

REVIEWS

Aarts, Bas and **Charles F. Meyer** (eds.). *The verb in contemporary English: Theory and description*. Cambridge: Cambridge University Press, 1995. Reviewed by **Peter Collins**, University of New South Wales.

This is a collection of quality essays fittingly dedicated to Sidney Greenbaum. It is organised into two parts entitled 'Theoretical approaches to the study of the English verb' and 'Descriptive approaches to the study of the English verb'. Here 'descriptive' is interpreted narrowly to mean 'based on corpus data': the concerns of several chapters in the first part (e.g. by Huddleston and Coates) are in fact as much 'descriptive' as they are 'theoretical'.

The editors have by and large achieved thematic unity across the collection. However, there are several chapters written on topics whose relationship to the verb, it must be said, is somewhat tenuous (particularly those by Hudson, and by Finegan and Biber, on complementisers). From time to time one comes across assertions which betray an almost self-conscious concern with thematic relevance, as on p.75: 'The relevance of secondary predicates to the study of the verb is that they are often closely related to verbs, and have been analysed by some linguists as verbal complements.'

The volume opens with a lucid introductory overview by the editors. This is followed by four chapters concerned in various ways with verb-related elements. The first, by Meyer, posits two levels of grammatical relations, clause-level (e.g. complement) and general-level (e.g. complementation). The general distinction, familiar from X-bar theory, between complements and adjuncts is dispensed with, a decision which will presumably make it difficult to account for the similarities between clausal-phrasal pairs such as *Sampras defeated Agassi* and *Sampras' defeat of Agassi*. Tantalizingly, Meyer leaves open the issue of 'what particular relations need to be posited at each level in order to yield a satisfactory description of English and other languages'. (pp 37–38). Hudson ('Competence without Comp?') argues forcefully that the category 'Comp' is an 'invention' of modern linguistics that has served as 'a dustbin for the words which can introduce a complement-clause but which do not seem to belong to any other word-class' (p. 42). It is true that complementisers share structural similarities with other word

classes, but surely a distinctive feature of *that* is its omissibility (while *whether*, *if* and *for*, though not omissible, contrast systematically with *that* in terms of the mood and finiteness of the clausal complement). Schlesinger ('On the semantics of the object') invokes the semantic notions of 'completion' and 'feat' along with various processing constraints, in order to distinguish direct objects from indirect objects and objects of prepositions. The value of this analysis no doubt resides more at the language-general level of description, than the language-particular, since categories such as object can be defined successfully only via structural criteria in a language-particular description. Bas Aarts investigates 'secondary predicates', phrases predicated of a constituent standing in a thematic relation to the main verb of a sentence (e.g. *raw* in *Jim ate the meat raw*), taking care not to be drawn into such theoretical debates as those over the existence of small clauses. One of Aarts' major conclusions, that object-related resultative predicates as in *Jim rinsed the cups clean* are adjunct-like rather than lexically selected, may well be correct, but the evidence offered (namely that 'resultatives' is an adverbial notion, and that paraphrases of such predicates involve adverbial clauses) needs supplementation.

The next three chapters deal with the verbal categories of tense, aspect, voice and modality. Huddleston argues convincingly for a distinction between the inflectional preterite and the analytic perfect in English, as respectively primary and secondary past tense (with the former being more highly grammaticalised). He observes that both, in their basic meaning, express the anteriority of the time referred to, to the time of orientation, and when the preterite is not available to express this relationship (in cases of modal remoteness and backshift), the perfect is used. Rosta ('The semantics of English mediopassives') presents a Word Grammar analysis of sentences like *The book reads well*, which explores semantic properties of their subjects, the collocation of mediopassive verbs with adverbs such as *well* and auxiliaries such as *will*, and the typically 'habitual' aspect of mediopassive verbs. Coates' chapter exploring the difficulties in applying the root-epistemic distinction to modals such as *can* and *may* when they express possibility will strike a chord in all those who have grappled with the semantics of the modals. Coates claims that there have been two linguistic developments resulting from the weakness of the root-epistemic possibility distinction, but she does not provide adequate supporting evidence. The first is a growing trend for *may* to express a merging of the root and epistemic possibility meanings (a contention which could quite readily be confirmed or

disconfirmed via inspection of an historical corpus), and the second is the emergence of epistemic meanings for *can* in spoken American English (for which the one example Coates provides, overheard at a recent conference, is insufficient documentation).

The corpus-based articles in Part 2 stand alongside their more theoretical counterparts in Part 1, symbolically betokening the coming-of-age of computer corpus linguistics. The first two chapters deal with aspects of verb complementation. Jan Aarts and Flor Aarts show how computer corpus analysis can combine a ‘microscopic’ examination of language use and a ‘macroscopic’ examination of numerical data, in a case study of complementation with the verbs *find* and *want*. The usefulness of this exercise compensates for some questionable analyses, such as the treatment of *I don’t think I want to find him for you* as ditransitive rather than monotransitive. Leech and Li demonstrate the value of syntactically parsed corpora in their study of the gradience between NPs and AdjPs in predicative complement function.

Two papers deal with aspects of verb collocation. Algeo’s paper on ‘expanded predicates’ (i.e. idioms such as *have a look* and *take a walk*, with a general or ‘light verb’ followed by a more specific NP) contains a revealing comparison of British and American usage. After examining LOB and Brown, Algeo confirms the popular view that British English favours *have*-predicates, but rejects the view that *take* is favoured in American English. Johansson’s paper on verb-adverb combinations, establishes that the distribution of the adverb *badly* is sensitive to the meaning of the verb, and is suggestive of how a distributional lexicon might be structured. Appropriately, Johansson concludes with an exhortation to colleagues to pursue larger-scale studies in this area by exploiting the vast corpora and computational tools that are now available.

The next two papers introduce an historical dimension. Finegan and Biber examine diachronic patterns of alternation between *that* and zero as complementisers in the historical corpus ‘ARCHER’, noting the recent reversal of a former trend to favour zero in the register of letters. The contrast between this finding and those of several recent studies showing a preference for zero highlights, as Finegan and Biber acknowledge, the pitfalls of making comparisons across corpora. Mair compares data from LOB (sampling year 1961) and from a corpus that he is currently compiling (sampling year 1991) to mount a fascinating argument that the verb *help* is currently undergoing grammaticalisation (as evidenced by an increase in textual frequency, and its semantic ‘bleaching’) from a distinct lexical item to a type of semi-auxiliary or infinitival marker.

The focus of the final two papers is on spoken language. Svartvik and Ekedahl find that differences in the types and frequencies of verbs in two spoken registers, 'private speaking' and 'public speaking', are greater than those between the latter and writing. This leads them to question the validity of the traditional speech/writing dichotomy. Stenström demonstrates the inadequacy of a conventional grammatical treatment, and the need for a discourse-functional description, of 'comment clauses' of the type *I mean* and *you know* in the London-Lund Corpus.

