

BOOK REVIEWS

Cognitive Models of Speech Processing. Edited by G.T.M. Altmann. Cambridge: MIT Press, 1991. 540 pp. \$55.00

Reviewed by
CAROL A. FOWLER*
Dartmouth College
and
Haskins Laboratories

Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives, edited by Gerry Altmann, reflects the contributions to a research workshop held in Sperlonga, Italy, in 1988. It consists of 23 chapters contributed by many of the researchers making important contributions to the study of linguistic processing of speech.

I have a prejudice about edited books (even though I have contributed to some) that led me to pick up the book with as much reluctance as interest. The prejudice is that an edited book is not likely to be worth the work of plodding through it. This is because collections tend to lack cohesion, and individual contributions tend either to be redundant with work already published in journals or, if not, to represent work that might be rejected from a reviewed journal. Finally, I suspect that the motives for publishing a collection do not necessarily include a conviction that the scientific world will benefit from its publication.

I mention all this only to point out what an edited volume has to overcome to meet with my approval. This volume succeeds on almost all fronts. I deem it a highly valuable contribution to the literature.

This volume does have an organization that is spelled out in Altmann's very useful introductory chapter, but that comes through clearly as well in reading the book. For the most part, the book consists of successive series of papers on common topics that are generally capped by a commentary on the series. The topics progress in an orderly way from discussions of earlier, lexical processing to discussions of later, syntactic processing. There is a common thread throughout most of the chapters consisting of an interest in the issue of whether various processes are modular or, alternatively, interactive.

Although there is some redundancy with contributors' work published elsewhere, this actually proves helpful rather than time-wasting. Very many chapters offer a comprehensive and illuminating look at investigators' research programs. Indeed, for

* Preparation of this review was supported by NICHD grant HD 01994 to Haskins Laboratories.

me, that is one of the real values of the volume. It provides an almost up-to-date, fairly comprehensive picture of cross-disciplinary research on language processing. There were very few chapters that I thought the volume might profitably have omitted.

The volume has some weaknesses. Not all the chapters are strong, not all are well-integrated with the volume as a whole – either because they do not fit well into the organization of the book or because the authors do not integrate the presentation of their own work with relevant work by other contributors to the volume – and not all the commentaries are as closely directed to the chapters they were meant to address as the reader might wish. (This was not necessarily the commentator's fault. In one case, a commentator had to write her commentary having at hand neither chapter she was to comment on.) A final occasional weakness derives from an otherwise great strength of the book – namely its cross-disciplinary nature. The list of contributors includes individuals with affiliations in departments of linguistics, psychology, speech, engineering and many others. The weakness stems from a failure of some contributors to take the probable readership of the volume into account when they wrote their chapters. To take two examples, readers of Frazier's chapter should bone up on "binding theory", "bounding theory", the "theta criterion" and "case theory" among other terms I stumbled over; readers of Steedman's chapter should know what $BFG \equiv \lambda x F(Gx)$, among other mysteries, means. However, for readers who do occasionally founder, Altmann's introduction provides a helpful integrative summary of many of the chapters.

LEXICAL ACCESS FROM SPOKEN INPUT

The first series consists of seven chapters and a commentary on lexical access from speech input. The chapters in this series are almost uniformly useful and interesting. They reflect a research domain that is undergoing rapid and exciting development.

Four models of lexical access are presented in the series. The development of these explicit models apparently has helped to sharpen the theoretical issues surrounding lexical access and has fostered the development of a number of useful experimental paradigms. I will consider the progress in each of these domains that the series of chapters presents.

The models

Properties of TRACE (Elman and McClelland, 1986) are explored by simulation in a chapter by Frauenfelder and Peeters, Marslen-Wilson tests some experimental predictions of his revised cohort model (e.g., Marslen-Wilson, 1987) and Luce, Pisoni, and Goldinger summarize previously published research addressing predictions of Luce's (1986) Neighborhood Activation Model (NAM). I will consider a fourth model, offered by Norris, separately.

Most striking are the points of agreement among the models. All agree that a number of lexical candidates is "activated" by acoustic speech input and that candidates other than the one matching the input serve as "competitors" for the target word. The models are also in rough agreement that other top-down information can affect word selection,

although the information deemed effective and constraints imposed on its roles differ across models. Finally, of the two models that address the issue of segmentation of the speech stream, TRACE and the cohort model agree that there is no explicit segmentation (in contrast to the proposal by Cutler and Norris (1988) that lexical search is initiated at onsets of strong syllables) other than that yielded by sequential identification of words in the speech input.

A unique feature of TRACE is its network of lateral inhibitory connections among lexical units. In her commentary, Bard points out that models with and without lateral inhibition generally make qualitatively the same predictions about the effects of competing lexical candidates on lexical access. Lateral inhibition serves a sharpening function but does not, typically, change the predicted ordinal patterning of response times or accuracies. She suggests one prediction that might distinguish a model such as TRACE from others. It is that lateral inhibition should create an interaction between the effects of target-word and competitor "strength" on response measures. For example, a low frequency and hence weak target will inhibit its competitors weakly; accordingly, there may be a large difference in the effect of high and low frequency competitors on time to identify the low frequency target; this difference should be attenuated if the target word is high in frequency and hence can inhibit its competitors strongly. Bard concludes that the evidence to date does not favor lateral inhibition.

A difference between the cohort model and NAM concerns the locus of frequency effects on performance. In the cohort model, frequency affects activation of lexical candidates, whereas in NAM frequency has a post-access influence on the decision rule that selects a final candidate. The disagreement rests in part on apparently incompatible experimental outcomes obtained by Marslen-Wilson (e.g., 1987) and by Goldinger, Luce and Pisoni (1989) and reported in the chapter by Luce *et al.* Given that lexicons arise from experience with word usage in the environment of the listener, it would not be surprising if frequency effects reflected differences in lexical memory itself. However, the data do not appear to decide this issue yet. Further differences between the cohort model and NAM arise in the insensitivity of the cohort model, but not of NAM, to the number of competitors activated by spoken input and the insensitivity of NAM, but not the cohort model to the locus of any phonetic overlap (at word beginning or later) between words in the lexicon.

