tive states (expressed by noun pairs such as *life* and *death*) as complements without a resultant change in meaning.

The last chapter, 'Consequences of a generative lexicon', further discusses the effects of co-composition and coercion on causative-inchoative verbs like *open* and *break,* stage level vs. individual level predicates, temporal conjunctions like *before* and *during,* and prepositions. Finally, P makes a brief excursion into discourse analysis, where coercion and co-composition as well as contextual effects such as the structure of rhetorical relations in discourse and pragmatics place constraints on coreference. The reader is left with the impression that the GL allows semantic structures to be unpacked like Russian dolls yielding an infinite number of meanings. The book successfully demonstrates the power and exciting possibilities of this new research paradigm in lexical semantics.

Cognitive Science Laboratory
Princeton University
221 Nassau St.
Princeton, NJ 08542
[fellbaum@clarity.princeton.edu]

English for the computer: The SUSANNE corpus and analytic scheme. By GEOFFREY SAMPSON. Oxford: Clarendon Press, 1995. Pp. ix, 499.

Reviewed by D. TERENCE LANGENDOEN, *University of Arizona*

This book documents the SUSANNE corpus,[1] a full syntactic annotation of the Gothenburg corpus (Ellegård 1978), comprising 64 of the 500 texts (approximately 130,000 words) in the Brown corpus (Francis & Kučera 1989 [1964]), 16 in each of the four Brown genre categories: press reportage, belles lettres, technical prose, and adventure fiction. It is a small corpus by contemporary standards (cf. Oostdijk 1991, Marcus et al. 1993), requiring only 5.3 megabytes of storage, but the detail of its annotation is unrivalled.

Each word and punctuation mark in the corpus is tagged with one of nearly 400 distinct 'wordtags' (105–20, 447–8) based on, but greatly extending, the 1985 version of the Lancaster tagset, with 166 members (Garside et al. 1987:165–78). Provision is also made for the further tagging of certain words with a reference to appropriate entries in Hornby et al. 1973, more precisely to the computer-usable dictionary based on that dictionary's typesetting tape (Mitton 1986).[2] Then, for each paragraph in the corpus (44), a tree diagram is constructed in which each phrase, clause, and other supraword constituent is tagged with one of several hundred 'tagmatags' (168–70)[3]. Moreover, all wordtags and tagmatags may be further elaborated by various suffixes (170–1). In addition, one of 23 'functiontags' (362–3) is assigned to each of certain constituents of clauses 'to

[1] SUSANNE stands for 'Surface and underlying structural analyses of natural English'. Work on the corpus began in 1988 and is continuing. The current version of the corpus and further documentation are available from the Oxford Text Archive.

[2] The scheme for providing sense codes tied to the electronic version of Hornby et al. 1973 (67–73) is not included in current SUSANNE releases.

[3] Since certain of the characters used to define the 79 listed tagmatags can be combined to create additional ones (e.g. given the tag *Ns* designating a noun phrase marked as singular, and *Nn* designating a proper name, one can form the tagmatag *Nns,* a proper name marked as singular), there are at least several hundred legal tagmatags.

represent logical properties of the respective constituents' (354), and '[e]xtra nodes dominating no wording ("ghost nodes") are added to the parsetrees to represent the logical position of elements that have been moved or deleted in surface structure' (353). Finally, '[n]umbers, called "indices", mark the relationship between nodes marked grammatically as counterparts, such as a ghost and the corresponding full surface constituent' (354).

For example, (1) gives the full annotation of example G06:0950 (42), and (2), the supraword annotation of example N04:0560 (356).[4] In the latter, *:G* is a functiontag indicating a 'guest node', a constituent of a tagma that corresponds to a ghost node in another tagma; *135* and *136* are indices; and *136* (occurring by itself) and *o135* are ghost nodes.

(1) [O [S? [Rq:q [RRQq *Where*] [P [II *in*] [Nns [NP1g *Europe*]]]] [Vosb [VBDZ *was*]] [Nas:s [PPHS1m *he*]] [Vrg [VVGi *going*]] [Rw:t [RTo *now*]]] [YQ ?]]
(2) [S *It was* [Ns:e135 *a* [J *terrible* 136] *thing* [Ti:G136 *to do* o135]]].