This collection has much to tempt those who prefer an approach to English grammar that is relatively 'theory-neutral' and informed by corpus-interrogation.

Geoffrey Leech, Greg Myers, and Jenny Thomas (eds.) *Spoken English on computer: Transcription, mark-up and application*. Longman, 1995. 260 pp. ISBN 0-582-25021-8. Reviewed by **Anne Wichmann**, University of Central Lancashire, Preston.

As the editors of this book rightly say, corpus-based research into spoken language is 'taking off'. If ever there was a moment for providing students, teachers and researchers with a view of the field, this is it. The book *Spoken English on computer* represents a timely and extensive survey of current practice in the transcription and analysis of spoken corpora. It seems, however, that the current practice still leaves a lot to be desired.

The book is divided into three sections, each with an introduction, and headed by an excellent general introduction. These editorial texts are extremely valuable; they almost become chapters in their own right and are well worth reading. The first section is the longest and centres round fundamental theoretical aspects of spoken language analysis. The second section offers papers on more specialised applications and the ways in which computerised texts could be encoded to meet specialised needs. The final section contains accounts of individual corpus projects, including the BNC, the LLC, the ICE and the SEC.

The first two chapters in **Section I** are among the best in the book. They are very readable and very salutary. Edwards, in chapter one,

considers the fundamental problem of transcribing spoken data. As she rightly says, 'a transcript is fundamentally selective and interpretative'. The resulting representation of the primary data must permit the questions to be addressed for which the data was intended, suggest new questions to be asked, and do both without misleading or hindering the analyst. In chapter two, Cook addresses the temptation to restrict the encoding of data to those features which are readily identifiable and categorisable, while neglecting those aspects of interaction which are gradient, vague or indeterminate. He recognises the paradox that an analysis of spoken interaction cannot be usefully made without reference to those elements which are hardest to transcribe.

Further chapters in this section deal with the representation of prosody (Chafe) and the complexities of the TEI encoding conventions (Burnard and Johansson). Chafe considers the use of conventional punctuation marks to indicate prosody in a transcription, which makes perfect sense for use by non-specialists. I would, however, have welcomed in this section a discussion of some of the theoretical problems with prosodic categories.

Burnard and Johansson give respectively a general overview, and then a detailed account, of features suitable for encoding transcriptions of spoken texts. Johansson admits that the coding is very complex and only has a chance of being really useful if there is some kind of interface which converts it into a readable form. This is echoed by Sinclair, who predictably has little time for elaborate encoding. In my view, however, his plea for keeping to the «raw» data rather misses the point made by several contributors that no transcription of a spoken text is in fact the raw data.

Section II deals with specialist applications of spoken language research. It includes a very interesting chapter by Perkins, containing a clear overview of users' needs in the analysis of disordered speech. Sebba gives a rather brief account of the main technical and theoretical issues to be addressed when transcribing bilingual data. The technical problems described are typical of a slightly anachronistic concern for the printed page, and I would have liked more discussion of the complex issues of categorising bilingual data.

A valuable contribution by Cheepen deals with the use of corpora for studying higher level discourse structures in conversational data. This is mentioned briefly by Monaghan in the previous section, but Cheepen makes a useful distinction between different types of 'conversation'. The most interesting feature of this chapter is the simplicity of the model

(signalling of global structures by lexical items), and particularly the idea that they can be analysed automatically.

Considering how little space is devoted in this book to phonetic aspects of speech analysis, Roach and Arnfield's chapter is disappointingly short. It deals with one of the most crucial issues for future work on speech corpora – the alignment of wave form and (phonemic) transcription. It also gives a brief introduction to the American TOBI system of prosodic transcription, but with little detail and without discussing how it relates to the British system.

Grammatical tagging is covered by Garside (tagging the spoken BNC), who gives an interesting account of how automatic tagging procedures are applied to the transcription of spoken language. This involves adding tags for slang, interjections, taboo words etc. Also, in the case of words with more than one possible tag, modifying the frequency information to account for spoken usage. Other problems to be solved are disfluencies and repetitions in speech, including filled pauses. As a brief account of how automatic tagging is carried out, and with the addition of the tagset as an appendix, this is a useful reference text.

Thompson *et al* describe a speech corpus based on the HCRC map task (a corpus of spontaneous speech elicited in a controlled context). This is a fact-filled account of the procedure – very useful for anyone thinking of embarking on a similar task (or calculating necessary resources e.g. in a research proposal).

Section III, the final section, contains accounts of transcription practices in the LLC (Peppe), Cobuild (Payne), SEC (Knowles), ICE (Nelson), BNC (Crowdy) and COLT (Haslerud and Stenström). The paper by Knowles might have been better placed in the first section. Although it deals with a specific project, turning the MARSEC corpus into a relational database, and in this respect belongs to the others in this section, the theoretical point being made here is at least as important as the project itself, and deserves to be highlighted. It is the only contribution which departs from the conventional linear representation of data, describing instead a data-handling process in which only non-derivable information is stored.

The book shares some of the features of a conventional edited work with the features of a compendium. The chapters, twenty in all, are very unequal in length, some amounting to no more than a couple of pages. This does not detract from the book as a whole; indeed, it probably avoids a certain amount of repetition. Just occasionally though, one does feel that the individual authors could have made a little more

effort to provide a rounded report, rather than a page out of a manual or a poster on a page, or even a transcript of their oral presentation. Nonetheless, I think this book will be appreciated by teachers, researchers and students alike. The wide coverage of techniques and applications will make it invaluable as a reference work. The remarkable clarity and comprehensiveness of the editorial introductions will make these sections alone valuable reading for those new to the field.

While the book succeeds in giving wide coverage to current practice, it is clear that the state of the art is not healthy. The book reveals some interesting problems:

1. that much more discussion is needed about the theoretical issues underlying encoding categories;
2. that we are still obsessed by the printed page;
3. that we need reminding that the transcription is NOT the primary data but 'selective and interpretative'.

There is still much to be done, and the book does not on the whole suggest a coherent way forward. That is not the fault of the book. Its strength is that the issues have been brought together in this way. It now presents a challenge to the future.

Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Arto Anttila (eds.). *Constraint Grammar: A language-independent system for parsing unrestricted text*. Berlin: Mouton de Gruyter, 1995. 430 pp. ISBN 3-11-014179-5. Reviewed by **John Hutchinson** and **Ruthanna Barnett**, University of Lancaster.

The automatic parsing of real language has been an elusive goal, and so a system which purports to do just that is very welcome. This volume uses the English Constraint Grammar (ENGCG) to illustrate the Constraint Grammar system, developed at Helsinki.¹ Successful Constraint Grammars have been produced, or are in development for Finnish, Swedish, Basque and German.

