might call the "world knowledge/deep reasoning" school of thought, which is the current paradigm in AI, engendered by Charniak's dissertation. He introduces frames, scripts and plans, MOPS, and information formats (from the Linguistic String Project, NYU). The chapter concludes with a discussion of handling dialogue. Although there is a good discussion of some points, this chapter is somewhat disappointing because many areas of research are left out or no examples are given: classic pragmatics is not discussed at all, story and text grammars are only mentioned, MOPs are only mentioned, and no mention is made of rhetorical structure theory. Perhaps the author felt that this area was not very well settled and therefore should not be treated extensively. Graduate students, however, should be treated to open problems as well as to (purported) solutions.

The language generation chapter is quite brief and treats sentence generation from a syntactic point of view. A short discussion on text generation completes the chapter. Clearly, more could be done with generation even if it is "the poor cousin" of automated language research.

Grishman spends little time trying to fill in the technical linguistic background of the reader. There are short introductions to the Chomsky hierarchy of grammars and predicate logic, but the reader is mostly presumed to be conversant with linguistic theory. The book is, after all, about *computational* linguistics. The amount of introductory material seems reasonable relative to the expository material. For computer science graduate students, however, this account would have to be supplemented with more introductory linguistics material. The book also contains little philosophical discussion. The author does not attempt to address the problem of defining meaning, let alone the role of meaning and language for the concept of mind (a hot topic for some cognitive psychologists and philosophers). Admittedly, such discussions can become abstruse and perhaps not very useful, and Grishman typically avoids such topics.

The author frequently writes short evaluations of previous efforts in an attempt to weigh the contribution of the work. The evaluations seem very fair; Grishman does not take doctrinaire stances toward different approaches to natural language, an excellent way to write an introductory text. One nice touch of the book is the pursuit of topics through several different chapters. For example, the problem of quantifier ordering is discussed both in the semantics chapter as a representation problem and in the language generation chapter. Various such themes are pursued throughout the book.

In summary, this introductory text takes a stand, a point of view, toward the computational linguistics field. It can be used most profitably in classes with students who already have some linguistics background. If supplemented with additional material, it could also be used for courses populated by computer science students. The book is interestingly written with many insightful discussions, and it is the only (introductory) computational linguistics textbook that looks at the field from a linguist's point of view.

References

Tennant, H. 1981 *Natural Language Processing.* Petrocelli Books, New York.

Winograd, T. 1983 *Language as a Cognitive Process: Volume I: Syntax.* Addison-Wesley, Reading, MA.

John Roach is an associate professor of computer science at Virginia Tech. He works on the influence of social factors on understanding in dialogue and text and with extensions of semantic grammars. Roach's address is: Department of Computer Science, Virginia Tech, Blacksburg, VA 24061. E-mail: roach@vtopus.cs.vt.edu

---

Speaking: From Intention to Articulation

Willem J. M. Levelt
(Max-Planck-Institut für Psycholinguistik)

Cambridge, MA: MIT Press, 1989, xvii + 566 pp.
(ACL–MIT Press Series in Natural- Language Processing)
Hardbound, ISBN 0-262-12137-9, $39.95 (20% discount to ACL members)

Reviewed by
M. Martin Taylor and Insup Taylor
Defence and Civil Institute of Environmental Medicine and University of Toronto

OVERVIEW

In *Speaking,* Pim Levelt has written a majestic book. It is majestic in its scope, in the power and elegance of its writing, and in the regal authority with which it describes cognitive structures and processes that are almost inaccessible to experimental study. It discusses in 12 chapters how a speaker produces fluent speech, from concept and intention to articulation. It is hard to believe that one person could attempt such a task, let alone that he could complete it so well. The book is intended primarily for advanced students in psycholinguistics, but we find it to be better suited to students of computational linguistics, since we think Levelt has described how speaking could be done, but not necessarily how it is done by humans.

The book contains many technical terms and concepts, clearly introduced and used, and avoids statistics and computer jargon. Very occasionally, traces of Dutch creep in, though not obtrusively. *Blackboardchalk, passingnote,* and *swimmingwater* are given as examples of English compound words; *bank* is said to have as its two meanings "financial institution" and "furniture". Despite these small lapses, Levelt's knowledge about English (as well as his ability to use it) far surpasses that of most English speakers untrained in linguistics.

The book is well designed and carefully proofread. There is a good index, and more than 600 references. We found only three misprints in 500 pages of text, as well as one "cognitive misprint" that is amusing because it occurs in a discussion of similar phenomena in speech (the commen-

tary on Example 8, p. 248). Didactically, every chapter starts with a discussion of what each section says, and finishes with a good summary. Many sections repeat this plan.

