AI Magazine Volume 10 Number 4 (1989) (© AAAI)

Cognitive Models of Speech Processing:

Psycholinguistic and Computational Perspectives

Gerry Altmann

The 1988 Workshop on Cognitive Models of Speech Processing was held at Park Hotel Fiorelle, Sperlonga, Italy, on 16–20 May 1988. Twenty-five participants gathered in this small coastal village, where the Emperor Tiberius once kept a summer house, to discuss psycholinguistic and computational issues in speech and natural language processing.

The main aim of the workshop was to draw together current research trends within the fields of human speech perception and natural language processing. Cognitive psychologists have attempted to model what goes on at the many different levels at which speech perception can be described; they have also been concerned with the interaction, if any, between these levels. The mechanisms that have been proposed have varied in the degree to which they are amenable to detailed computational modeling. Recent developments involve the availability of new and more powerful computational frameworks within which to model cognitive processes (for example, parallel distributed processing [PDP] and the active chart).

To attempt some form of integration between the different types of behavioral data and the different computational approaches to modeling the data, scientists from both the psycholinguistic and computational domains were brought together for five days to discuss their work and share their perspectives. The talks, which covered a wide variety of topics, are summarized in the following paragraphs.

Richard Shillcock of the Centre for Speech Technology Research (CSTR), Edinburgh, Scotland, presented data on the recognition of words that contain other spurious words within them, as in trombone, which contains bone. Evidence from human studies suggested that the spurious word is activated, even though in principle it would be possible to prevent this activation by only accessing the lexicon at the offset of some previously found word (trom is not a word, so access would not be reinitiated at its offset).

This finding was further discussed when Uli Frauenfelder of the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, presented a computational simulation of the lexical access-segmentation process using the interactive-activation model TRACE. The TRACE simulation was particularly useful because it gave some insight into the way that the TRACE architecture and the precise contents of its lexicon might help suppress spurious words within words.

Still on the subject of when—and how—words are accessed, Anne Cutler of the Medical Research Council Applied Psychology Unit (APU), Cambridge, England, presented data to support the theory that new words are hypothesized to start immediately prior to strong syllables (that is, syllables which contain unreduced, full vowels; in the case of Shillcock's data, all the spurious words started with a strong syllable). Although the evidence is compelling for the English language, this theory (developed with Dennis Norris, also of APU) generated some controversy. Jacques Mehler of EHESS, Paris, and Juan Segui of CNRS, Paris, argued that Cutler and Norris's metrical segmentation strategy must be language specific because in a language such as French, all the syllables are full and unreduced, and applying this strategy to French would lead to an explosion of segmentation errors. They argued that the syllable, irrespective of its strength, is the natural unit of lexical access.

The main aim of the workshop was to draw together current research trends within the fields of human speech perception and natural language processing

Relative to the discussion of the role of strong syllables in lexical segmentation, Gerry Altmann of CSTR reviewed some of the evidence based on computational studies of large computerized lexicons (20,000+ words). This evidence suggested that a stressed syllable conveys more information about the identity of the word in which it occurs than an unstressed syllable. By applying techniques borrowed from information theory, it is possible to show that this fact is not the result of some fortuitous assignment of lexical stress to the most informative syllable. Rather, it is because more categories of stressed vowel exist than unstressed vowel (as calculated from a frequency-weighted lexicon), and more words can be eliminated from the search space when the set of categories with which to discriminate between the competing words is larger than when it is smaller.

Stressing the computational aspect, Ellen Gurman Bard of CSTR reconsidered the theoretical implications of the original Cutler and Norris findings. She argued that an experiment by Cutler and Norris, which had been interpreted as supporting their segmentation strategy, could be reinterpreted within an interactive-activation model so that the main effects were not the result of lexical segmentation but of inhibitory effects between lexical competitors. As in the case of the Frauenfelder-Shillcock discussions, the consideration of a computational framework within which to view the data on lexical segmentation enriched the possible explanations of the empirical data.

William Marslen-Wilson of APU also considered the effects of competition among lexical hypotheses during lexical access. He showed that the recognition of a word can be affected by whether frequent words exist which resemble the word in question. This result was taken as support for Marslen-Wilson's Cohort model of speech perception.

Paul Luce of New York State University at Buffalo (SUNY Buffalo) presented work in which he calculated the similarity neighborhoods of each word in a large computerized dictionary. These neighborhoods simply consisted of the phonetically similar words with which the target word might be confused, defined as those words which differed from the target in just one phoneme (beach and peach are, therefore, neighbors). Luce showed that the time to recognize a word was dependent on both the density of the neighborhood for this word and on the frequencies of the members of the neighborhood (compare Marslen-Wilson). Considerable discussion took place on just how such effects might be manifested within the interactive-activation framework.

Arty Samuel of Yale University, New Haven, Connecticut, and Cindy Connine of SUNY Binghampton reviewed the data on the effects of lexical and sentential context on lexical access. At issue was whether these effects resembled the effects of lexical context on phonemic identification: Do they interact in the same way, with higher-level information being used top-down to assist perception? Using different methodologies, they both concluded that whereas *lexical information* (information about words in the mental lexicon) can actually affect the percept of a sound making up a word, *sentential context* (syntactic likelihood, semantic plausibility, and so on) cannot influence the percept. Instead, these sentential effects occur at a later stage in processing and do not interact with the lower-level stages. These findings argue for a different kind of relationship between the phonemic and lexical levels—an interactive one—than between the lexical and sentential levels—a noninteractive one. The implications for models of speech perception—and the architectures required to model perception—were discussed at length.

