Grosz, B.J., Joshi, A., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics* (pp. 44–50).
Kay, M. (1985). Parsing in functional unification grammar. In D.R. Dowty, L. Karttunen, & A.M. Zwicky (Eds.), *Natural language parsing*. Cambridge, England: Cambridge University Press, 251–278.
Moore, R.C. (1980). *Reasoning about knowledge and action*. Menlo Park, CA: Stanford Research Institute Artificial Intelligence Center Report 191.
Moore, R.C. (1984). A formal theory of knowledge and action. In J. Hobbs & R.C. Moore (Eds.), *Formal theories of the commonsense world*. Norwood, NJ: Ablex.
Sacerdoti, E. (1977). *A structure for plans and behaviour*. Amsterdam: North Holland.
Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.
Sidner, C.L. (1983). Focussing in the comprehension of definite anaphora. In R. Berwick & J.M. Brady (Eds.), *Computational models of discourse* (pp. 331–371). Cambridge: MIT Press.
Turner, R. (1987). A theory of properties. *Journal of Symbolic Logic, 52*, 455–472.
Wallen, L. (1987). Matrix proof methods for modal logics. *International Joint Conference on Artificial Intelligence 87*, 917–923.

Arbib, M., Conklin, E., & Hill, J. *From Schema Theory to Language*. New York: Oxford University Press, 1987. $39.95

Reviewed by *Suzanne Mannes, University of Colorado, Boulder.*

In their book *From Schema Theory to Language*, Arbib, Conklin, and Hill have made a bold attempt at synthesizing several areas of cognitive science: biological control theory, neural modeling, artificial intelligence, cognitive psychology and brain theory. Their book includes the embellishment of several AI models which are representative of the current status of these areas. The three models discussed at length regard sentence comprehension, discourse generation about visual scenes and language acquisition in the child.

The main premise of the book is that we can view higher-level cognitive processes in a way similar to the way we understand visual processing as an action-perception cycle. We see something and our interpretation of that scene or object influences what we will see next. As we take in a visual scene we are constantly updating our representation of that scene and this in turn changes the future potential interpretations of that and other scenes. The world knowledge we have that allows us to make interpretations of relationships is stored in units called schemata, which share many features with semantic nets and production rules (p. 17). However, the chapter on knowledge representation is merely a rehashing of the well-known concepts of semantic nets and the differences between semantic networks and production rule formalisms and the relation between these representations and the schema theory to be developed is never made clear enough. A fairly detailed description of the KL-ONE network representation is given although this formalism plays a role in only one of the three models. On the whole, this treatment of knowledge representation offers little to the reader's understanding of the book's models. (On the same note, there may be a case to be made for the analogy between prey-selection in the frog and object-naming in humans but this really ought to have appeared as a footnote or appendix rather than as a chapter, unless the connection between the two was made more direct and explicit.)

As is often the case (e.g., Anderson & Hinton, 1981; Zipser, 1986), these researchers have tried to show the relation between their work on higher-level cognitive processes and the cooperative computation style of the brain. The problem with this is that the phenomenon which they claim to be most easily interpreted in neuronal terms is vision but, the computation of optic flow used as their example "is a far cry from language processing" (p. 37). Because the work reported is still so far from being interpretable in neural terms, the authors might be better off not making such elaborate hypotheses about how this work might be neuronal when they still have to conclude that it probably is not (p. 80).

The Gigley model of sentence comprehension, HOPE (Gigley, 1982) can, when presented with a sentence containing an ambiguous word, decide on the appropriate meaning of that word given the context in which it occurs (e.g., decide between the meaning of *bark* as a verb and as a noun). The goal of this model is not simply to process sentences, assigning the correct meaning to ambiguous words, but to allow the experimenter to "lesion" it by selectively disabling particular rules in order to model normal subjects as well as subjects exhibiting a variety of aphasic disorders. HOPE operates on three levels of description, the grammar level, the phonetic category meaning level and the simple phonetic level. Thus, at any time in the comprehension process, some of the schemata at these variant levels are activated and have associated with them activation values. When the level of activation exceeds some threshold, the schema fires and propagates its activation to the other schemata to which it is attached in both an excitatory and inhibitory fashion.

The model is grammar driven with rules such as

if a determiner occurs
then
postulate that there is a common noun following it.