The preface tells us of the four major objectives: language independence, robustness (an ability to handle unedited text), very high quality standards, and precise evaluation. Chapter 1, by Karlsson, develops in detail the 24 central goals which constitute the basic philosophy of the CG approach (p. 2). Some of these goals would be common to any parsing system, whilst others are CG-specific.

In its simplest form, the system involves morphological/morpho-syntactic analysis, followed by surface syntactic analysis. The former is based on the lexicon, which is 'designed according to Koskenniemi's (1983) two-level model, TWOL' (p. 11). Initially, each word is assigned all of its possible labels, and then the core of the system (the constraint grammar) is used for disambiguation. In this sense, constraints define the analyses which may or may not occur.

In chapter 2, dealing with the 'formalism and environment', Karlsson provides a full documentation of the CG parsing. The version of the system which provides the most comprehensive environment for developing and testing constraints (implemented in Lisp) is capable of processing 3–4 words per second on a Sun SPARCstation 2, while a production version, written in C without the support environment, processes 400–500 words per second on a Sun SPARCstation 10/30 (p. 45).

The system is complicated; after pre-processing, the maximum CG parsing set-up is:

- (a) morphological analysis
- (b) morphological heuristics
- (c) deactivation heuristics
- (d) clause boundary mapping
- (e) context sensitive disambiguation (Up to 5 cycles through (d) and (e).)
- (f) heuristic disambiguation
- (g) text-based disambiguation
- (h) quantitative disambiguation
- (i) morphosyntactic mapping
- (j) syntactic analysis (Up to 5 rounds.)
- (k) heuristic syntactic analysis

Most of these stages will be discussed briefly below.

Used in default mode (i.e. with default values for the 26 optional parameters), constraint grammar parsing proper starts with step (d). This means that input to the parser is a morphologically analysed text, and a *constraint file* which contains 'disambiguation constraints, clause boundary mappings, and surface-syntactic rules, plus the requisite declarations

needed for defining the proper contexts of the constraints' (p. 45). The bulk of chapter 2 describes the contents of the constraint file, dealing with *inter alia* the definition of sentence delimiters, the mapping of syntactic labels to morphological features, and the formalism used to describe the various disambiguation constraints.

Chapters 3, 4 and 5 introduce the ENGTWOL lexicon; chapter 3 deals with the lexicon's compilation and testing, chapter 4, its feature system and chapter 5 looks at solutions and problems.

As one of their general, guiding principles, the Helsinki team state that 'Morphological and lexical analysis is the basic step of parsing.' (p. 11). It is not surprising therefore that the lexicon is a crucial part of the overall system. After basic pre-processing, the morphological and lexical analysis stage introduces possible interpretations for later disambiguation. Voutilainen and Heikkilä (p. 94) claim that, typically, '35–50% of all words in running text receive more than one analysis' and that each word is given on average 1.7–2.2 alternative morphological readings.

The 'lexicon' actually consists of 154 sublexicons, access to which, via pointers, is as required to produce a reading of a given word of text. Associated with each reading is a part-of-speech and, optionally, other features which may include syntactic-function labels.

Dealing first with the construction and coverage of the lexicon, which includes approximately 56,000 entries, *A comprehensive grammar of the English language* (Quirk *et al.*: 1985) formed the basis for the treatment of morphologically irregular lexis, including irregular verbs and adjectives, etc., and closed class parts-of-speech. Adding the more regular words involved analysis based largely on the tagged Brown and LOB corpora. Further additions were made using dictionary and other sources. Voutilainen and Heikkilä (p. 92) give the main lexis selection criteria as: '(i) no misspellings; (ii) domain-generic material should be preferred; (iii) all morphological alternatives should be accounted for.'

A number of 'descriptive policies' are discussed (pp. 94–97) which explain how the demands of a practical parser have been balanced with lexicographic standards. The consistency with which the CG team approach their main aim, a robust working parser, makes their arguments persuasive. When introducing the <Rare> feature, then, to allow for the inclusion of significant, but infrequent, syntactic ambiguities (exemplified by this example 'The chauffeur did not 'sir' his employer.' taken from LOB and given on page 96), it was slightly disappointing that an email version of the ENGCG parser (set up to allow wider access to the system) returned the analysis:

```
"<*the>"
  "the" <*> <Def> DET CENTRAL ART SG/PL @DN>
"<chauffeur>"
  "chauffeur" N NOM SG @SUBJ
"<did>"
  "do" <SVO> <SVOO> <SV> V PAST VFIN @+FAUXV
"<not>"
  "not" NEG-PART @NEG
"<$'>"
"<sir>"
  "sir" <Title> N NOM SG @NN>
"<$'>"
"<his>"
  "he" PRON PERS MASC GEN SG3 @GN>
"<employer>"
  "employer" <DER:er> N NOM SG @OBJ
"<$.>"
"<$2-NL>"
```

To check the master lexicon, a 12,548 word extract from a British newspaper was processed by ENGTWOL. 97.7% of the input words were given a correct reading, possibly amongst others. Even with this coverage, Voutilainen and Heikkilä conclude that a new version of the lexicon should include 'virtually everything in a large machine readable dictionary ... to guarantee exhaustivity' (p. 101), with each entry furnished with frequency information derived from tagged corpora of the scale of the *Bank of English*.

The feature system, described in chapter 4, uses 139 morphosyntactic features. Heikkilä (p. 111) divides these into three main types: part-of-speech features, inflectional features and auxiliary features. The system has a base set of 16 POS categories which roughly follows CGEL with additional features defined for abbreviations, adjectives, adverbs, determiners, nouns, numerals, pronouns and verbs. These range in number from 5, for numerals, to 39 for verbs and include such things as number (e.g. SG – singular), person (e.g. PL1 – 1st person plural) and verb complementation pattern (e.g. <SVO> – monotransitive). Maybe a hint of what is to come is the mention of the 'domain feature' (p. 130), e.g.: ('rheumatoid' <D:med> A ABS)

The feature system, and its application, is defined with the subsequent syntactic analysis in mind; ambiguity is not introduced if it cannot be

resolved later. Because the Helsinki team refrain from making bold claims about the wider applicability of their annotation scheme, it is difficult to criticise. Where policy decisions appear to be compromises, those decisions are convincingly argued, usually on both theoretical and practical grounds.

Chapter 5, which deals with ‘solutions and problems’, gives more detailed treatment of the handling of, for example, fixed syntagms, both ‘idioms’ and ‘compounds’, and systematic closed-class ambiguity. The section on remaining problems really amounts to further consideration of the limitations imposed by the practical system. Within this (p. 158), Heikkilä states that ‘the system under discussion is not concerned with semantics’, but eventually concludes that ‘ignoring semantics totally may ... not be an ideal solution in the long run’.

