of schematic generality—the extent to which the collocations reflect a particular or more general set of cultural assumptions.

Conclusion: text analysis and discourse interpretation

So, to take up the question raised at the end of the last chapter, how far does text analysis contribute to an understanding of the discourse process and enable us to assign communicative significance to what is said or written? As we have seen, the computer analysis of corpora provides us with profiles of the occurrence and co-occurrence of textual features and these serve as a norm of what is customary against which any particular instance of usage can be compared. Where there is deviation from the norm, and therefore a departure from what is expected, this is likely to be taken as acting contrary to the co-operative principle and will create an implicature—there is some underlying significance here, some effect intended beyond what is actually said.

Just what this significance may be, however, is a different matter. And here, in this last chapter, we return to the points made in the first about the nature of language in use and the

distinction between text and discourse. Textual analysis can only tell us about texts, the language that people produce (or more strictly, have produced) in the process of communication. It cannot tell us about the process itself, about how people negotiate a relationship between text and context in order to bring about a degree of discourse convergence appropriate to their purpose. As we have seen, what people intend to mean by the texts they produce, and what they are interpreted as meaning, cannot be directly inferred from the texts themselves, no matter how precisely these are analysed. For texts only have reality for the language user as a means to an end, as a way of mediating discourse, and they are not normally produced as an end in themselves. But they can, of course, be analysed as an end in themselves, and this is what corpus analysis so effectively does. The result is a wealth of new information about how texts are constructed. These textual facts cannot account for all the other factors we have considered in this book that come into play in making meaning, and it would be a mistake to claim that they can. But what they can do is to alert us to possible intentions and interpretations which we might otherwise not be aware of, and so provide a basis, and a stimulus, for further empirical enquiry into the pragmatics of discourse and the nature of human communication.