Architectures and Mechanisms for Language Processing

Matthew W. Crocker, Martin Pickering, and Charles Clifton, Jr. (editors)
(University of the Saarland, University of Glasgow, and University of Massachusetts, Amherst)

Cambridge University Press, 2000,
x+365 pp; hardbound ISBN
0-521-63121-1, $64.95

Reviewed by
Amy Weinberg
University of Maryland

The concerns of psycholinguists will look very familiar to people engaged in CL research. A major area of investigation in psycholinguistics is determining how a listener uses frequency and other experience-based information during on-line sentence understanding. This book gives computationalists a distinguished guide to the current issues in the field. It is very clear that psycholinguistics has been influenced by recent work in CL, and is ripe for further cross-fertilization. Many of the papers recognize that more rigorous modeling is called for and that these models will require insights from computational linguistics. I will highlight some of the issues that shape current debates, with particular emphasis on areas of common interest between psycholinguistics and CL.

The major source of data in the theory of human sentence processing comes from the disambiguation of temporarily ambiguous sentences as in examples (1) and (2):

(1) The man believed the woman {**implicitly** | **was unhappy**.}

(2) The girl heard by the window {**that John was coming** | **was speaking too loudly**.}

The boldface material represents two possible continuations of the first part of each sentence, but the choice of the appropriate continuation depends on how the parser initially structures the preceding material. In (1), the parser might choose to make the postverbal noun phrase either the direct object of the matrix verb, compatible with the first continuation, or the subject of an upcoming complement sentence, compatible with the second. In (2), the ambiguity centers upon whether *heard* is taken as an active main verb or as a passive participle inside a relative clause (*the girl who was heard*). Early experiments such as those of Frazier and Rayner (1982) suggested that speakers had a clear preference for the first continuation in each case, as verified by an increase in reaction time when the disambiguating word, underlined in examples above, forced the second continuation. In a case such as (1), the initial preference seems easy to reanalyze, while in (2), the correct analysis seems unrecoverable and is commonly referred to as a garden path. These facts were initially explained by a theory that assumed a serial parser. The choice function guiding the initial analysis was independently justified either by considerations drawn from linguistics (Gibson 1991, Weinberg 2000) or theories of memory (Frazier and Rayner 1982). Importantly, these theories assumed that listeners delayed consideration of semantic, pragmatic, or frequency-based factors until after an initial analysis was constructed.

The present book highlights recent extensions and challenges to these theories.

Chapters by Richard Lewis (based on SOAR [Newell 1990]), by Paola Merlo and Suzanne Stevenson (based on a symbolic connectionist architecture), and by James Henderson (based on an extension of recurrent networks using Temporal Serial Variable Binding) try to derive initial parsing preferences and to distinguish possible from impossible reanalyses by means of architectural constraints defined in their underlying (implemented) models. The common idea here is that the distinction between an easy reanalysis and a garden path should fall out as a side effect of constraints that need to be imposed on the system in order to allow it to perform analysis on "normal" unambiguous input in an efficient manner. These models suggest that restrictions on how local the initial misattachment point is to the point of disambiguation, or how many constraints need to be specified in order for attachment to occur, determine the possibility of reanalysis. The fact that these factors characterize models with such different architectural bases suggests that these notions will play an important role in the design of any parser.

Chapters by Charles Clifton, Jr., by Michael Tanenhaus, Michael Spivey-Knowlton, and Joy Hanna, by Gerry Altmann, by Steffan Corley and Matthew Crocker, and by Martin Pickering and Matthew Traxler debate the correctness of the simple statement of the preference function that chooses initial analyses in terms of a set of noninteracting nonsemantic or pragmatic constraints. Evidence in favor of a radically interactive constraint-satisfaction model comes from the nonstationarity of preferences during sentence comprehension. For example, the relative frequency with which a verb appears as either an active main verb or a passive participle, the compatibility of the subject noun phrase with interpretation as the agent of a clause (as required by the verb of a main-clause analysis), and the presence of a post-verbal *by* phrase (normally associated with a passive participle), contribute to the availability of the reduced relative clause, as shown by the ease of interpretation in (3).

(3) The evidence examined by the lawyer turned out to be unreliable.