Overall, despite the deeply technical nature of the content, the book is easy to read and persuasive, perhaps seductive, in its argument.

THE MODEL

The model Levelt describes for the production of speech consists of three major modules plus a store of declarative knowledge called the **Lexicon**. In addition, the speech recognition system is called into play for monitoring the resulting speech. A module called the **Conceptualizer** takes the ideas to be transmitted and structures them into a formal grammar much like a phrase-structure grammar, that uses types such as EVENT, PLACE, and PATH as its elements. The result is a preverbal message, which is passed to another module called a **Formulator**, which has two components, a grammatical encoder that creates a sentence pattern (surface structure), and a phonological encoder that produces a "phonetic plan." The phonetic plan is the output of the Formulator.

The Formulator (and only the Formulator) has access to the Lexicon, which contains all known words in the form of **Lemmas** and **Forms**. Lemmas are defined as bundles of declarative knowledge about a word's meaning and grammar (p. 236), and are used by the grammatical encoder. Forms contain similar knowledge about the morphology and phonology of the words, and are used by the phonological encoder. The phonetic plan that the Formulator produces is the input to the final module, the **Articulator**, which turns it into audible (overt) speech that an experimenter or observer can study.

An important feature of the design is that no feedback is allowed to occur between the modules. Grammatical and phonological encoding are placed within a single module (the Formulator), because there must be a small amount of feedback between them. A second design feature is that the model works in an incremental fashion. Items are output by a module as soon as the required information is available, and are (normally) used immediately by the next module; considerable ingenuity goes into making designs for the different components that minimize the amount of look-ahead required.

Speech errors are a major source of information about the way we speak, and Levelt uses error data to check many aspects of his model. But he makes little mention of the apparent fact that errors that satisfy the context at more than one level are preferred to errors that do not (e.g. Freudian slips tend to satisfy conceptual, syntactic, and phonological contexts). Such an acknowledgment might cast some doubt on the modular separation of the speaking processes. Consider *I'll go shut up the darn bore* for *I'll go shut up the barn door* (Dell 1988). At the phonological level, two initial speech sounds of the two words, *barn* and door, are exchanged; at the word level, *darn* and *bore* are real words, not nonsense words; *darn bore* is a cliché; and finally, the speaker was thinking of, or wishing to shut up, a darn bore.

The book, especially the central section, is based on "theoretical notions . . . [that] mostly stem from linguistics and computer science" rather than from psychology. Levelt might well have added "from the symbol-manipulation school of computer science." From time to time, especially in the later chapters in which the processes are near to overt speech, he does refer extensively to psychological studies that are consistent with the machinery being constructed, but it is never clear that the studies force particular constructions as opposed to alternative possibilities, even within the symbolist paradigm.

PLAN OF THE BOOK

Chapters 1 and 2 form an introductory section, presenting an overview of the model and setting the activity of speaking into its conversational context.

Chapter 1 describes the modular structure of the speaking machine and shows the actions of the different modules through the analysis of one utterance in a conversational fragment. Chapter 2 concludes the introduction to the book. It deals with the interaction between two talkers, showing the context in which most speech is produced, and introducing the concepts of communicative intention and of shared situational context between the conversational partners. The chapter describes such "standard" topics as turn taking, Grice's (1975) cooperative principle and its four maxims (e.g. quality: do not say what you believe to be false), and speech acts (a speaker's utterance such as "A train is coming" communicates her intention such as warning or informing, as well as conveying a proposition that can be true or false). However, the chapter is not strong on the local and global coherence found in conversation.

The rest of the book is concerned with how concepts within a talker are turned into sound patterns. Chapters 3 and 4 deal with the Conceptualizer, 5 to 10 with the Formulator and its interaction with the Lexicon, 11 with the Articulator, and 12 with the problem of self-monitoring and self-repair. It is only in Chapter 12 that the listener reappears, in a trivial way, to provide signals to the talker that self-repair may be necessary.

Chapter 3 describes the structure of a preverbal message, and relates the message to the nonlinguistic representation and manipulation of concepts. Preverbal messages are constructed only out of propositional representations, not from spatial, kinaesthetic, or other forms. These representations must be turned into a propositional form before the speech production system can do its work. A cognitive concept contains, in parallel, several relations among a group of entities. *John visited Mary in hospital and gave her flowers* is an ordered presentation of a situated event that could easily have been presented as *The flowers Mary got in hospital were given to her by John.* The processing modules frequently must order as a sequence things that are presented in parallel.