My general impression of these models is that the agreements among them (particularly that "activation" is an apt metaphor to describe the effect of spoken input on lexical entries and that competitors can affect ease of lexical identification) are fundamental. Of course the disagreements will fuel further research and further development of the models.

Norris' model is more restricted in its focus than the others, but it tackles a major implausibility in TRACE and a major problem that must be handled ultimately in any model of speech perception. The problem is the temporal dimension of speech. In TRACE, features, phonemes and words span fixed numbers of "time slices" and time slices are spatialized. Because, in a spoken sequence, a word can start at any point in time, TRACE's whole lexicon is replicated repeatedly over time slices. This is not only uneconomical, it is, ultimately, not a workable solution. As Norris points out, the model cannot generalize a pattern learned at one point in the network to other points in time

where the pattern might also occur. Further, since each phoneme occupies a fixed number of time slices, the model can only handle words spoken at one rate.

Norris' solution is to adapt a Jordan network (Jordan, 1986) to speech recognition. This network has been successful in simulating sequencing with coarticulation in speech production. Like ordinary feedforward nets, it has input units (called "plan" units in the speech production model) that are linked to hidden units; hidden units are linked to output units. But, in addition, it has recurrent links from output units to a set of state units so that the state units store the state of the network. Both input units and state units feed to hidden units. In that way, the output of the network is determined by activation both of the input units and of the state units, which have encoded the network's history. Jordan's net learned to output sequences of phonemes. In Norris' adaptation, plan units become input units that receive acoustic/phonetic information and output units are lexical units. Initial results suggest that the system learns sequential patterns, with rate variation, considerably better than does a straight feedforward model while evading the massive replication of the lexical network in TRACE.

Theoretical issues

The following is a sampling of the issues considered in the chapters. They reflect most directly the characters of the TRACE, cohort and NAM models.

As noted earlier, none of the models has an explicit segmentation process. Speech input continuously activates lexical candidates, and segmentation is a consequence of word identification. In contrast, in her chapter, Cutler summarizes research of her own and her colleagues that she interprets as favouring a distinct segmentation strategy. (A later chapter by Segui, Dupoux, and Mehler examines crosslinguistic differences in the strategy.) In English, the strategy is to initiate lexical access whenever a strong syllable is encountered. Estimates suggest that, in conversation, this strategy will succeed about 90% of the time. Bard suggests, however, that at least some of the evidence interpreted by Cutler as favoring the segmentation strategy might be explained instead as an indication that strong syllables activate words compatible with themselves more than do weak syllables. This interpretation, if it is viable, would preserve the view of segmentation as a consequence of lexical access built into the TRACE and cohort models. However, on the side of Cutler's view is the consideration that, before a child has a lexicon, there must be a way for him or her to acquire one. Acquiring one must require segmentation of the speech signal without the help of a lexicon. More about this a little later.

Issues related to that of segmentation concern the activation of lexical competitors of a target word – in particular, their nature and their number. Shillcock reports that embedded words such as "bone" in "trombone" are activated (in that "trombone" is an effective prime for "rib"). This outcome does not occur in the simulations using TRACE reported by Frauenfelder and Peeters unless the longer word is composed of two embedded words (in their example: succeed = suck + seed). In any case, Shillcock's results are complex in that embedded-word priming does not occur for prefixed words such as "report", and the reasons why it does not have not yet been worked out. As for the number of competitors, Luce and colleagues find that "density" of competitors

affects ease of identification of a target word in a variety of paradigms. The cohort model does not predict this, since in that model, lexical candidates are activated in parallel, and time to identify a word is determined by the similarity and frequency of the most similar competitor. Again there seems to be some disagreement in experimental findings; Shillcock does not get effects of the number of competitors, but only of the highest frequency one. Bard suggests that a resolution may require looking at the set of words in dense and sparse neighborhoods. Dense neighborhoods offer greater opportunity than sparse ones for one of the neighboring words to be a high-frequency word and hence a strong competitor. Perhaps, then, it is not really the number of competitors, but the most activated member of sparse and dense neighborhoods that affects target-word ease of identification.

Experimental paradigms

I was very impressed with the variety of experimental paradigms that researchers use in this domain to investigate lexical access and by the extent to which some of them seem rather beautifully suited to their function. In my own field, investigating the front end of speech perception, I have complained (Fowler, 1990) that experimental procedures are not always carefully chosen for the hypotheses they are supposed to test. It strikes me that there is more attention to this matter in this research area than in my own. Research reported by Cutler, by Luce *et al.*, and by Marslen-Wilson's work in particular is notable for the use of a variety of converging measures. Marslen-Wilson attests to the importance of using converging (or apparently converging) measures. Moreover, he reports a very clever use of a cross-modal priming technique in which auditory primes are whole or truncated.

Some final comments

None of the models presented in this section has a front end – that is a component in which features, phonemes or anything else, is extracted from the raw acoustic signal for lexical access. This is not a complaint. Had the researchers allowed themselves to be bogged down by this problem, they could not have made significant progress in studying the role of the lexicon itself in speech perception. I mention it because a constraint on any realistic model of speech perception is that the lexicon and procedures for lexical access have to be acquirable by a novice lacking a lexicon. I am convinced by the research I have read about in these chapters that the lexicon figures importantly in speech perception by competent users of the language. However, one should not conclude from this that the lexicon is necessary for identification of the phonetic label that constitutes a word or for the segmentation of (at least child-directed) speech into its component words. The only way for children to learn what to do with their vocal tracts to produce a word they have heard spoken is to get the requisite information, bottom up, from the speech signal. The only way to learn which sequences are words includes segmenting the signal based on information for segmentation in the signal. The point is that an acoustic speech signal *can* be adequate for identifying the phonetic labels of words even though it may not be adequate as produced by a talker who knows that his or her audience can supply top-down support.