In Chapter 6, Voutilainen introduces morphological disambiguation. In the Constraint Grammar system, the output of the morphological analyser consists of all the possible analyses. The disambiguator acts as a filter, discarding those analyses that the context prohibits.

The Helsinki team evaluate the success of the disambiguation process in terms of *recall* and *precision*. 100% recall would mean that all legitimate readings survive the grammar, and 100% precision would mean that all illegitimate readings are discarded. For a system to have perfect recall and precision, it would have to produce all and only the correct analyses for the input.

Difficulties in achieving perfect recall and precision can be traced further than the system. Firstly there needs to be agreement on the correct analysis for the system to produce. In some cases of part-of-speech assignment, and in many more cases with higher level syntactic assignment, the correct analysis will be theory-based. The Constraint Grammar system is largely based on CGEL, but since this is a descriptive grammar, there will be instances in which other decisions have been taken. These decisions should be documented, a point made by Voutilainen. A small test carried out on persons already familiar with the ENGCG parsing scheme showed that 100% agreement was possible on the analysis within a given parsing scheme, i.e. once the scheme was developed, high consensus on its application could be reached.

The constraints are often constructed in a negative form; the correct description is reached or approached by ruling out what cannot be the case in a given context. Constraints may refer to morphological or syntactic contexts, e.g.:

Morphological constraint –

A reading containing verbal features (PRES, PAST, IMPV) is to be discarded if it occurs after an article.

Syntactic constraint –

Noun is subject if followed by active finite verb form, and no nouns intervene.

The writing of the disambiguation grammar is described as an incremental process. Recall can be increased by testing each constraint against a corpus, and precision may be improved by adding new constraints, whose need will be illuminated by errors in the test run. There are at present 1,100 constraints for morphological disambiguation, and each of these constraints expresses part of one of 23 grammatical generalisations on which the system is based. These generalisations are divided into six types:

1. Nominal phrase, e.g. *determiners and premodifiers are followed by a nominal head. In between, only certain (potentially co-ordinated) determiners and premodifiers are legitimate.*
2. Verb chain, e.g. *To the right of an infinitive marker, there is an infinitive.*
3. Clause, sentence, e.g. *A sentence contains at least one (potentially co-ordinated) main clause.*
4. Agreement, e.g. *An accusative is preceded by a main verb or a preposition or a co-ordinated accusative.*
5. Co-ordination, e.g. *Only likes co-ordinate.*
6. Complementation, e.g. *A preposition is immediately followed by a co-ordinated preposition of a noun phrase acting as a complement.*

In tests, the morphological disambiguation grammar produced output in which 93–97% of all words were fully unambiguous, and at least 99.7% of all words retained the legitimate reading.

For evaluation, ENGCG was compared with two probabilistic systems (CLAWS1 (Garside *et al.* 1987) and PARTS (Church 1988), in the analysis of five sample texts. As Voutilainen points out, this type of comparison is quite difficult to evaluate, since each system uses a

different annotation scheme; thus the ‘correct’ answer will actually be different in some cases for each system. The results are given in terms of *recall* and *precision*. The sample texts contained a total of 1267 word-form tokens, and the results for the three systems are shown below:

TOTAL	CLAWS1	PARTS	ENGCG
recall	96.95 %	96.21 %	99.77 %
precision	96.95 %	96.21 %	95.54 %

ENGCG recalls more, but loses a little on the precision figures (i.e. some word-forms are assigned more than one analysis). The probabilistic systems assigned the wrong analysis about 4% of the time, and although ENGCG was incorrect only 0.3% of the time, it also left some 4% of cases partly unresolved. However, the 4% of errors for PARTS and CLAWS are not the same 4% as cause problems for ENGCG. Many of the cases in which ENGCG was unsure of the final analysis can be solved using heuristics. A combination of ENGCG and CLAWS might be expected to give an unambiguous precision rate of 99.1% (p. 282).

In Chapter 7, ‘Testing and Modifying the Disambiguation Grammar’, Voutilainen and Savolainen describe the process of improving the lexicon and grammar according to errors committed in the morphological assignment. The improvement in error rate is reported as 0.3% to 0.04%. The remaining errors may be due to such phenomena as ellipsis, unconventional uses, non-standard usage, spelling errors, and other more serious problems such as recognition of adjectival heads of noun phrases, and deferred prepositions. Of course, these are the very things that make the processing of unrestricted text a significant problem.

ENGCG at least partly provides solutions to these errors by introducing *heuristics* (Chapter 8; Voutilainen). The system is flexible and, depending on the task at hand, priorities may be different for recall and precision. These varying needs can be catered for using a combination of ‘more reliable’ grammar-based descriptions (constraints) and ‘less reliable’ descriptions (heuristics): heuristic disambiguation constraints, text-based heuristics, noun-group learning.

Heuristic morphological disambiguation constraints are formally similar to grammar-based constraints, but are used only when no grammar-based

constraint has been able to resolve the ambiguity. In general, the default is to assign a noun reading, unless specific criteria are filled; e.g. words ending in *-ly* will also be given an adjective and adverb reading. Text-based disambiguation involves using statistics derived from annotated corpora, and Voutilainen also mentions the possibility of using ENGCG to derive information from a fully disambiguated part of a text, and utilising that information for the analysis of the remaining part of the same text. Noun-groups (nominal heads with one or more nominal premodifiers) may be identified and this collocational information used for resolving remaining ambiguities. With preliminary experiments, this seems to work well in conjunction with the other heuristic rules, within limited domains.

In conclusion, (p. 314) Voutilainen says that the experiments show that the heuristics used were potentially less error-prone than traditional probabilistic techniques (95–96% recall as compared to 85% for probabilistic systems).

In the final chapter, Anttila introduces the syntactic analysis process through the example case of the English Subject. Syntactic function labels are assigned in a similar way to morphosyntactic labels; each word is assigned all of its potential syntactic functions, and constraints are used to reduce the ambiguity. Syntactic constraints are both positive and negative. The latter, which define contexts in which the function is not possible, are more common, and although less informative, are generally simpler. These negative constraints remove inappropriate tags, and could therefore be seen as constituting a grammar of, for example, non-subjects rather than subjects (p. 338). Anttila admits that ‘this is something linguists may find awkward’ (p. 318). An example of this negative type of constraint might be represented as:

- a) In prose: *A word is not a Subject if no inversion auxiliary precedes and no finite verb follows.*
- b) In CG rule notation: (@w = s0 (@SUBJ) (NOT *-1 INV) (NOT *1 VFIN))

Syntactic constraints are both constraints and principles. For example, the *uniqueness principle* (UP), introduced by Anttila as a ‘potentially universal constraint’ (p. 341), states that:
*A clause may contain at most one instance of each principal function.*²

Principles are like constraints in that they are independent; they are not like principles in a theory such as HPSG, in which they simplify rules. In CG in fact, the inter-relation of constraints and principles will result in redundancy, but ‘CG [was developed] as a tool for the descriptive grammarian rather than a theoretical linguist’s workbench’ (p. 350).