These serve as the predictive schemata for the comprehension process. During normal operation, words are fed into the model one at a time and a cycle of decay/activation/updating takes place before a new word is introduced. One particular type of lesioning tested involved disabling the described determiner-common-noun rule. When this was done, the model could not recognize determiners and hence did not interpret words which followed them as common nouns. Therefore, in trying to understand the sentence, *The dog barks, dog* does not receive activation from the grammar space because the determiner-common-noun rule has been removed (i.e., no common nouns are able to receive activation due to the disabling of this rule). It only receives activation from the phonetic level which recognizes the word as *dog*. Unfortunately, the input from a single source is not enough to allow *dog* to reach threshold and hence it never enters the pragmatic space, where sentence interpretation actually takes place. As a consequence, when the model is fed the third word of the sentence *barks* and tries to find an agent for it in the pragmatic space, none can be found and HOPE flags an error (p. 92). By deleting the determiner-common-noun rule the model produces an error when asked to interpret the remainder of the sentence following the determiner, rather than producing an erroneous interpretation as one might expect from a true aphasic. This supposedly simulates the loss of the class of closed words exhibited by Broca's aphasics (p. 90) though the authors point out that

> One might expect such an aphasic to be able to interpret the sentence "The dog barks." quite satisfactorily using semantic information to make up for the lack of ability to use the determiner. (p. 95)

but HOPE has a highly impoverished semantics.

The authors present no data to support the claim that their model behaves as a human aphasic. Because no comparison between actual aphasic data and the model output is presented, it is difficult to know exactly what to expect and what *would* be appropriate behavior of the model. A comparison of human data and model performance has been demonstrated by another researcher (Cottrell, 1985) so sufficient data do exist to allow such comparisons to be made. Arbib, Conklin, and Hill claim, however, that

> Unfortunately, the state of the art is such that we cannot yet correlate the performance of the model at all well with aphasiological data. This is not only because of the preliminary stage of modeling in neurolinguistics, but also because aphasiological data have been gathered for clinical diagnosis in a way that pays little attention to the issues of performance modeling. (p. 49)

It is commendable that this model incorporates a very small number of limited rules. However, it is immediately apparent and readily admitted (p. 80) that this model in its current form cannot deal with sentences containing more than one noun phrase or multiple instantiations of a single word.

> Serious problems [arise] in the analysis of more complex sentences where a particular word (even so common a word as "the") occurs more than once.... (p. 80)

Also, this would be a highly inefficient means of comprehending sentences in a larger system because, for example, when a determiner is encountered, activity is propagated to all category meaning pairs for common nouns at the phonetic category meaning level; there is no way of focusing on a particular set of words to activate.

In addition, the mathematical procedures followed for obtaining activity levels in the system are not those which the more current connectionist models employ (McClelland & Rumelhart, 1986). Namely, nodes in the network receive activation in an additive manner, with no normalization. Hence, activation values for a node in this network are absolute and not relative to the other nodes with which it shares relationships. This is slightly problematical for implementational and evaluative purposes. Implementationally, it adds the burden of modeling some sort of flexible threshold determining factor, as the network gets larger and the activation values get bigger, the threshold will have to be adjusted accordingly. Also, this unusual way of accruing activation levels makes it difficult to compare this model with others at the implementational level.

Totally irrespective of these small problems is the fact that several working models of the same sort have already been developed (Cottrell & Small, 1983; McClelland & Kawamoto, 1986; Waltz & Pollack, 1985). These models have their own shortcomings but still accomplish at least as much as HOPE and can deal with more complicated texts (e.g., Waltz & Pollack's system has no trouble parsing the sentence "the astronomer married the star," although this contains multiple occurrences of the determiner "the"). Of these alternative models, the Waltz and Pollack (1985) and Cottrell and Small (1983) models are mentioned, but receive no attention and participate in no comparisons with the current model.

The second model simulates language acquisition in a single 2-year-old, Claire. This model uses templates representing relations derived from examples as the basic language unit. At first, these templates are very simple, consisting of a single word for a relation to be expressed and a single slot which may be filled with an object. After a number of examples, there comes to be a regularity apparent in the object set that a particular relation may take on, and generalization about that class of objects can then take place. Thus, words are classified through their usage (the Classification through Word Use hypothesis). Although this is called a model of language acquisition in the child, many rudimentary processes, such as phoneme and word segmentation, and the learning of word meanings are sidestepped. The task which the model actually performs is the classification of words into lexical classes. Given input in the form of an adult sentence, the model either produces an answer in the case that the input was a question or merely repeats it otherwise, according to the child's current grammar.

The implementation of this model is slightly troublesome in that it runs counter to the empirical data which are presented by the authors themselves.

