

## Foreword

Richard Hudson<sup>2</sup>

This book deserves a prominent place in the growing international literature on dependency grammar and computational linguistics. The nature of syntactic structure is one of the most disputed questions in linguistics because science and tradition are so hard to separate in one of the most fundamental disputes.

An ancient tradition in Europe and the Middle East gives priority to the word as the basic unit of syntax, which means that syntax is primarily a matter of defining the relations between individual words—what have come to be called “dependencies”. For instance, in the sentence “Small children often cry”, the syntactician identifies just three dependencies that relate *small* to *children*, *children* to *cry*, and *often* to *cry*; once these dependencies have been identified, and the words and dependencies have been classified, nothing more remains to be said about the sentence’s structure.

A much more recent tradition started with Leonard Bloomfield and the American structural linguists in the early twentieth century, and has come to dominate syntactic theory. In this tradition, the structure of a sentence consists of a more or less elaborate hierarchy of “phrases” in which the word has no particular priority. In “phrase-

---

<sup>2</sup> Fellow of the British Academy. Emeritus Professor, University College London. Founder of Word Grammar.

structure grammar”, in contrast with “dependency grammar”, the four words of our example are combined with at least three phrases (*small children, often cry* and *small children often cry*) and possibly more—for example, *cry* would typically be classified not only as a word but also as a one-word phrase.

Unfortunately for scientific progress, this tradition was built from scratch, with very little reference to the existing dependency theory, and continues to ignore the dependency alternative. The result is that the very foundations of the scientific study of syntax are unstable, with an unresolved conflict between phrase structure and dependency structure. The main influence on syntactic theory is not debate and research, but geography. Linguists trained in America adopt phrase structure, while the more independent syntacticians of Europe favour dependency theory. This cannot be good for our discipline.

This background explains why a European dependency grammarian like me is pleased to see dependency theory being so ably developed by Haitao Liu outside the traditional “battle-field” of Europe and America, in the People’s Republic of China. His dependency analyses of Chinese are a particularly welcome contribution to dependency theory. However, what is most exciting about his work is the way in which he has applied dependency analysis to large corpora in different languages, something which is possible nowadays thanks to the use of computers.

A corpus of naturally occurring sentences is the ultimate test of any theory of language precisely because it shows how important it is, in theorizing about language, to go beyond mere grammar. For instance, Liu reports that his Chinese corpus contains a very similar pro-

proportion of nouns to the proportion that I reported some years ago for several English corpora: about 41%. This is, indeed, an extraordinary finding; but it demands an explanation. Why should this figure emerge from such different corpora? One thing is clear: the explanation cannot lie only in grammar. To understand usage, we need a much broader range of theories: not only linguistic theories of grammar, vocabulary and genre, but also psychological theories of working memory. Liu's studies address many of these questions, though it is surely too soon to expect satisfying answers to many of them.

Perhaps the most interesting topic discussed in this book is the statistical measure of syntactic difficulty called "dependency distance". This measures the load which a word places on working memory, on the reasonable assumption that a word is kept active in working memory until all its outstanding dependencies have been satisfied. Returning to our earlier example, "Small children often cry", most of the words are very easy to process because their dependencies are satisfied by the next word; for instance, *small* needs a "parent" word, but this is immediately provided by *children*; and the same is true of *often*, which depends on the next word *cry*. But *children* is slightly harder because it is the subject of *cry*, from which it is separated by *often*. This increased load is still trivially easy for adult English speakers, but as the dependency distance between *children* and *cry* increases, the difficulty increases, and most English speakers struggle with really long subjects such as "Small children with anxious parents who keep trying to get them to smile and be happy even when they have tummy ache or when they are teething often cry".

Earlier work on dependency distance in languages such as Eng-

lish suggest that the limitations of working memory keep the average dependency distance quite low, and one would expect the same to be true in other languages. But Liu has found evidence for considerable variation among languages. In particular, he reports that the average dependency distance in Chinese is at least twice as great as that in English. This is an extraordinarily important finding which should stimulate a great deal of productive research. Do other corpora in English and Chinese show the same differences? If they do, why are the effects of working memory so different in the two languages? Is it because Chinese words are easier to hold in memory, so that more words can be kept active? Or is it because Chinese speakers have less limited working memories? I, for one, look forward very much to the light that Liu's future work will certainly cast on these fascinating questions.