At times, this book reads a little like a collection of papers, with some repetition of the general ideas in place of a progressive explanation of the system. Apart from this, our only real grumble is the treatment of the syntactic analysis, details of which are largely limited to the final chapter, and, even there, only the subject function is discussed in depth. Having said that, this last chapter is a valuable explanation of how Constraint Grammar models the language. As Anttila explains, the description was chosen for both practical and theoretical reasons. From a practical point of view, CG was ‘developed for disambiguating and parsing large corpora for information retrieval. The important considerations were for robustness and computational efficiency not linguistic felicity and descriptive convenience – emphasising one will usually compromise on the other’ (p. 346). And on a theoretical level, the choice came ‘from deep scepticism concerning the value of many linguistic concepts to disambiguation and parsing (Karlsson 1990)’ (p. 346). Karlsson’s original idea was to parse real text without high-level theoretical concepts by using low-level distribution facts derived from data.

On the basis of this book, we think that the Helsinki team have made fine progress towards not only achieving their goals but in some ways transcending them. We heartily recommend this book as a significant contribution to the field of NLP.

Notes

1. Much of this work was carried out under ESPRIT II project no. 2083, Structured Information Management: Processing and Retrieval (SIMPR); with funding provided by the Technology Development Centre of Finland (TEKES).
2. A clause being ‘a string of words that occurs between any two UP-boundaries, such that the string itself does not contain a UP-boundary’, a UP boundary may be for example, a comma or an explicit clause boundary (CLB), and the principal functions include e.g. finite main verb, subject, object complement.

References

- Church, K. 1988. A stochastic parts program and noun phrase parser for unrestricted text, *Proceedings of the Second Conference on Applied Natural Language Processing, ACL*, 136–143.
- Garside, R., Leech, G., Sampson, G. (eds.) 1987. *The computational analysis of English*, London: Longman.
- Karlsson, F. 1990. Constraint Grammar as a framework for parsing running text, in Hans Karlgren (ed.) *COLING-90: Papers presented to the 13th International Conference on Computational Linguistics*, Vol.3 168–173.
- Koskenniemi, 1983. *Two-level morphology: A general computational model for word-form production and generation*, (Publications of the Department of General Linguistics, University of Helsinki, no.11), Helsinki: University of Helsinki, Dept. of General Linguistics.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik 1985. *A comprehensive grammar of the English language*, London: Longman.

Dieter Mindt. *An empirical grammar of the English verb: Modal verbs.* Berlin: Cornelsen Verlag, 1995. 192 pp. ISBN 3-464-00649-2. Reviewed by **Vera Kamphuis**, University of Nijmegen.

Reading *An Empirical Grammar of the English Verb* is an ambivalent experience. The virtues of the book are, in many places, the same factors that create some of its disadvantages. This comment applies to the content of the book as well as to its presentation. Yet, the final impression that the English verb phrase is a clearly structured phenomenon is a virtue of the description. In what follows, I shall attempt to clarify the matters that contribute to this impression in more detail.

In the introductory chapter, the author gives a detailed account of the basic principles underlying the grammar, and of the procedures that were used in creating it. The grammar is based on large amounts of language data derived from authentic English. The approach is inductive, i.e. from language to grammatical generalization rather than from pre-stated rule to example. The grammar is not based on previous grammars, but makes a fresh start with the description of English. It is learner and teacher oriented.

The grammar consists of three parts. The first part (pp. 13–55) presents a new description of the English verb phrase. Its syntactic structure is represented in three-dimensional form; each of the subparts (or triangles) of the form is described in detail. The syntax of the different elements (main verbs, auxiliaries, catenative verbs, modal verbs and DO) is described, as well as their function. Finally, the features used in the analysis of the individual modal verbs are described. The second part of the book (pp. 56–162) describes modal verbs in terms of *a.* modal meaning (17 categories); *b.* time and temporal meaning; *c.* real and non-real states or events; *d.* progressive versus non-progressive meaning; and *e.* grammatical contexts. Each chapter contains a section on Essentials, a description of Prototypes, and a section describing Details about form, meaning and contexts of the modal. Frequency data are supplied both in numbers and diagrams. In the third part of the book (pp. 163–184), the procedure of the analysis is reversed. It uses features *a-e* above as a basis to describe the distribution of different modals. Again, frequency data are supplied both in numbers and diagrams. Finally, an index is provided.

Despite the detailed description of the strategy and procedures used in compiling this grammar, a number of questions arise. The first remarkable fact is that there is no reference whatsoever to the source of the materials that were used. Although this in itself does not affect the reliability of the grammar, it is quite a surprising hiatus for a work of reference that emphasizes its corpus-based nature. The passage ‘... the number and size of available corpora increased steadily. During the final stages of our work we had access to more than 80 million words of English’ (p. 6) could have been supplemented with factual information.

A second comment with respect to the data regards the selection of subcorpora. As the author argues, ‘[t]he selection of subcorpora depends on the purpose of a grammar’ (p. 7). He goes on to explain that his grammar, which has been designed for learners and teachers of English, is primarily based on fictional texts of British English. ‘Fictional texts combine elements of spoken English (in the passages containing direct speech) and of expository prose. They are especially rich in grammatical variation.’ (p. 7). However, one might wonder whether the expository prose and dialogue occurring in fictional texts are in fact of the same register as, say, expository prose in official documents or dialogue in spoken conversation. No formal underpinning is presented for the claim that the type of material occurring in fictional texts is in fact the most beneficial for learners of English.

A third point relating to the use of corpus data regards the exclusively inductive approach. In the preface, the author emphasizes that the grammar is based on original research and that it is not based on previous grammars. While the objectives of his approach are clear, it does have potential consequences for the completeness of the grammar. This is revealed for instance by the description of contexts in which modal verbs occur. The occurrence in tag questions is only mentioned for the modals WOULD, SHALL, NEED, OUGHT and USED TO. Obviously, this is due to the fact that examples of these cases appear in the data. Yet, it would be wrong for a student to infer that, for other modals, no such use is possible. This is precisely the point that is so often made with respect to corpus linguistics in general: a grammar that is based *exclusively* on corpus data can only profess to be a grammar of the corpus, not of the language. To achieve any level of completeness, intuitive data have to be used as well as corpus data. Or, to put it differently, inductive and deductive methods should supplement each other.