DEVELOPMENT OF SKILL IN SPEECH ANALYSIS

Mehler, Dupoux and Segui make almost the same argument in their chapter located later in the book, and they propose a conceptual model SARAH that itself develops from a system suitable for speech analysis by infants to one suitable for speech analysis by adults. In the initial state, the model has an array of "syllable analyzers", one for each possible syllable of any possible language. A coarse-grained unit such as a syllable is proposed as primitive for a variety of reasons. Infants begin life processing speech differently from nonspeech (in showing a right ear advantage only for the former) and so they must have some way to distinguish speech from nonspeech. A continuous fine-grained representation such as that yielded by LPC analyses performed on successions of short time windows would not readily make a speech/nonspeech distinction, but a coarser chunking into syllable-sized units might. Further, infants normalize for speech rate (Miller and Eimas, 1983), and syllable analyzers could incorporate some kind of time warping routine to achieve that. Finally, in their other chapter (Segui *et al.*), the same authors interpret their findings as evidence that a syllable-based segmentation of the speech signal is used by adults to access the lexicon.

As experience with the native language grows, unused syllable analyzers atrophy as it were, leaving just those used in the language. Likewise, universal word-segmentation strategies are pared down to just those that are useful in the native language. The analyzers and word boundary detectors provide the means by which a child can begin to build a lexicon.

In the steady-state of the model, the lexicon is accessed based on syllable-sized units, but lexical entries themselves are sequences of phonological segments. (Mehler *et al.* do not offer a motivation for this last proposal. Other things equal, in a system like the one the authors outline, one might expect syllable-sized primitives in the lexicon as well. This feature of the model presumably is driven by experimental evidence for phoneme-sized segments rather than by *a priori* theoretical considerations.)

The chapter by Mehler *et al.* is the only one to address the problem of development directly, and, in my view, it lays out the problem just right. How can a child, lacking a lexicon, get one if, in adults, both segmentation of the signal and identification of words require a lexicon? I am less satisfied with the solution that the authors propose. My major objection is that the model accepts the general view, characteristic of the models that lack a developmental component, that to perceive some *x* the perceiver must have a representation of *x* in the head already (an *x* analyzer). Since it is hopelessly implausible that infants are born with detectors for all possible words of all possible languages, Mehler *et al.* suggest syllables as primitive innate representations. However, another possibility is that there are no pre-existing linguistic-unit representations at all. Rather, syllables and other components of a word are real components of speech utterances that the acoustic signal informs about in a way that infantile auditory (or specialized speech) perceptual systems can detect. I find the proposal of innate detectors unsatisfactory because it seems to hand to the evolutionary biologist a problem of the same order as the one that Mehler *et al.* reject for the developmental psychologist — that is, one of explaining how representations could have arisen from organisms lacking them if prior representations are needed to make sense of perceptual input.

LEXICAL STATISTICS

One chapter, by Altmann, stands on its own in the series. It describes work on "lexical statistics". One aim of the research in this area is to discover lexical search strategies that can considerably reduce the set of candidate lexical entries that might match the input. Psychologically, this is an important area of research, in part because spoken word inputs to a perceptual system may well be incomplete due to noise or casual speech articulation, and in part because the perceptual system does not appear to wait until a whole word has been uttered before initiating lexical search. A second aim of the research is to develop a realistic measure of search-space reduction — a measure that accurately reflects the size of a space of candidates to match an input word. Altmann favors two measures developed by Carter (1987). One, labeled PIE (for "percent of information extracted"), is based on information theory and specifies how much information is required to achieve a given reduction in the search space. The second measure, PINE (100% minus PIE), is the information that remains to be found to identify a word uniquely.

Among the findings Altmann reports using these measures is one consistent with Cutler's proposal that listeners access the lexicon based on perceived strong syllables. Altmann finds that knowing stressed syllables reduces the search space more than knowing weak syllables. The major reason, as Altmann shows, is that less information is provided by unstressed than stressed vowels, because the former are frequently schwa. An interesting finding is that contiguous segments that do not compose a syllable and, especially, noncontiguous segments reduce the search space more than do strong syllables, because the phonotactic constraints among segments of a syllable reduce their information content. (An inference that Altmann draws is that it may be unfortunate that talkers coarticulate, because that increases the likelihood that contiguous sequences will be extracted. This seems somewhat misguided to me, since the best search string of all must be the one that itself specifies the lexical item uniquely.)

The work that Altmann reports is quite interesting. However, there is as yet little psychology in it (and so the title "lexical statistics and cognitive models of speech processing" is somewhat misleading). That is, the investigators do not restrict their analyses to partial-word descriptions that are likely products of incomplete information made available to the perceptual system. Perhaps that is good because preconceptions about the kind of information that perceivers may receive and the kinds of partial information that they are likely to be able to use may preclude discovery of counter-intuitive search capabilities. However, it does make it difficult for me to see clearly what of psychological import should be made of these findings.