It is a well-accepted fact that as a child learns language he or she often has available concepts which have no lexical item associated with them (p. 105). Likewise Claire has available to her words which have not yet gained their meaning through association but, are merely mimics of adult language. In the model,

> Adding a word to the lexicon causes a concept to be added to the concept-space and adding a concept to the concept-space prompts the model to ask for a lexical item corresponding to it. (p. 121)

Also, it has been argued that a child's competence in comprehension may often be more advanced than his or her production competence but the model uses the same set of templates for both processes (p. 122). These templates which make up the template grammar represent relations between words. The processes by which these words come into being, such as phonetic recognition and segmentation, have been ignored in this model as a simplification, but in order for this to even begin to be a neural model as the authors hope, it is essential that these lower level, precursory processes not be omitted.

It is ironic that the model deals with the language acquisition of a 2-year-old child because that is precisely of what ACT* has been demonstrated as being a good model (Anderson, 1983, pp. 261–305). Anderson's work has not been cited, leading the reader to wonder whether the authors are unfamiliar with Anderson's work or have chosen to omit it for some principled reason.

The third model is a refreshing departure from the first two. This model proposes to simulate the generation of scene descriptions by human subjects. Particularly interesting is that the authors have performed experiments and compare the behavior of their system to that of their subjects! Also, they have taken care to not sidestep too many of the important processes involved. By coordinating two subsystems GENARO and MUMBLE, Arbib, Conklin, and Hill begin modeling at a point where the visual scene to be described has been perceived and end with a paragraph-length description of that scene; the contents and form of which closely parallel human descriptions. A very interesting starting point for GENARO evaluates the potential salience of different items in the scene to be described. It is this measure of salience which determines what items in the scene will be mentioned, in what order and in relation to which other objects. GENARO uses the salience information and operates with rules such as,

> mention the most salient item first,
> then
> mention things with which it shares a relation or characteristic

to derive a series of *r*-specifications, representations of the salience information in a symbolic, LISP-like format, which are then sent off to MUMBLE.

Once an item has been used in an *r*-spec, it is removed from the stack of things to be mentioned, the completed *r*-spec is sent to MUMBLE and then the next item on the list is tested to see if its salience value warrants its mention. If it does, the cycle is repeated.

MUMBLE takes the *r*-specs it is given as input and produces English sentences which describe the semantic information it has been passed in a non-language format by GENARO. Once an *r*-spec reaches MUMBLE there are four potential relationship constructions which can be used to produce the English equivalent of the symbolic, but semantic not lexical *r*-spec. These constructions include syntactical formulations for describing a variety of spatial relations between the items in the scene. The application of these rules results in the production of some very realistic protocols.

When so much effort has gone into producing a work of this size, it is unfortunate when small errors occur in the final product. Arbib, Conklin, and Hill claim that their correlation of .52 between salience rating data and textual data (written protocols) means that "roughly 50% of selection and resultant object ordering was due to the use of salience-based selection" (p. 214). The value is really 25%.

The work presented in this book is recognized as being of an important and sophisticated nature and the descriptions of the sentence generation and visual scene description generation models are quite clear and concise. However, other more well-known models already exist to deal with two of the proposed phenomena and do so quite well. The language understanding program is quite reminiscent of HEARSAY, which is described in detail at the outset of the book and a model to deal with language acquisition in the young child has been implemented by John Anderson, using his ACT* formalism. Although the first two models presented are scholarly in their own right, they do little to further our understanding of or our competence in modeling these phenomena. The scene description generation model does provide us with a new perspective on this phenomenon though and should not be overlooked. If the reader is already familiar with the works which have been cited in this review, he or she might benefit most from concentrating on the well-thought-out introduction and the visual scene description system.

REFERENCES

Anderson, J.A., & Hinton, G.E. (1981). Models of information processing in the brain. In J.A. Anderson & G.E. Hinton (Eds.), *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Cottrell, G.W. (1985). Implications of connectionist parsing for aphasia. *Proceedings of the 9th SCAMS* (Symposium on Computer Applications in Medical Care). Institute of Electrical and Electronics Engineers, New York, NY.

Cottrell, G.W., & Small, S.L. (1983). A connectionist scheme for modelling word sense disambiguation. *Cognition and Brain Theory, 6,* 89–120.

Gigley, H.M. (1982). A processing model of English language comprehension. Unpublished doctoral dissertation. University of Massachusetts, Amherst.

McClelland, J.L., & Kawamoto, A.H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In J.L. McClelland, D.E. Rumelhart, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.*. Cambridge, MA: MIT Press.

McClelland, J.L., & Rumelhart, D.E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press.

Waltz, D.L., & Pollack, J.B. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science, 9,* 51–74.

Zipser, D. (1986). Biologically plausible models of place recognition and goal location. In J.L. McClelland, D.E. Rumelhart, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press.