The description of the verb phrase presented in *An Empirical Grammar* differs from that usually found in learner grammars in that it reserves a separate structural position for the class of *catenative verbs*. This class, based on formal criteria, comprises verbs like SEEM TO, WANT TO, etc.; lexical verbs taking a verbal complement (KEEP followed by present participle; BEGIN followed by present participle, etc.); catenative auxiliary constructions (BE GOING TO, BE ALLOWED TO); and catenative adjectives (BE ABLE TO, BE LIKELY TO, etc.). The position in the verb phrase, distinct from other auxiliaries, seems to make for a useful structural generalization. To mention an example, the structure of a verb phrase like *can be seen to have started happening* (p. 26) is naturally accounted for in terms of the classes mentioned. The global relationship of the elements within the three-dimensional representation may not be immediately apparent; in fact, a two-dimensional representation is possible. However, the separate sections on each of the triangles make good use of the form.

Applying formal criteria as a basis for classification sometimes results in artefacts in the description. This problem is illustrated by the description of the auxiliary DARE, which is classified as a marginal modal, a catenative verb or a main verb (p. 33). However, given the formal criteria, there are some cases where DARE can be taken either as a catenative verb or as a marginal modal (note 2 on p. 33). In other words, in these cases the formal criteria do not provide sufficient

discriminatory power. In the chapter on the function of elements in the verb phrase, the description of the function of modal verbs and catenative verbs (pp. 42–43) is in fact identical. This, again, may be seen as an artefact of the description, since the classification that distinguishes between them in the first place is based on formal criteria. Such artefacts are hard to avoid, and the fact that they become visible in these cases is a correlate of the clarity of the description.

The description is abundantly illustrated with examples, and that is one of the obvious virtues of the grammar. Due to the fact that the chapters on individual modals contain sections on Essentials, Prototypes and Details, there is some repetition, both of explanatory text and of examples. The terminology used sometimes deviates from what is common; for example, the term ‘auxiliary’ is used only for perfective, passive and progressive auxiliary, and the term ‘modal auxiliary’ is not equivalent to ‘modal verb’. However, terminology is carefully introduced and well-defined. This holds in particular for the definition of the categories used in the analysis of modals, which is given in a separate chapter.

The percentages in the grammar can, but need not necessarily, be used by the reader. For cases where the criteria of analysis do not permit a clear-cut classification (e.g. where meanings are on a ‘gradience scale’, p. 68), one might wonder what the effect of the analysis of percentages is; again, however, this is a problem that cannot be avoided. In the description of the syntactic contexts of the modals, the percentages relate more to formal characteristics, and the issue therefore does not arise. In the third part of the grammar, where the analysis procedure was reversed, the frequency data shed more light on the choice of modal in certain (syntactic and semantic) contexts. It is probably in the latter context that they serve their intended purpose best.

The presentation of the material in *An Empirical Grammar* deserves some special attention. In some ways, it is presented almost as a kind of ‘formal’ grammar. There is a strategy of ‘refinement’: first the more global distinctions are made, then within each global class more specific subtypes are described. Sometimes, the refinement strategy creates a little bit the impression of an ‘inventory’, which on the one hand contributes to the clarity of presentation, but, on the other, has a rather numbing effect on the reader. A similar situation arises in the layout, where different strategies are used to reflect different aspects of the description. This results in a presentation on the page which is very clear in structure (i.e. it is very easy to see what type of information is presented on the page) – but at the same time has the odd consequence

that reading it may be experienced as a somewhat tiring activity. This is particularly the case when explanatory text is accompanied by examples. The effect of layout here is either that the reader has to interrupt the grammatical description in order to process the example, or (more likely) that, after reading the explanatory text, the reader has to move back on the page in order to read the examples. Where a lot of these instances occur, there is a strong inclination to continue reading 'down' and therefore to skip the relevant examples. To me, this example illustrates what I mentioned at the beginning of this review: the very aspect that leads to clarity in one respect, turns out to be a potential disadvantage in another.

An Empirical Grammar of the English Verb is a descriptive grammar based on authentic materials. The description of modals is embedded in a very clear structural description of the verb phrase. Although the emphasis on three-dimensionality sometimes seems somewhat superfluous, the graphs are put to good use in the explanatory text. In some cases, the formal classification leads to artefacts in the description. Presentation and layout convey an immediate impression of clarity, although this clarity, too, has a potentially disadvantageous counterpart. The description of modal verbs and their form, meaning and context is supplemented by a characterization of semantic and syntactic contexts with respect to the choice of modal. Particularly for the learner, this latter component can be a very useful addition to the grammar. Although the exclusively corpus-based nature of the description has consequences for the level of completeness, the book can certainly serve as a useful work of reference for the audience it aims at.

Christine Johansson *The relativizers whose and of which in present-day English: Description and theory* (Acta Universitatis Upsaliensis. Studia Anglistica Upsaliensia, 90). Stockholm: Almqvist & Wiksell International. 1995. 276 pp. ISBN 91-554-3501-7. ISSN 0562-2719.

Christer Geisler *Relative infinitives in English* (Acta Universitatis Upsaliensis. Studia Anglistica Upsaliensia, 91). Stockholm: Almqvist & Wiksell International. 1995. 269 pp. ISBN 91-554-3502-5. ISSN 0562-2719.

Reviewed by **Pieter de Haan**, University of Nijmegen.

The two volumes before me are both reports on studies of different aspects of related constructions, viz. relative clauses in English. Both volumes were presented for public disputation within one week from each other in the spring of 1995, at the university of Uppsala. It seems therefore logical to review the two volumes together.

They have a few things in common. They are both corpus studies in which the authors have studied the Brown Corpus, the LOB Corpus, the London-Lund Corpus (LLC) and the Birmingham Corpus. Both volumes are basically descriptive studies of the phenomenon of particular relative constructions in English, viz. those containing the relativizers *whose* and *of which*, and the so-called relative infinitives.

One difference between them is that Johansson bases her quantitative study on Brown and LOB for written English, and LLC and the spoken part of the Birmingham Corpus for spoken English. She uses the written part of the Birmingham Corpus in order to provide further examples. She has been obliged to do this, one feels, because the constructions occur very rarely in the corpus material, and Brown, LOB and LLC did not provide her with a sufficient number of observations.

Geisler, on the other hand, has found so many instances of relative infinitives in Brown, LOB, and LLC that he did not need to use the Birmingham Corpus other than to provide the odd additional example to illustrate a particular point. He therefore bases his quantitative results on Brown, LOB and LLC only.

I will start this review with a discussion of Johansson. Johansson sets out to provide a comprehensive description of a relatively rare syntactic construction in English, viz. the relative clause with the relativizer *whose* or *of which*, and to relate the results of a corpus study to the theoretical description of this construction within the framework of GB.

The book starts with a fairly extensive introductory section, in which

aim and method, terminology, the corpus texts examined and previous research are discussed. As to terminology, apart from well-established terms like antecedent and relative clause, Johansson introduces a number of terms that she uses to describe specific parts of the construction under investigation. The explanation of these terms is not very clear, but fortunately there is a diagram on p. 17 which clarifies matters. In the sentence *This is the house whose roof / the roof of which was damaged* the **relative group** is *whose roof / the roof of which*. The **relative complex** is *the house whose roof / the roof of which* (= antecedent + relative group). The **relative clause complex**, finally, is the entire sentence, which, according to Johansson, consists of the **antecedent clause** and the relative clause. So apparently the antecedent clause in this example is *This is the house*.