MODULARITY VERSUS INTERACTION IN WORD RECOGNITION

Two chapters, by Connine and Samuel respectively, and a commentary by Tyler focus on possible limits on the interactions among levels of representation in word recognition. The two chapters consider the roles both of the lexicon and of sentential context in phonological analysis of spoken words. Connine's stimulus materials were

acoustic continua that varied in VOT so that one endpoint was a word or nonword beginning with /d/ and the other a word or nonword beginning with /t/. Using continua in which the endpoints were both words, Connine examined the effects of a preceding sentence context on identification. (For example, identifying a word as "dent" or "tent" in the contexts: "She drives the car with the _____" and "She saw the show in the _____".) Measures of response time to identify the word did show an effect of context, but only for stimuli near the endpoints of the continuum. For two reasons, Connine concludes that the effect of sentence context was not on phonological analysis or lexical selection, but rather was a post-perceptual bias effect. One reason is that Connine and Clifton (1987) had found a similar response patterning when word identification was biased by pay-offs – a source of influence that is not likely to be perceptual. Second, Connine argues that, in interactive models in which sentential context can affect lexical selection, the effect of context builds up over time and so should be most evident in the middle of the continuum where response times are long.

Using continua in which one endpoint was a nonword (e.g., *dice/tice* and *dype/type*), Connine and Clifton (1987) obtained a different response pattern. In this case there was a response time advantage for the real word, but that advantage now was evident only near the continuum boundary. As just noted, Connine interprets this patterning in the data as evidence of an interactive relation between a higher level process and phonological analysis.

Samuel draws exactly the same pair of conclusions as Connine about the differential sources of influence on phonological analysis provided by lexical and sentential contexts. His paradigm uses the "phonemic restoration" phenomenon (Warren, 1970) whereby a phoneme is perceptually restored in a word when its acoustic correlates have been removed and replaced by a potential masker. In Samuel's version of the paradigm, subjects listen to two tokens of a word or nonword. In one, the acoustic consequences of production of a phoneme have been eliminated and replaced by noise (the "replaced" condition); in the other, the acoustic consequences remain, but are masked by noise (the "added" condition). Subjects attempt to discriminate the two types of stimuli. Samuel (1981) found that accuracy (*d'*) for doing so was lower for words than for nonwords, signaling that restoration of the missing phoneme is more compelling in words. In the present chapter, Samuel reports additional tests, some very clever, all leading to the same conclusion that the lexical context of a phoneme affects *d'* and so may directly affect perceptual analysis of a speech utterance. In contrast, *d'* is not reduced when words are presented in a predictive sentential context. Rather, β , the bias term is increased, signaling a bias to respond that a word is intact.

In her commentary, Tyler focuses on just the one conclusion that Connine and Samuel draw, that effects of sentence context on phonological analysis are post-perceptual and noninteractive. She may have restricted herself to this one conclusion because she found it more controversial than the conclusion that lexical influences are interactive. Perhaps because she focuses on just half of the outcome, however, her assessment of the research described in the chapters by Connine and Samuel is, in my view, more negative and pessimistic than the work warrants.

Generally, Tyler finds fault with the paradigms described in both chapters with regard to their chances of discriminating between modular and interactive systems. For example,

she points out that Connine's response times are very long (averaging a second or more), too long perhaps to provide a reliable index of perceptual processes uncontaminated by post-perceptual ones. As for Samuel, she faults his paradigm on several grounds. For example, she argues that without ROC curves, it is not possible to know whether d' and β are independent. In addition, it is not known how to map d' and β onto the three ostensible phases of word recognition. (According to Marslen-Wilson, 1987, they are access to a set of candidates, selection of one of them and integration of the word into ongoing high-level analyses of an utterance.) Accordingly, Samuel's assumption that d' indexes perceptual phases (access and selection) and β indexes post-perceptual phases cannot be justified. Finally, lacking speeded responses, there is ample occasion for post-perceptual processes to contaminate the data.

Although these arguments may have some force, I judge the research more positively and optimistically than does Tyler. Any experimental research on cognitive processes must use indirect measures, because the processes of interest are covert; so no procedure is risk-free. However, there is reason to believe that the paradigms of Connine and Samuel can provide outcomes as interpretable as other procedures. Both paradigms are capable of providing distinct response patterns, one suggestive of modular processing and one suggestive of interactive processing. Moreover, both do so in the research summarized by Connine and Samuel. The fact that both can provide an outcome suggestive of modular processing is most striking since Tyler's objection that response times were slow would lead to a prediction that all outcomes would look interactive. The fact that the findings converge so neatly is striking and pleasing too.

Even so, the next chapter by Thompson and Altmann may provide some justification for doubting that any final resolution of the question of sources of influence on lexical access and selection will be as neat as the findings of Connine and Samuel suggest. Using implemented and hypothetical models as examples, Thompson and Altmann show that the relation between the syntactic parser (if any, see next section) and the lexicon can plausibly take a variety of forms. The relation may be modular or interactive, but there are multiple plausible alternatives that are varieties of hybrids between the conceptual extremes.

REPRESENTATIONAL STRUCTURE IN CONNECTIONIST MODELS

I found Elman's chapter on "Representational structure in connectionist models" one of the most exciting in the book, largely because it offers a perspective on language learning that seems to me considerably more realistic than, for example, the metaphor of the child as a little linguist or hypothesis tester. Elman's aim, however, was not to develop a theory of language acquisition, but rather to investigate how connectionist models that learn come to represent grammatical information.

Elman describes two simulations. Both take sentences, word by word, as input; their task is to predict a forthcoming word to follow its predecessor. During training, when there is an error in prediction and the actual next word is input, back propagation is used to change network weights to improve prediction accuracy. Like Norris' model, Elman's is a Jordan net (Jordan, 1986), adopted to enable exploitation of the infor-

mation available in word order. The model has input and state units that feed into hidden units. The hidden units feed to output units and also, in a 1:1 fashion with weights 1.0, back onto state units. In this way, the hidden units serve to map new inputs and prior states of the system onto the next output.