In (3), *evidence* is a bad agent, and thus would disfavor the main-verb analysis of *examined*. The *by* phrase, which is associated with the passive-participle reading, also disfavors the main-clause reading. Trueswell, Tanenhaus, and Garnsey (1994) found no significant difference in reading time at the main verb (*turned*) when compared with an unambiguous control (formed by replacing *examined* with a verb like *chosen*). Noninteractive theorists claim that this means that the parser initially tries the wrong main-clause analysis, on the basis of structural factors, but that this analysis is rapidly revised. Interactive theorists claim, however, that this result follows simply from allowing factors like "agentivity" and "presence of *by*" to interact in the choice of the initial analysis. The chapter by Tanenhaus, Spivey-Knowlton, and Hanna presents a constraint-satisfaction implementation that allows us to test predictions of the interactive theory. They claim that the model can explain a range of hitherto contradictory findings. For example, they claim that experiments, such as that of Ferreira and Clifton (1986), using stimuli like (3) but without the *by* phrase, that were interpreted as demonstrating a stage where the parser abstracted away from issues of agentivity, could actually be predicted by an interactive model. This is because Ferreira and Clifton's stimuli did not include *by* phrases that could further bias the model to the relative-clause reading of the ambiguous verb, because the language as a whole is biased to treat main-verb/past-particle ambiguities as main verbs, and because the stimuli used verbs that were equibiased between a relative-clause and main-clause reading. These factors conspired in their simulation to outweigh any evidence from the nonagentivity of the subject, which would favor the relative-clause reading. By contrast, since

the stimuli in Trueswell, Tanenhaus, and Garnsey (1994) contained more factors that biased in favor of the relative-clause reading, an interactive model would predict that relative clauses would be easier to understand in these cases.

On the other hand, two major criticisms of constraint-based models are leveled by Clifton in this volume. The first is that, while there is much evidence that one can make the dispreferred analysis a more fit competitor, one doesn't seem to be able to make the analysis preferred by purely structural constraints unfit (e.g., blocking the preferred interpretation in cases where it turns out to be correct). This is predicted by a model that gives particular constraints a first crack at analysis, but is not predicted by a fully interactive model.

The second criticism is that it is too easy to produce models that can handle any range of data if there is no limit to the number and type of constraints that can interact. As more sources of constraint are proposed, this becomes a real problem. For example, Altmann's and Pickering and Traxler's chapters emphasize the contribution of plausibility as a source of constraint. Corley and Crocker try to deal with this second line of attack. While it is reasonable to believe that the mental lexicon contains entries for verbs that specify the likelihood of each possible complement type, considerations of sparse data make it less feasible to suppose that every lexical choice for possible adjective-noun pairs is stored. Nonetheless, there seems to be a clear difference between the ease of understanding of cases such as (4) and (5):

(4) The warehouse fires many of its employees every summer.

(5) The corporation fires many of its employees every summer.

MacDonald (1994) suggested that this was due to the greater frequency of *warehouse fires* as an adjective-noun pair and the relative rarity of this part-of-speech analysis for *corporation fires*.

Corley and Crocker try to show that these data do not force word-by-word bigram conditioning, and propose an alternative where only word to part-of-speech-category (unigram) and part-of-speech-category to part-of-speech-category (bigram) parameters are computed when assigning probabilities to ambiguous lexical items. Their model is inspired by standard part-of-speech taggers. This balancing between enriching the repertoire of constraints and keeping their number tractable and learnable should also occupy the field in the near future.

Topics of the rest of the book include expanding the experimental paradigms used in sentence processing (to ERP work, as discussed in Colin Brown and Pater Hagoort's chapter), expanding the cross-linguistic coverage of the field (in chapters by Barbara Hemforth, Lars Konieczny, and Christoph Scheepers and by Marica De Vincenzi), and, in an interesting final section, semantic processing (chapters by Lyn Frazier, by Linda Moxey and Anthony Sanford, and by Amit Almor).

This book represents the state of the art in sentence processing, with interesting examples and opportunities for computational modeling. The themes that it presents are likely to occupy the field for some time.

References

Ferreira, Fernanda and Clifton, Charles Jr. 1986. The independence of syntactic processing. *Journal of Memory and Language*, 25: 348–368.

Frazier, Lyn and Rayner, Keith. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14: 178–210.

Gibson, Edward. 1991. *A computational theory of human linguistic processing*. Unpublished doctoral thesis, Carnegie Mellon

University.
MacDonald, Maryellen C. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9: 157–201.
Newell, Allen. 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.
Trueswell, John C., Michael K. Tanenhaus, and Susan M. Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33: 285–318.
Weinberg, Amy. 2000. A minimalist parser. In Samuel David Epstein and Norbert Hornstein, editors, *Working Minimalism*, The MIT Press, Cambridge, MA.

Amy Weinberg is an Associate Professor in the Department of Linguistics and the Institute for Advanced Computer Studies at the University of Maryland. She has worked on computational and experimental models of sentence processing, as well as models for foreign-language tutoring and machine translation. Weinberg's address is: UMIACS, University of Maryland, College Park, MD 20742; e-mail: weinberg@umiacs.umd.edu.