As these examples show, constituent structures in the SUSANNE corpus are relatively flat. They posit neither the intermediate phrase nodes of the X-bar theory nor the functional projections of the minimalist program,[5] and certain apparently discontinuous constituents such as *was ... going* in (1) are not tagged as such.[6] In addition, phrase nodes corresponding to certain lexical nodes are also not posited (175–6, 302–4).[7] Sampson's structural conservatism is based on his observation that there is little agreement about the correct assignment of constituent structures to naturally occurring sentences of English (4–5). He contends that with the exception of the treatment of coordination, the higher level structures marked in the SUSANNE corpus are relatively uncontroversial.[8]

S's one structural innovation is to assimilate coordination to subordination: 'the second and any subsequent conjunctions of a co-ordinate structure are treated as subordinate to the first, "main" conjunct' (310–1). For example, (3) shows the annotation of the phrase N13:0490.

(3) [N *the heat* [Ns+ *and the dust*]]

Thus, 'within the co-ordination, *the heat* is not a tagma' (311), nor is *the dust*. Whether or not S's innovation is well-motivated structurally, it is easy to recode his analysis into a more standard one such as (4) in which both *the heat* and *the dust* are constituents of *the heat and the dust*.

(4) [N [Ns *the heat*] [Ns+ and [Ns *the dust*]]]

[4] Example numbering in the SUSANNE corpus corresponds to that of the tagged Brown corpus (Francis & Kučera 1989). The notation in examples 1 and 2 is not precisely that of the SUSANNE corpus file but a shorthand that S uses for text displays. I also omit labels on the right brackets.

[5] S does not consider constituents based on X-bar theory or the minimalist program but does discuss the highly articulated constituent structures of generative semantics (352) and concludes that they are not appropriate for a syntactic encoding standard such as SUSANNE.

[6] Though the separate phrase-level tagging of *was* and *going* as *Vosb* and *Vrg* respectively could be used to reconstruct them as a unit by a program which uses SUSANNE tagging as input.

[7] Some of S's rules for determining whether a single word constitutes a phrase of its own are rather delicate and seemingly arbitrary. For example, a single noun is analyzed as a noun phrase if it is an immediate constituent of a clause but not if it is an immediate constituent of a prepositional phrase (176).

[8] 'The treatment of co-ordination is probably the most nonstandard aspect of the SUSANNE parsing scheme' (310).

Another area of innovation in the SUSANNE tagging scheme is its treatment of names. S observes that the category of proper name is appropriately assigned at the phrase, rather than at the word, level. Accordingly, the wordtag that is assigned to an entry in Hornby et al. 1973 which occurs as (part of) a proper name is the one appropriate for that entry; for example the word *Flagstaff,* when occurring as a proper name, is tagged both as a singular common count noun, NN1c, and as a proper noun phrase, Nn ... (87). SUSANNE encoders have, in fact, exercised considerable ingenuity in the tagging of proper names, as illustrated by how they have tagged the names of the American states (146–8).

One aspect of the SUSANNE parsing scheme which 'has proved (unexpectedly, from the author's point of view) to be quite controversial' (80) is that 'a SUSANNE wordtag is a string of characters which is intended as a single atomic symbol, not as shorthand for a set of grammatical features' (79). Although 'sets of tags for words having some common grammatical property do often have a character in common [,] ... this is a matter merely of practical convenience; it does not imply that there is some specific set of binary- or multiple-valued grammatical features underlying the tagset such that any individual wordtag can be completely translated into a distribution of values over the various features' (80). As a matter of fact, however, the SUSANNE wordtag set CAN be translated into a set of FEATURE STRUCTURES, each of which is a bundle of features whose values are binary- or multiple-valued features or other feature structures.[9] The latter is significant since it overcomes S's major objection to the use of sets of features to encode SUSANNE wordtags, namely that a common grammatical feature may have different interpretations in different contexts; for example, that 'for a possessive pronoun to be plural (e.g. *our* v. *my*) has nothing at all to do with verb agreement' (81). Moreover the overall design of the SUSANNE tagging scheme incorporates something like feature structures since the class of tagmatags is defined to allow certain characters, which can be understood as feature values, to be combined (see note 3); and other symbols, also interpretable as feature values, can be suffixed to both wordtags and tagmatags.