In the first simulation, sentences were two or three words long and consisted of a subject noun, followed by a verb and sometimes by an object noun. Some verbs in the lexicon require a direct object, some take an object optionally and some do not take an object. Nouns are animate or inanimate. Following training of the network, Elman assessed what it had learned about grammatical word class by performing a cluster analysis on the hidden unit activation patterns characteristic of each word in the lexicon. In the outcome of the analysis, words clustered into two major categories corresponding exactly to the class of verbs and the class of nouns. There were three subclusters in the verb category corresponding to the verbs that did not take a direct object, those that take an object optionally and those that take an object obligatorily. Subclusters in the noun category, likewise reflected word types (for example, animate *vs.* inanimate nouns).

This outcome, of course, was obtained in the absence of explicit instruction about grammatical category or about the meanings of words. The finding evokes the "syntactic bootstrapping hypothesis" (Gleitman, Gleitman, Landau and Wanner, 1987), which suggests that language learners can learn something about the meanings of words from the constructions in which the words appear in sentences. Testing the hypothesis, Naigles (1990) found that children would look selectively at a videotape showing a causative action having heard, for example, "Look! The duck is gorging the bunny", and at a tape showing a noncausative action given, "Look! The duck and the bunny are gorging". (Similar results were obtained in another study by Hirsh-Pasek, Naigles, Golinkoff, Gleitman and Gleitman, 1988, with real-word verbs, e.g., "Big Bird flexes with Cookie Monster".) Elman's model has a similar capability.

The second simulation is even more impressive in that it focused specifically on the model's ability to deal with grammatical features that depend on serial order constraints. Elman looked at subject-verb agreement and on relative clause formation. Ambitiously, he used a grammar that permits center embedding, so that nouns and verbs that must agree may be separated from each other by several words.

Interestingly, the model did not learn the grammar if complex sentences were introduced immediately. Analogous, perhaps, to children, the model benefited from simpler ("model-directed"?) speech, initially consisting only of simple sentences. Over four training phases, complex sentences were introduced with increasing frequency until they constituted 75% of the input. Following training, the model handled verb agreement and relative clause formation very well. Given a sentence-initial word, "boys", the model predicted only "who" (the only way in the grammar to introduce a relative clause) or else verbs in the third person plural form. Given "boys who Mary", the model not only predicted third-person verbs exclusively, but it restricted its prediction to verbs that take a direct object! Having received that verb as further input, the model now predicted a third-person plural verb to agree with "boys". The model handled up to three center embeddings before running into difficulties, performing somewhat better than typical human subjects.

Although, as noted, Elman's major interest was in the steady state of the models, I think that they may serve as interesting models of language acquisition. That is not to say, of course, that children learn by predicting successive words of a sentence, but rather that they may acquire grammatical structure without having to hypothesize explicit rules or set parameters in an innate language acquisition device. Children do not *appear* to be little linguists; perhaps they are not. (Perhaps they are little Jordan nets.)

LANGUAGE COMPREHENSION

Two chapters, by Tanenhaus, Garnsey and Boland and by Frazier, and a commentary by J.D. Fodor focus on sentence parsing. A central issue, not unexpectedly in this volume, concerns the autonomy of syntactic parsing. A major question, then, concerns the point in sentence comprehension at which world-knowledge is brought to bear. A second question, which arises in contrasting the views of Tanenhaus *et al.* and Frazier, concerns the source of any world-knowledge that is brought to bear — from within or outside of the language system.

Tanenhaus *et al.* and Frazier agree that "thematic structure" (in the definition of Tanenhaus *et al.*, the way in which the arguments of a verb participate in the event denoted by the verb) can guide parsing. However, Tanenhaus *et al.*, who do not address themselves to the modularity issue explicitly, consider thematic information part of "combinatory lexical information" and hence available in the lexical entry of a word; in contrast, Frazier assumes that it derives from outside the language system and hence reflects a lack of encapsulation of some kinds of syntactic processing.

Tanenhaus et al.

Tanenhaus *et al.* report a number of interesting and mutually consistent findings regarding "gap filling" in sentence comprehension. As interesting as the findings, are the experimental methods used to obtain them.

"Gaps" are said to exist in sentences in which a constituent ostensibly has been moved from its underlying location. In (1a) and (1b), the gap is marked by " Δ " (and the moved constituent is "which customer"). The gap filler "article" renders (1b) anomalous.

(1a) The businessman knew which customer the secretary had called Δ at home.

(1b) The businessman knew which article the secretary had called Δ at home.

In one experiment, Tanenhaus *et al.* asked when a gap is filled in sentence processing. That is, do readers, having read "called", immediately posit and fill a gap, because a gap might have occurred there, even though the next word, not yet presented, might turn out to be "about", with the gap following it instead of following "called"? The researchers used understandable sentences such as (1a) and anomalous sentences such as (1b), and they recorded evoked potentials associated with each word in the sentence. The measure of interest was N400, a negativity occurring 400 msec after stimulation.

N400 is large if the subject detects something unexpected, here an anomaly in a sentence such as (1b). Words were presented slowly (one each 500 msec) to permit association of an N400 with each word in the sentence. The researchers found a large difference in the N400 associated with verbs such as "called" in sentences such as (1a) vs. (1b). The finding suggests that subjects immediately filled the gap after "called" even though a gap need not occur there; therefore, reading "called" and filling the gap with "article", they immediately detected the anomaly in sentences such as (1b).

This technique strikes me as rather wonderful in not requiring any artificial judgment from the listener that might disrupt ordinary processing. The only drawback is the necessity of presenting the words slowly to permit N400 to develop in association with presentation of each word.

The remaining studies summarized by Tanenhaus *et al.* used a somewhat less elegant, but still "on-line", technique of having subjects make judgments after each word of a sentence as to whether the sentence still made sense. Dependent measures were response category ("yes"/"no") and response time. Using this technique, the authors found that not all verbs are like "call" in promoting immediate gap filling. Other verbs, such as "remind" that are commonly followed by complements (as in "Which child did your brother remind to watch the show") only promote immediate gap filling if the potential filler is plausible. (For example, in the sentence, "child" is a plausible object of "remind", but "movie" in "Which movie did your brother remind to watch the show" is not.) The finding that filling a potential gap is affected by the real-world plausibility of the potential filler is one piece of evidence offered by Tanenhaus *et al.* that combinatory lexical information, including thematic structure, guides syntactic parsing.