Curiously enough, Johansson does not use a separate term for the string *the house whose roof / the roof of which was damaged*, in other words the noun phrase in which the relative clause functions as the postmodifier, even though extensive reference is otherwise made to Quirk *et al.* (1985) and de Haan (1989), both of whom distinguish this category. This makes the description of the function of these noun phrases later on needlessly confusing, for when Johansson speaks of the noun phrase she does not mean the superordinate noun phrase, but the noun phrase of which the relativizer forms a part, in other words, the string for which she has created the term relative group.

Johansson points out that a comparative study of *whose* and *of which* can only be carried out with reference to non-personal antecedents, as these are the only antecedents that can take either relativizer. Personal antecedents take only *whose* in comparable constructions. All this means that the number of observations on which the study is based is very small. It took me some time to work out exactly how many there are in the corpus texts mentioned above, but I think that the total number of non-personal antecedents with relative clauses containing *whose* or *of which* is 196 for the written parts and 33 for the spoken parts. What is a pity is that Johansson constantly shifts the focus from the use of the two relativizers with non-personal antecedents to a comparison of *whose* with personal vs. non-personal antecedents, and then again to the differences between speech and writing, and so on. This makes it hard, at times, to keep track of what she is doing.

Chapter 1 is devoted to a discussion of the concept **genitive**. A clear distinction is made between genitive constructions and so-called **partitive** constructions (as in *the passengers, 72 of whom were injured*). Johansson

provides ample examples to illustrate the various types of relations that obtain between a noun phrase head and its genitive determiner (or post-modifier). Johansson argues that certain relationships are best described in semantic terms, while others should be described in syntactic terms. She distinguishes seven types of semantic relationship, and two types of syntactic relationships. Among the semantic relationships are part-whole relationships (as in *the house whose roof*), kinship (as in *the dog whose ancestors*) locative (as in *Russia, whose technology*) and genitive of attribute (as in *a building whose significance*). The two syntactic relationships are subjective (as in *cells whose response*) and objective (as in *slogans, whose repetitions*).

Johansson makes a number of interesting observations. For instance, she finds that the relativizer *whose* occurs in 75% of the cases expressing semantic relationships, but that the two relativizers occur with equal frequencies in the cases expressing syntactic relationships. *Whose* occurs in 75% of the subjective genitive cases, whereas *of which* occurs in 75% of the objective genitives. In the semantic genitives, the relativizer *whose* is more often used the higher up the gender scale the antecedent is.

Chapter 2 is a discussion of the two relativizers in written English, whereas chapter 3 discusses the two relativizers in spoken English. Johansson points out that spoken English uses the two relativizers *whose* and *of which* far less often than written English. Therefore, in chapter 4, she discusses a number of alternatives to these constructions that are used in spoken English. The book is concluded with a six-page summary (Conclusions) which sums up the major findings, a bibliography, and an appendix with text categories of the corpora studied.

Chapter 2 provides information about the frequency of occurrence of the two different relativizers, related to different antecedents, text types, form and function of the antecedent, form and function of the relative group, and function of the relative clause complex, by which, of course, is meant the function of the superordinate noun phrase in its next higher constituent. These are exemplified with many corpus examples.

One of the conclusions is that the two relativizers occur most frequently in informative prose, and that *whose*, especially, is found most frequently in category J (both in Brown and in LOB). As to form and function of the antecedent, it is observed that the antecedents of both relativizers are, especially, inanimate abstract nouns, often functioning as objects or prepositional complements, and often indefinite.

Section 2.6 is partly a repetition of section 2.4, in that some of the conclusions of 2.4 are presented again. What is interesting to observe

is that relative groups most often function as subjects in the relative clause, and that the relative clauses in question are for the greater part non-restrictive clauses. Johansson also comments on the fact that relative clauses with *whose* / *of which* do not refer to the antecedent in the same way as those with *who* / *which*. The use of the genitive introduces a new entity. Therefore, these clauses cannot refer to or restrict the reference of the antecedent in the same way as other relative clauses.

In section 2.7, the various positional variants of *of which* are discussed against a theoretical description of this phenomenon within the framework of GB. Johansson distinguishes three variant positions:

- a. (the car) the brakes of which he repaired
- b. (the car) of which he repaired the brakes
- c. (the car) which he repaired the brakes of

The (a) variant occurs most frequently overall, and any amount of variation is found only in subject relative groups. The (c) pattern is not found in the corpus material. A further examination of some of the Birmingham material suggests that complex, modified relative groups favour the (b) pattern.

Chapter 3 has a similar set-up to chapter 2. The last section of chapter 3 compares written English with spoken English. Johansson observes that *whose* and *of which* are especially found in those spoken text categories which resemble writing most (i.e. lectures and prepared speeches). On the whole, the findings for spoken English are not so different from those for written English. The most important difference seems to be that the two relativizers in question occur roughly five times as often in written English. Johansson suggests that this is almost entirely due to the fact that spoken English uses *that* far more often, and that, had *that* been included in the study, the numerical differences between the two media would have been negligible.

Chapter 4, finally, discusses possible alternatives in spoken English to the relative constructions under discussion, both relative and non-relative. Relative alternatives comprise both possessive and non-possessive relativizers. Johansson discusses some of these alternatives, which she finds in the corpus, and shows that they can be replaced with relative constructions with *whose* or *of which*. However, she admits that the alternatives are often not the same with respect to informational value.

All in all, the book is a solid report on the study of an interesting syntactic construction. However, it suffers from the fact that the material

studied yielded so few observations. This means that not much statistical significance can be attached to the figures presented in the book. It also means that the reader cannot help feeling at times that s/he is presented with the same examples over and over again.

There are also a number of flaws. I have already mentioned the somewhat muddled presentation of the terminology in the introduction. Furthermore, the tables are at times confusing. Some of the tables contain subtotal scores which are sometimes, but not always, represented in bold type. In other tables, bold type is reserved for a specific variable which is compared to other variables. In yet other tables, the composite figures for the subtotals are presented in brackets, but again this is not always the case. Also, in some tables, percentage scores are presented beside the raw or standardised scores. But there are also tables where the odd percentage score is thrown in.

What I personally also find confusing, and, frankly, a little off-putting, is the presentation of the spoken examples. Most of them contain all the prosodic information as it is found in the corpus. As someone who is not accustomed to studying this material, I find this confusing and not seldom have I misread */and* for *land*, for instance. Besides, since the prosody of the examples is not considered in the first place, I wonder if the examples should not have been presented without the prosodic information.