The Conceptualizer turns a propositional representation of a concept into a "message," which refers to events in the world and predicates (e.g. assertions, denials) about them. It has propositional structure, function/argument structure, and thematic structure.

Chapter 4 deals with the generation of a preverbal message from a communicative intention, and it is with this chapter that the serious work of describing the speaking process begins. If Simon wants to ask Hanna *Will you buy a stamp?* he uses procedural knowledge of how to convert patterns of goals into messages. Simon's procedural knowledge is formally analyzed as:

if the goal state is
    KNOW (H, INTEND (S, INTENDED (H, DO (A))))
then encode message
    ? (FUTURE (DO (H, A))))

where A is the intended action and H the hearer. That Simon has this procedural knowledge is inferred from the request he produced.

Chapters 5 to 7 deal with the first part of the Formulator, the grammatical encoding from preverbal message to surface structure. Grammatical encoding takes a preverbal message as input and delivers a surface structure as output. It is lexically driven; it generates incrementally, from left to right; and it generates major sentence constituents in parallel.

Chapter 6 describes the structure of entries in the Lexicon. Each word has a four-component entry: meaning, syntactic use, morphology, and phonology. Meaning and syntax are used by the grammatical encoder, and morphology and phonology by the phonological encoder. A word is selected because it fits semantically with the concept to be encoded. For example *give* (p. 189) is

```
EVENT (CAUSE,
        PERSON (X),
        EVENT (GO[[poss]],
                THING (Y),
                PATH (FROM/TO,
                        PERSON (X),
                        PERSON (Z))))
```

where it is important that the three arguments X, Y, and Z be PERSON, THING, PERSON. But then in a footnote, Levelt says that this might be too restrictive, and cites *The bright lights gave Santa's arrival a colorful appearance,* which has none of the required argument types.

The Santa example shows one of the difficulties of Levelt's symbolist approach: the concept represented in the Santa example as *give* has little in common with the *give* in *John gave Mary a book, John gave Mary a cold, John gave Mary a hard time, John gave up,* or *John gave up the ghost.* There is, to be sure, some semantic overlap among the six concepts, but it is hard, if not impossible, to give (a seventh use of *give*) a context-free definition of *give* that will satisfy all of them without at the same time applying to concepts that are not well represented by *give.* There is no context-free mapping from concept to word in natural language. Therefore the matching of conceptual specification such as that presented above for *give* to the lexicon entry for *give* can be neither a necessary nor a sufficient condition for the selection of *give* in overt speech. But in Levelt's design, all hinges on this selection from the Lexicon. One worries.

Chapters 8 to 10 move from the surface structure generated by the grammatical encoder to the **phonetic plan**, an internal form of speech in which all the phones are sequenced preparatory to feeding the Articulator module. In this section of the book, there is much more consideration of real psychological data, largely observations of natural or induced speech errors. The types of errors that do and do not occur are illuminating, and do provide tests of possible models, though the tests may not be as strong as Levelt seems sometimes to claim.

Chapter 11 deals with the actual process of articulating the phone string output by the Formulator. It is here that there is most experimental evidence, and it is here that Levelt acknowledges most frequently doubt about what actually happens. Perhaps for this reason, the chapter is one of the most interesting in the book. There is perhaps a little more emphasis on cognitive control as compared with mechanical dynamics than one might like; for example, CV formant transition rates do not change much, if at all, with speaking rate, and yet Levelt uses lack of control of transition rates as an argument against some theories.

The final chapter, Chapter 12, has a very interesting discussion on the linguistic constraints that govern the forms of self-repair. It is, of course, necessary that any self-repair be marked as such for the listener. To ensure that this marking occurs, the talker must use signals that are known to the listener. These signals must either be universal and inbuilt or culturally and linguistically determined. Whichever they are, they must be readily detected, and thus should be possible to describe linguistically. Levelt shows that they are, and describes a few of the principles that allow the listener to determine which parts of the already-received message should be altered. He also uses some of the timing effects of self-repair to assist in determining at what stage of the speech process the error is detected and at what stage it is repaired.

And then the book stops.

An epilogue is probably not required, but when one turns from page 499 to page 500, one is surprised to find not a summation, but a table of IPA phonetic symbols. To complain about this abrupt stop may seem like carping, but in such a well-written and well-paced book, one expects an elegant completion. Rather, it feels as if the contract said that there should be no more than 500 pages of text, and that's where it had to stop.

PROBLEM AREAS

Speaking is a book that lies between engineering and science. It describes brilliantly how an intelligent speaking machine might be built using concepts that computer scientists now understand, and it points out ways in which the speech of the machine would be like human speech. But the book purports to be more than that: it claims to describe in considerable detail how humans do speak. To support this claim requires access to more data than present-day experiments can provide.