Frazier

Frazier cites work of her own with Rayner (Rayner, Carlson, and Frazier, 1983) to argue that some aspects of syntactic parsing are generated by real-world knowledge. These investigators used another fairly nonintrusive measure, eye movements during reading, to examine processing of sentences such as (2a) and (2b).

(2a) The florist sent the flowers was very pleased with herself.

(2b) The performer sent the flowers was very pleased with herself.

They found that (2a) was associated with more indications of "garden pathing" (because "sent" is interpreted as a main verb) than was (2b). For reasons I do not wholly appreciate, this kind of information distinguishing florists from performers is supposed not to be coded linguistically and stored in the lexicon. Accordingly, its influence on parsing is identified as an influence of real-world (nonlinguistic) knowledge.

Rather than supposing that all of sentence parsing is penetrated by world-knowledge, Frazier summarizes evidence that two kinds of module subserve syntactic analysis of sentences. Constituent-structure ("c-structure") analysis is strictly modular in J.A. Fodor's (1983) sense. It pursues one syntactic analysis at a time and it chooses the parse based on local, strictly grammatical, information. In contrast to this, other evidence suggests that for analyses recovering the syntactic relations among potentially non-adjacent constituents — including reference assignment and identifying thematic roles

for verb arguments — analysis proceeds differently. The thematic role analyzer, for example, may delay its parse until all verb arguments are in, and it uses real-world knowledge as well as grammatical information provided by the c-structure module to guide its formation of a “thematic frame” for the sentence.

One piece of evidence for the system that constructs the thematic frame having the characteristics just described is provided by an experiment in which sentence-reading times were collected. Frazier points out that, given sentence fragments such as (3a) and (3b), we are inclined to treat the first noun as “theme” rather than as “agent”:

- (3a) The building’s excavation
- (3b) The city’s excavation
- (3c) The city’s excavation of the historic building.

However, that noun *can* be the agent, as in (3c). If the processor were to assign thematic roles immediately, then “city’s” role would have to be reassigned in (3c) once “historic building” had been read. However, no differences in reading times were found in sentences in which words such as “city” served either as theme or agent. This suggests to Frazier that, unlike c-structure determination, which provides a parse for each word as it comes in, role assignment is delayed until all arguments have been read. However, earlier assignment did occur with words such as “archaeologist” that are not plausibly themes in the phrase “the archaeologist’s excavation”. This suggests a second critical difference between thematic role determination and c-structure determination, namely that real-world plausibility can affect parsing of thematic structure.

In short, then, Frazier proposes that J.A. Fodor’s dichotomy between modular input systems and the nonmodular central processing system needs revision. In the revision, serving as a kind of interface between linguistic input modules (including the c-structure system), are almost-modular systems (such as the thematic-role system) that are modular except that they are not encapsulated from the nonlinguistic “central system”, the source of real-world knowledge.

J.D. Fodor

In her commentary, Fodor focuses on the implications for modularity that the chapters by Tanenhaus *et al.* and Frazier raise. Her own inclination is to maintain a view that the language system is strictly modular with respect to nonlinguistic general world knowledge (the central system alluded to above). She argues that findings by Tanenhaus *et al.* do not distinguish compellingly between an interpretation that subjects use syntactic information about verbs, and the one that Tanenhaus *et al.* prefer, that subjects use thematic information. Further, she suggests that their word-by-word sense judgment task may tend to enforce use of real-world knowledge at points where that knowledge might not be brought to bear in natural language processing.

Fodor mentions in a footnote that she did not have access to the chapter by Tanenhaus *et al.*, but rather was sent another chapter, by Tanenhaus and Carlson (1989). That may explain why she does not mention their use of the nonintrusive evoked potential measure that may be less subject to the criticism she raises about their methodology.

Perhaps for the same reason, she does not allude to the fact that Tanenhaus *et al.* consider the thematic information to be available in the lexicon. Given that, their findings do not challenge a theory that language processing is encapsulated from nonlinguistic stores of world-knowledge (as Boland, Tanenhaus and Garnsey (1990) point out explicitly in a recent publication). Perhaps, however, even knowing that the information might be available in lexical entries, Fodor would still find the formulation of Tanenhaus *et al.* objectionable. She characterizes thematic information as "semantic/pragmatic" in nature, and she wants to exclude that kind of information from influencing syntactic parsing — even, perhaps, if that kind of information were available in the lexicon.

Commenting on Frazier's chapter (also unavailable to her in its present form when she wrote the commentary), Fodor points out that Frazier's proposal of an unencapsulated syntactic module is considerably more radical than Frazier's characterization of it implies. In the conception of modularity of language, there is a language module that is encapsulated from the nonlinguistic central system. Then, within the language module, there are submodules roughly corresponding to the autonomous levels of linguistic theory. Given that view, Frazier's proposal suggests not only that some of syntactic parsing is penetrated by nonlinguistic knowledge of the world, but that, therefore, the language module itself is penetrable. Fodor judges the data too sparse to warrant as yet such a radical weakening of the idea of module. (She offers a counterinterpretation of the "excavation" experiment summarized above, but, perhaps as an unfortunate consequence of her not having the final version of Frazier's chapter, she predicts an outcome of the experiment that is different from the outcome that Frazier reports.) She suggests an alternative to Frazier's model that preserves the modularity of the language system, but that can accommodate most of the data. It is to suppose that the syntactic parser does not pursue just one parse at a time (as findings of garden-pathing have suggested), but that it pursues some constrained subset of all parses consistent with the input. Real-world knowledge then can dispose of all parses except one among those that the syntactic parser offers. Evidence of garden-pathing is then reinterpreted as reflecting the weightings initially given by the central system to the set of parses offered by the syntactic system.