To end this review on a positive note: Johansson has studied her material very extensively, against the theoretical description of the phenomenon. Moreover, she has considered her examples with reference to such matters as textual structure and informational value. For this reason alone, her study is a valuable contribution to the literature on relative clauses.

Geisler's is a corpus-based study of relative infinitives in spoken and written English. After an introductory chapter, in which the aim, the corpora, the statistical tools and the terminology used are explained, and a historical perspective is presented, the study goes on, in chapter 1, to a typology of relative infinitives. A number of antecedent functions are distinguished, to which I will return below. Each of the following three chapters is devoted to a discussion of one of the three major functions. Chapter 2 discusses the subject relatives, chapter 3 the object relatives and chapter 4 the adverbial relatives. The book is concluded with a five-page summary (Conclusion), and a bibliography.

Geisler introduces a few terms which are relevant to this study. Unlike Johansson, he uses the term **relative complex** to refer to the noun

phrase consisting of the antecedent and its infinitive postmodifier. The term **antecedent function** is used for a very specific reason. As the relative infinitive usually has no overt relativizer (unlike finite relative clauses), the relationship between the antecedent and the infinitive is established directly. This is in most cases the same as the function of the overt relativizer *who* or *which* in the corresponding finite relative clause. For instance, in the sentence *He was the last man to enter*, the antecedent function is subject (cf. *He was the last man who entered*). This distinction enables Geisler to distinguish properly between the function of the relative group as a whole (called **matrix function** in this study) and the function of the antecedent within the relative infinitive.

The term **control** is used to refer to the relation between the cognitive subject of the infinitive and its referent. Geisler rejects the term **antecedent**, which has sometimes been used for this purpose, and suggests the term **controller** instead. In the sentence *I'll have plenty of time to deal with them*, the matrix subject *I* serves as the controller of the understood subject of the infinitive.

Discourse functions of relative clauses are discussed in terms of **grounding** and **anchoring**. Grounding is a more general term. It refers to the fact that a referent is functionally related to another referent. Anchoring is a special type of grounding: it refers to the linking of a new referent in the noun phrase to some other referent inside the same noun phrase.

Topic continuity, finally, refers to noun phrases inside the infinitive clause, whose reference is uniquely identified. This notion is used in the light of hypotheses about information flow and the organization of given and new information.

Chapter 1 distinguishes four major types of antecedent functions. They are **subject**, **object**, **adverbial** and **prepositional complement**. On p. 33 there is a useful table (1.1) which summarises the various syntactic patterns of these function types. It is shown, for instance, that subject antecedents (as in *the latest dominant type to arise*) do not allow any syntactic variation, whereas the others all do, to a greater or lesser extent. Object relatives, for instance, allow passive infinitives and the presence of an overt subject introduced by *for*. It should be noted that Geisler uses the term **object** also to refer to what might otherwise be called the subject of passive infinitives. This is the only way, he argues, in which these constituents can be distinguished from 'real' subjects, i.e. subjects of active infinitives.

Subject relatives, on the other hand, are the only ones that allow

non-modal interpretations. Adverbial antecedents are further subdivided according to their meaning into locative, temporal, manner and instrument adverbials. A number of theoretical possibilities of realization types are discussed. Three of them are found in the corpus material, and are referred to later on as patterns 1 (as in *the place in which to stay*), 4 (as in *the place to stay in*) and 6 (as in *the place to stay*). Again, a table (1.2, on p. 46) shows how the various adverbial categories are distributed by their structural patterns.

Chapter 2 discusses particularities of the subject relatives. Attention is paid to the different patterns presented by intransitive and transitive infinitives. The latter outnumber the former by 87% to 13%. Also the status of the relative complexes, in terms of definiteness or indefiniteness, is discussed. Next, the various constituents of the relative complex are discussed, viz. the determiner, the premodifier, the head, and any other postmodifier (usually a prepositional phrase). This same basic set-up is used for chapters 3 and 4. Of course, the various special features of the individual types are only discussed in the chapters which are relevant to these types.

For instance, chapter 2 contains a discussion of the modality of the subject relatives, as we have just seen that they were the ones that also allowed non-modal interpretations. Chapter 3 has a separate section devoted to the possibility of having either active or passive infinitives with object relatives. Chapter 4, finally, has a separate section discussing the structural patterns involving occurrence of and the various positions of the preposition and the relativizer.

Geisler provides a detailed analysis of the material he has found in his corpora. He presents his findings in a clear way, but I have come across a number of rather serious flaws. They all bear on the analysis of the structures discussed. I must mention some of them.

On p. 41, he confuses a phrasal verb with a prepositional verb and calls *another job to take on* an example of an active verb phrase with postposed preposition. On the same page, he calls *anything for which to be hated* an example of a prepositional verb with a preposed preposition. On p. 62, he analyzes *thinking it the ideal climate in which to begin battle and bloodshed* as an example of anticipatory *it*, which, moreover 'refers back [sic] to the object of the previous clause'.

On p. 128, the following example is quoted: *Among the newest fabrication methods to enter the display field are expandable styrene molding and blow molding*. This is said to be an example of a subject relative in a matrix subject, whereas it is clear that the matrix function

of the noun phrase in question is prepositional complement. Exactly the same type of mistake is made on p. 196: *Among the first to react was the headmaster...*

On p. 162, finally, he mentions finite and non-finite clauses that are found following the relative infinitives in some cases. However, he fails to distinguish between those cases where the following clause is embedded **within the relative infinitive** and those where the relative is embedded **within the matrix sentence**, which is quite different, of course.

These flaws spoil an otherwise solid description of relative infinitives to a certain extent. It makes the reader a little suspicious of the reliability of the statistics presented.

The same point I raised in connection with Johansson about the presentation of the spoken examples also applies here. In not a single case is the prosody of these examples discussed. Therefore, I would have thought that the prosodic marking might as well have been left out, but as I said before, this may be a very personal thing.

What I looked for in vain in Geisler's book is a discussion of any possible consistent differences between restrictive and non-restrictive relative infinitives. Geisler does not mention that the non-restrictives have been excluded from the study. Especially in connection with the object relatives, I would have expected there to be a consistent pattern showing that, in non-restrictive relative infinitives, the passive would be preferred, or perhaps even obligatory, as this is what I have frequently found myself.

Despite these criticisms, I think that Geisler's study, like Johansson's, has made a valuable contribution to the description of relative clauses in English. We now have much more detailed information about various syntactic and pragmatic aspects of relative infinitives. Geisler's study is particularly useful as it has been based on the study of more material than Johansson's, although he has been shown to be somewhat inaccurate in his syntactic analyses at times.

References

- de Haan, Pieter. 1989. *Postmodifying clauses in the English noun phrase: A corpus-based study*. Amsterdam: Rodopi.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.