It is intrinsically hard to study experimentally how humans speak in a natural conversational context. The mere fact that experimental conditions are imposed on the talkers is enough to ensure that the conversation is unnatural, since the talkers are pursuing goals imposed by the experimenter rather than goals of their own. This difference has an immediate impact on many aspects of the conversation, especially on the emotional involvement of the participants, which may be reflected in aspects ranging from intonation and word choice to the global coherence of the topics and the linking structures used to maintain that coherence. Researchers in automatic speech recognition are well aware of these effects.

If the study of speaking cannot be an experimental science like physics and chemistry, it must be an observational science like astronomy and geology. But just as physics and chemistry can help us to understand the observations of astronomy and geology, so can experiments on the microstructures of speaking help us to understand our observations of natural speech. Physics is a much simpler science than psychology, and the interior of a supernova much simpler than that of a human brain. How likely is it, then, that the complex structures and processes described by Levelt actually occur in the brain when people speak?

If we treat the book as an engineering design for a speaking machine, rather than as a scientific description of how humans really speak, we are on safer ground. Levelt, perhaps unconsciously, acknowledges as much in the difference between his treatment of processes close to the overt production of sound and that of deeper, more "cognitive" processes. For the peripheral processes, about which experimental results are available, he presents several models and frequently comments that we just don't know the real answer; but for the deeper processes that require much speculative interpretation of observations, he presents a formal structure and in most cases asserts that this is the way things are. It is hard to accept that we can be more assured of the processes for which we have little information than about those for which experimental data are available.

The first main problem area is thus that the descriptions in the book can be considered only as plausible, not definitive as Levelt seems to claim.

The machine Levelt has designed is a parallel and incremental production-rule system. Rules are executed as soon as their conditions are true and not otherwise. This commitment to a binary logic is weakened in places, when he talks about levels of activation and priming effects, but it is an underlying theme of the whole structure. On the few occasions he does mention connectionism, he confuses it with spreading activation and semantic networks, and even these (still symbolist) ideas are given short shrift. Distributed representations are simply not mentioned. As a result, some of the complex mechanisms of the design remind one of Ptolemaic epicycles, which did describe planetary motions quite well. Epicycles had to be proliferated *if* one required that all motion be circular; rules must be proliferated *if* one requires that all cognitive function be symbolic, logical, and binary. We know that distributed processes probably underlie any brain function, and may well work in parallel with the symbolic processes at an overt level. So, the second main problem area is the concentration on symbolic representations and processes to the exclusion of distributed ones that might make some effects easier to understand and to describe.

The third main problem area is that although Chapter 2 is devoted to the conversational context of speech, only a little attention is given to that context as a determiner of the forms of speech. For example, the underlying base unit of speech is taken to be the "message," which has a one-to-one correspondence with the grammatical "sentence." If a message yields an utterance that is less than a well-formed grammatical sentence, it is considered to be an elliptic message, and the evidence that a message is elliptic is that the utterance lacks some elements of a literary sentence. But in conversation, an utterance is supposed to convey information or affect the listener's behavior. To include sentence elements that are already in high focus for the listener is to convey a different message than to omit them. A string that is elliptic in literary grammar may well be complete in conversational grammar.

FINAL COMMENTS

In summary, Levelt's book should be important for computational linguists. Its main flaw, as seen from the psychological viewpoint, is contained in the statement on page 39: "No finite set of rules can ever delineate the full set of well-formed conversations." This statement probably applies to every aspect of natural language. If so, it denies the premise on which the whole production-rule system of the book is based.

We do think that rule-based systems are a part of natural language production (and reception), but we do not think they are the whole of it. Distributed representations and processes probably are at least as important. The toy connectionist systems implemented to date must be unlike those that underlie brain behavior, but even they have shown that simple networks can perform in a robust way functions that are not easy for rule-based systems. NetTalk (Sejnowski and Rosenberg 1987), for example, can do a tolerable job of turning an orthographic string into a phone string, using no phone-specific or lexical information at any node in the net, after being exposed to the correct pronunci-

ation of many strings. Rule-based systems can do better, but the rules are based on years of study by phoneticians, and are not only complex, but are supplemented by large lists of lexical exceptions.

Presumably a connectionist system that could be taught rules as well as learn from example would outperform both versions; and that is what a human seems to be—a system that can be taught rules and that can learn by example how to apply and when to bend the rules. Levelt's machine is not human-like in this respect. It knows and uses rules. It may even abstract rules from multiple examples. But the rules of the machine do not allow easy context-dependence, analogy, or the use of similarity.