Comments

I have devoted a great deal of space to these chapters because the enterprise they reflect is central to understanding language comprehension, and I admire the ingenious paradigms that Garnsey, Tanenhaus, and colleagues and Rayner, Frazier, and colleagues have developed to study language comprehension. However, I did find the collection of chapters disappointing in a few ways that concern the contribution they make to the book itself. I summarize those disappointments briefly below.

Communication across disciplines. As noted earlier, the collection of papers in the volume itself represent contributions from people having affiliations in departments of psychology, linguistics, communication disorders, cognitive science, engineering and others. For the most part, I (a psychologist) could read the papers without major difficulty. However, I had real trouble with Frazier's chapter which I attribute to her heavy use, without illuminating definition, of terminology special to syntactic theory.

Missed confrontations or crosstalk. The contribution made by the collection of three chapters might have been greater had the editor asked Tanenhaus *et al.* and Frazier each to address explicitly some of their points of agreement and disagreement on critical issues. The crosstalk that was so welcome and apparent in the eight chapters on the lexicon is mostly absent here. Tanenhaus *et al.* might have been asked to discuss their view of modularity, given the central importance of this topic to Frazier and Fodor. For her part, Frazier might have been asked to consider how the research of Tanenhaus and colleagues fits her theoretical proposal. Whereas Tanenhaus *et al.* do cite work by Frazier and colleagues, Frazier makes just one passing reference in a footnote to research by a colleague of Tanenhaus *et al.* and none at all to the fairly extensive research by the authors on, it seems, issues of common importance and interest to Frazier and to Tanenhaus *et al.* Interestingly, the major disagreement that I detected between Tanenhaus *et al.* and Frazier concerned the source of real-world knowledge that ostensibly affects syntactic parsing. Would the proposal by Tanenhaus *et al.* that the source is lexical, were it acceptable to Frazier, have allowed her to evade weakening the idea that the language system is encapsulated from nonlinguistic knowledge of the world? Unfortunately, we do not get to learn her response to their idea or Fodor's either.

Fodor's commentary was very helpful in addressing the research summarized in both chapters from the perspective of modularity theory. However, sadly missing from the book, was a second commentary addressing the research from the perspective of an interactive, and especially, connectionist, theory. Following are two topics that I wish had been addressed.

Modular systems with leaks. More than a decade ago, Garrett (e.g., 1980) published a model of language *production* based largely on patternings of spontaneous errors of speech production. The model was of a type that we now call modular. It had two syntactic stages, one conforming approximately to construction of a deep structure or propositional representation and one conforming to construction of the surface structure. In addition, there were following stages of phonological processing elaborated by Shattuck-Hufnagel (e.g., 1979). The modular nature of the model reflected an apparent autonomy among linguistic levels that speech errors suggest. Word errors sometimes occur apparently as a deep structure representation of a sentence is being constructed. Comparison of intended (target) words and words produced in errors revealed semantic similarity of target and error (e.g., "winter" (target) → "summer" (error)) and identity of syntactic class; in contrast, the sound properties of the words appeared irrelevant to the occurrence of an error. Sound errors do occur, of course, and when they involve interaction between two words in a sentence (for example "forage in the mountains" → "morage in the fountains"), the syntactic class of the interacting words appears irrelevant, and so does the lexical status of the sound sequences that resulted from an error; that is, sound errors frequently create nonwords.

Dell (1980; 1986) found, however, that there were leaks in the modularity of the language system as it is reflected in speech errors. Whereas target-error pairs of words such as "winter" → "summer" and "pepper" → "salt" do not appear to be phonologically related, they are more closely related than are pairs resulting from a random repairing of targets and errors (e.g., "winter" - "salt"). Compatibly, whereas sound errors

do frequently create nonwords, they create real words with a statistically higher-than-chance frequency. The errors suggest a mostly-modular system, but one that has leaks.

Dell proposed a connectionist model of language production with distinct levels of representation for words, morphemes, and syllabically-coded consonants and vowels, and distinct processing levels as well. The domain-specificity of the levels of representation and the syntactic, word-formational and phonological processing gave the system its near-modularity, as reflected in speech errors. The vertical connections between levels provided the leaks. Dell's model (see, especially, 1986) provides a good fit to patterns of speech errors, both spontaneously produced and experimentally elicited, and a better fit than does Garrett's model.

Given that the book repeatedly addresses the issue whether language processing is modular or interactive, it would have been of great value had a commentary also been offered from someone with the opposite theoretical inclination from Fodor's. Such a commentator could have addressed the question whether language comprehension is really modular or else whether the findings of Frazier and Tanenhaus *et al.* suggest that it is modular with leaks, and so, perhaps, interactive.

Must parsing be explicit? The three chapters follow Elman's in which he shows that syntactic structure emerges implicitly in a connectionist learning model in which sentences are serially input and in which the uneven frequencies with which words in the sentences precede and follow one another affect network weights. I would like to know whether, in a system like this, in which syntactic structure is implicit in network weights, there needs to be an explicit parser at all. Can the findings of Tanenhaus *et al.* and Frazier be simulated in a model in which knowledge of syntax is implicit?

INTONATION AND PHRASE STRUCTURE

The last set of papers focuses on intonational phrasing in relation to syntactic structure. For me, these papers surpass Frazier's in the barriers they offer to a naive reader. Accordingly, I can only suggest in a general way what they are about.

Steedman addresses the observation that intonational phrases frequently do not conform to the syntactic phrases that conventional analyses produce. For example, in the following sentence, "was avidly reading about" may be an intonational phrase, but it is not a syntactic phrase:

- (4) The very eminent professor was avidly reading about the latest developments in geomorphology.