Within the computational domain, Jeff Elman of the University of California, San Diego, and Dennis Norris of APU presented computational simulations based on PDP models that were sensitive to temporal variation. Both used recurrent networks based on the work of Michael Jordan. Norris demonstrated that a simplified version of a network designed to model word production could adequately model word recognition. A number of simulations were presented that demonstrated the ability of this network to recognize words from among their competitors (including cases similar to the trombone-bone case investigated by Shillcock and discussed by Frauenfelder).

Elman moved to the sentence-processing domain and demonstrated that a variant of the Jordan network could predict the next word in a simple three-word sentence (for example, man break glass, cat eat sandwich). He then showed that the internal representations constructed by the network to represent each lexical item grouped the items according to certain natural properties (for example, sandwich and cake would be grouped together but separately from the grouping containing glass and plate; however, both groupings would be placed together by virtue of their syntactic properties).

After dealing with lexical access and PDP architectures, the main focus of the workshop shifted toward the syntactic level of analysis. The emphasis here was on the notion of

Much of the discussion reflected the increased emphasis . . . on (1) the flow of information between the different modules . . . under study and (2) the modularity . . . of the models which have been advanced to explain data on human speech and sentence processing.

syntactic constraint and the immediacy with which syntactic information could be used to constrain the interpretation of sentential input. Mike Tanenhaus (Rochester, New York) discussed some recent experiments that examined the time course with which verb control and verb argument structure are accessed and used in sentence processing. The results demonstrated that word recognition makes information available about the argument structure of the verb and that this information is used immediately during parsing. Lyn Frazier of the University of Massachusetts, Amherst, explored the relationship between different proposed postlexical modules that contribute to language comprehension. She discussed four modules, two that are primarily concerned with structural properties of the input (the Binding module, which operates according to linguistic principals such as c-command, and the constituent structure module, which operates according to sisterhood relations in the syntactic structures) and two that are concerned with more pragmatic aspects of the input (referential semantics and thematic-predication assignments).

In the computational domain but still within the realm of syntactic processing, two presentations were concerned with the power of syntactic information to constrain the vast number of spurious word candidates that might be hypothesized by an automatic speech-recognition system. Domenico Parisi of CNR, Rome, described some experiments using the Olivetti isolated word recognizer. He showed that performance was significantly improved if the hypotheses' output for each word took into account the sentential relations which could be expected to hold between the different words in the utterance.

Henry Thompson of CSTR and Gerry Altmann described the manner in which syntactic information could be made to constrain the lexical hypotheses' output by an automatic continuous speech recognizer. Using the active chart framework, they described a number of different implementations by which syntactic information significantly reduced the number of word strings that would be entertained by the processor. The implementations were equivalent to the extent that they led to the same degree of constraint. They differed, however, in the degree to which they conformed to Fodor's modularity hypothesis concerning the autonomy of the processing of different information types. Thompson and Altmann concluded that modularity can be a useful theoretical construct but that a point comes when one has to abandon modularity for the sake of computational efficiency.

The final topic considered in the workshop was intonation and the relationship between intonational structure and the syntactic processor. Mark Steedman of the University of Pennsylvania, Philadelphia, pointed out that the recent theories of prosody and intonation within the metrical framework postulate a level of intonational structure independent of syntactic surface structure. He argued that such an autonomous intonational structure is unnecessary and that an identical metrical prosody can be driven directly from syntactic surface structure. This technique requires a radical revision of the concept of surface structure and is achieved by replacing the standard syntactic component with a syntax based on a combinatory generalization of categorial grammar. This theory of grammar is independently motivated by the purely syntactic phenomena of coordinate structure and unbounded dependency. At the level of surface structure, it also appears to provide exactly the structural units that are required by recent metrical accounts of intonation.

Whereas Steedman described a grammar for intonational structure, Mitch Marcus of the University of Pennsylvania, Philadelphia, described a deterministic parser that could use obligatory intonational boundaries to help it assign the right structure to a sentence. Unlike Steedman, Marcus assumed a separate component capable of identifying the intonational boundaries. His parser would then use this additional input to construct partial descriptions of otherwise standard tree structures.

It is impossible to catalog perhaps the most important aspect of the workshop, namely, the discussions that arose, whether during the sessions or away from the formal workshop structure. The formal discussions were led by Jan Charles-Luce of SUNY Buffalo; Ellen Gurman Bard; Juan Segui; David Pisoni of Indiana University; Lolly Tyler of Cambridge University, Cambridge, England; Janet Fodor of City University, New York; and Bonnie Webber and Aravind Joshi of the University of Pennsylvania, Philadelphia. Much of the discussion reflected the increased emphasis within psycholinguistic and computational modeling on (1) the flow of information between the different modules that compose the (sub)systems under study and (2) the modularity or otherwise of the models which have been advanced to explain data on human speech and sentence processing.

The workshop would not have been possible without the generous financial support of British Telecom International. Additional financial assistance was provided by the American Association for Artificial Intelligence. The proceedings of the workshop are being published by The MIT Press/Bradford Books and edited by Gerry Altmann. This workshop was the first of a series of biennial international meetings based in Europe.

Gerry Altmann is a lecturer in the Laboratory of Experimental Psychology at the University of Sussex, teaching AI and cognitive science. He received his BSc. in experimental psychology from the University of Sussex in 1981 and a Ph.D. from the Centre for Cognitive Science at the University of Edinburgh. His Ph.D. work on the resolution of syntactic ambiguity was followed by four years of postdoctoral work at the Centre for Speech Technology Research, University of Edinburgh, on the effects of context on word recognition in continuous speech and computational models of such effects.