S points out that the SUSANNE corpus includes a parsing SCHEME, not a parsing SYSTEM, since SUSANNE tagging is too complex to be done automatically (4). Moreover, S does not provide a grammar for defining the class of well-formed SUSANNE annotation structures. The latter is, perhaps, achievable, and it would be a useful exercise to attempt to come up with one, if only to discover hidden inconsistencies in the present scheme, and to provide validation for the tagging of other corpora using the scheme. As it stands, however, the SUSANNE corpus is a very useful resource for the linguistic research community, and the SUSANNE Team and the funding agencies which have supported it richly deserve our thanks.

REFERENCES

ELLEGÅRD, ALVAR. 1978. The syntactic structure of English texts (Gothenburg studies in English 43). Gothenburg, Sweden: University of Gothenburg.

FRANCIS, W. NELSON, and HENRY KUČERA. 1989. Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers. Corrected and rev. edn. Providence, RI: Brown University (1st edition, 1964).

[9] See Sperberg-McQueen and Burnard 1994:515–9 for a feature-structure re-encoding of part of the wordtag set currently being used for the encoding of the British National Corpus. For objections to the use of sets of atomic tags for the encoding of grammatical structure, see Langendoen and Simons 1995.

Garside, Roger, Geoffrey Leech, and Geoffrey Sampson (eds.) 1987. The computational analysis of English: A corpus-based approach. London: Longman.

Hornby, A. S., E. V. Gatenby, and H. Wakefield. 1973. Oxford advanced learner's dictionary of contemporary English, 3rd edn. Oxford: Oxford University Press.

Langendoen, D. Terence, and Gary F. Simons. 1995. A rationale for the Text Encoding Initiative recommendations for feature-structure markup. Computers and the Humanities 29.191–205.

Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. Computational Linguistics 19.313–30.

Mitton, Roger. 1986. A partial dictionary of English in computer-usable form. Literary and Linguistic Computing 1.214–5.

Oostdijk, Nelleke. 1991. Corpus linguistics and the automatic analysis of English. Amsterdam: Rodopi.

Sperberg-McQueen, C. Michael, and Lou Burnard (eds.) 1994. Guidelines for electronic text encoding and interchange (TEI P3). Chicago & Oxford: Text Encoding Initiative of the Association for Computers and the Humanities, Association for Computational Linguistics, and Association for Literary and Linguistic Computing.

Department of Linguistics
University of Arizona
Tucson, AZ 85721–0028

Beyond names for things: Young children's acquisition of verbs. Ed. by Michael Tomasello and William E. Merriman. Hillsdale, NJ: Lawrence Erlbaum, 1995. Pp. vi, 421.

Reviewed by M. Lynne Murphy, *University of the Witwatersrand*

Semantic acquisition is a slippery area to study since the objects under investigation, meanings, are much less observable than other aspects of language such as phonology or syntax. In studying meaning, acquisitionists have traditionally focused on object names thus limiting their study to meanings that have concrete, temporally stable counterparts in the real world. This collection emphasizes the role of verbs and verb-like words in early language and the problems that verb-learning creates for theories of acquisition. While the authors draw upon the literature on verb-learning in the past fifteen years, this is not a collection of greatest hits. Rather, it is current work and, in some ways, future work, since a consistent theme is how much and what kind of work must be done in order to answer the questions that the book raises.

The fourteen chapters are quite diverse, right down to their definitions of *verb*. In their introduction ('Verbs are words too', 1–18), the editors provide some history, examining why verbs have been neglected and why they are suddenly of interest. While verb acquisition is more challenging to study than concrete noun acquisition, recent developments in linguistics and psychology make it more necessary and easier to take on verbs. In linguistics in particular, syntactic issues are increasingly resolved in the verb's lexical entry, and relations between syntax and semantics in acquisition (and elsewhere) have been realized to be more intricate than previous theories had allowed. Thus, verb acquisition is a major part of syntactic acquisition.

From here, the book is divided into three sections. Part I, 'Early words for action', begins with Patricia Smiley and Janellen Huttenlocher's 'Conceptual development and the child's early words for events, objects, and persons' (21–61), a dense but well-organized overview of the roles of input and conceptual development at the one-word stage. While they find evidence that both elements guide and constrain word learning, they suggest that mismatches between child and adult meanings reflect clashes