The metrical phonological approach of Selkirk (e.g., 1984) simply accepts the mismatch and supposes that intonational phrases map onto "sense units" rather than syntactic phrases. However, Steedman judges it implausible that languages would adopt two distinct, but simultaneously implemented, paths from sound to meaning.

He goes on to show that his own approach to a grammar, a "combinatory categorial grammar" (and the reader will have to discover for him- or herself what that is) sometimes generates nonstandard syntactic phrases. Interestingly, the fit between the phrases

it produces and intonational phrasing is very close.

This finding, if general and special to a combinatory categorial grammar (and, in his commentary, Joshi raises both of those questions), is extremely interesting. I wish I had not been prevented by the notation from understanding the grammar itself. Another interesting feature (according to Altmann in his introductory chapter to the volume) is that the grammar is right branching so that sentences can be interpreted "left-to-right" as their words come in.

In their chapter, Marcus and Hindle suggest a simple role for obligatory intonational boundaries in sentence parsing. They point out that obligatory, as opposed to optional, intonation boundaries are obligatory because they select among alternative parses of a sentence. In order for "they all knew" to have "We only suspected . . ." as its complement in the following sentence, it must be spoken parenthetically, flanked by intonation boundaries (notated in the sentence with "%"):

- (5) We only suspected % they all knew % that a burglary had been committed.

Marcus and Hindle propose that intonation boundaries are input as lexical items to the syntactic parser, which fails to recognize them as words. In response to receiving something uninterpretable, it closes off any open constituents, even ones such as the verb phrase beginning "suspected" above. This prevents it from identifying "they all knew" as the required complement of "suspected". (Notice that, in the absence of intonation boundaries, "they all knew" is identified as the complement and "that a burglary . . ." is the complement of "knew".) A later process with longer oversight finds verb arguments that a word's "theta grid" indicates are required for that verb but but they have not yet been identified.

CONCLUDING REMARKS

I came away from this book with considerable respect and admiration for most of the research endeavors I had read about, including especially the work I could most readily appreciate, the experimental and modeling work of psychologists. Theoretical ideas are being put to rigorous test, and research methods are undergoing development and refinement. Indeed, comparing the methodological and theoretical psycholinguistic state-of-the-art as represented in the volume to that, say, in the 1960s, when experimental work on psycholinguistics was just taking off, can even foster the view that scientific understanding might really be cumulative. I recommend the book to readers interested in the study of language comprehension.

(Received August 12, 1991)

REFERENCES

- BOLAND, J., TANENHAUS, M., and GARNSEY, S. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, 29, 413-432.

- CARTER, D.M. (1987). An information-theoretic analysis of phonetic dictionary access. *Computer Speech and Language*, 2, 1-11.
- CONNINE, C., and CLIFTON, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 291-299.
- CUTLER, A., and NORRIS, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- DELL, G.S. (1980). *Phonological and Lexical Encoding in Speech Production: An Analysis of Naturally Occurring and Experimentally Elicited Speech Errors*. Ph.D. thesis, University of Toronto.
- DELL, G.S. (1986). A spreading-activation theory of retrieval in speech production. *Psychological Review*, 93, 283-321.
- ELMAN, J., and MCCLELLAND, J. (1986). The TRACE model of speech perception. *Cognition*, 18, 1-86.
- FODOR, J.A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.
- FOWLER, C.A. (1990). Sound producing sources as objects of speech perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, 88, 2154-2164.
- GARRETT, M. (1980). Levels of processing in speech production. In B. Butterworth (ed.), *Language Production, 1: Speech and Talk* (pp. 177-220). London: Academic Press.
- GLEITMAN, L., GLEITMAN, H., LANDAU, B., and WANNER, E. (1987). Where learning begins: Initial representations for learning language. In F. Newmeyer (ed.), *The Cambridge Linguistics Survey, vol. 3* (pp. 150-193). Cambridge, U.K.: Cambridge University Press.
- GOLDINGER, S., LUCE, P., and PISONI, D. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-518.
- HIRSH-PASEK, K., NAIGLES, L., GOLINKOFF, R., GLEITMAN, L., and GLEITMAN, H. (1988). Syntactic bootstrapping: Evidence from comprehension. Paper presented at the Boston University Child Language Conference.
- JORDAN, M. (1986). *Serial Order: A Parallel Distributed Processing Approach*. Institute for Cognitive Science Report 8, 604. University of California, San Diego.
- LUCE, P. (1986). *Neighborhoods of Words in the Mental Lexicon*. Ph.D. thesis, Indiana University, Bloomington, IN.
- MARSLÉN-WILSON, W. (1987). Functional parallelism in spoken word recognition. In U. Frauenfelder and L. Tyler (eds.), *Spoken Word Recognition* (pp. 71-102). Cambridge, MA: MIT Press.
- MILLER, J., and EIMAS, P. (1983). Studies on the categorization of speech by infants. *Cognition*, 13, 135-165.
- NAIGLES, L. (1990). Children use syntax to learn word meanings. *Journal of Child Language*, 17, 357-374.
- RAYNER, K., CARLSON, M., and FRAZIER, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- SAMUEL, A. (1981). Phonemic restoration: Insight from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.
- SELKIRK, E. (1984). *Phonology and Syntax*. Cambridge, MA: MIT Press.
- SHATTUCK-HUFNAGEL, S. (1979). Speech errors as evidence for a serial-ordering mechanism in speech production. In W.E. Cooper and E. Walker (eds.), *Sentence processing: Psycholinguistic Studies Presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates.
- TANENHAUS, M., and CARLSON, G. (1989). Lexical structure and language comprehension. In W. Marslen-Wilson (ed.), *Lexical Representation and Process* (pp. 529-562). Cambridge, MA: MIT Press.
- WARREN, R. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393.