One of our annoyances with this fine book is that very often, a point is settled, without evidence, by a comment that so-and-so is very unlikely. In many of these cases we find so-and-so to be quite probable. A typical example is the comment (p. 18) that it is "an unlikely assumption that the speaker articulates sentence i while formulating sentence $i + 1$." On the contrary, it seems to us unlikely that the speaker ever articulates just one sentence without at the same time formulating future sentences that contribute to a developing argument. This is a difference of opinion that is unlikely to be resolved by experiment, but it illustrates the general point.

In sum, our reaction to the book as an engineering design for a machine that might speak like a human is "magnificent," but as a description of how people speak, "we remain unconvinced."

References

Dell, G. S. 1988 "The Retrieval of Phonological Forms in Production: Tests of Predictions from Connectionist Model." *Journal of Memory and Language*, 27:124–142.

Grice, H. P. 1975 "Logic and Conversation." In Cole, P. and Morgan, J. (eds.), *Syntax and Semantics 3: Speech Acts*. Academic Press, New York.

Sejnowski, T. J. and Rosenberg, C. R. 1987 "Parallel Networks That Learn to Pronounce English Text." *Complex Systems*, 1:145–168.

M. Martin Taylor obtained his B.A.Sc. in Engineering Physics at University of Toronto, his M.S.E. in Industrial Engineering at The Johns Hopkins University, and his Ph.D. in Psychology at The Johns Hopkins University. He holds the position of Senior Experimental Psychologist at the Defence and Civil Institute of Environmental Medicine in Toronto. *Insup Taylor* obtained her B.A. at Seoul National University, and her M.A. and Ph.D. at The Johns Hopkins University, all in psychology. She is the project director of McLuhan–Nanjing Cooperative Project on Literacy. They are the co-authors of *The Psychology of Reading* (Academic Press, 1983) and *Psycholinguistics: Learning and Using Language* (Prentice-Hall, 1990). Martin Taylor's address is: DCIEM, P.O. Box 2000, North York, Ontario, Canada M3M 3B9. Insup Taylor's address is: McLuhan Program, University of Toronto, 39A Queen's Park Cres. E., Toronto, Ontario, Canada M5S 1A1. E-mail for both authors: mmt@zorac.dciem.dnd.ca

This review is also published as DCIEM Technical Report 89-N-52.

Briefly Noted

English Syntax

C. L. Baker

Cambridge, MA: The MIT Press, 1989, xv + 500 pp.
Hardbound, ISBN 0-262-02287-7, $27.50

The Syntactic Phenomena of English

James D. McCawley
(University of Chicago)

Chicago: University of Chicago Press, 1988, liii + 768 pp. in two volumes
Hardbound, ISBN 0-226-55623-9 and -55625-5, $60.00 per volume; softbound, ISBN 0-226-55624-7 and -55626-3, $19.95 per volume

Although intended as textbooks for a course in English syntax, both Baker's and McCawley's books could also serve as useful reference books on orthodoxies of, and constraints upon, the structures of English, including the more esoteric and interesting ones. Both books have roots in generative transformational syntax, though they attempt to be relatively theory-neutral. Thus, in contrast to the surface-oriented analysis of Quirk et al.'s *A Comprehensive Grammar of the English Language* (1985), these treatments emphasize structure and constituency.—*G.H.*

References

Quirk, R.; Greenbaum, S.; Leech, G.; and Svartvik, J. 1985 *A Comprehensive Grammar of the English Language*. Longman, London.

The Recognition of Speech by Machine—A Bibliography

Arthur S. House
(Institute for Defense Analysis)

London: Academic Press, 1988, vii + 498 pp.
Hardbound, ISBN 0-12-356785-8, $49.00

Approximately 4500 works on automatic speech recognition are listed, indexed by author and subject. The compiler has tried to include as much up-to-date material as possible, with some historical coverage as well. Comprehensiveness is not claimed, particularly for material not published in English.

Frequency Analysis of English Vocabulary and Grammar, Based on the LOB Corpus. Volume 1: Tag Frequencies and Word Frequencies. Volume 2: Tag Combinations and Word Combinations

Stig Johansson and Knut Hofland
(University of Oslo and Norwegian Computing Centre for the Humanities)

Oxford: Clarendon Press, 1989, Vol 1: vii + 400 pp., Vol 2: v + 380 pp.
Hardbound, Vol 1:ISBN 0-19-824221-2, Vol 2: 0-19-824222-0

This book is based on a grammatically analyzed ("tagged") version of the Lancaster–Oslo/Bergen (LOB) Corpus, which is a