The Contribution of Clinical Intelligence and Artificial Aphasiology to Clinical Aphasiology and Artificial Intelligence

Helen M. Gigley and Joseph R. Duffy
University of Massachusetts, Amherst, Massachusetts

Few people would argue with the observation that the computer is beginning to have a major impact on our lives. Clinically, it has facilitated record keeping and data analysis, has contributed to our stock of augmentative communication devices, and—as is apparent in this conference (Mills, 1982; Katz and Tong Nagy, 1982)—is appearing more frequently as a tool in the management of aphasia.

In addition to developing these practical, fast, and reliable tools, computer scientists have been using the computer to mimic, explore and explain the nature of human processes requiring intelligence (for a general overview of such activities, see Simon, 1981). These efforts are conducted within the branch of computer science known as Artificial Intelligence (AI). Of special interest is the fact that AI has developed computer programs or models which can comprehend language (e.g., Erman et al., 1976; Marcus, 1978; Newell, 1978; Rieger and Small, 1981; Small, 1980; Wilks, 1976; Winograd, 1972; Woods, 1982; Woods et al., 1976). However, while these programs can "do the job" of comprehension, they rarely can claim correspondence with human language processing.

The disciplines of Clinical Aphasiology and Artificial Intelligence begin to meet when we ask how programs can be developed which comprehend language in the same way that is hypothesized for people. One possible answer is to build a neurologically plausible model that is consistent with linguistic theory and psycholinguistic observations of normal language users. In addition, powerful evidence supporting the psychological validity of such a model would be gained if, when a lesion is simulated in the model, the resultant computer performance is aphasic-like. That is, if aphasia consistent with clinical reality can be induced in the model, the validity of the model receives additional support. To date, this clinical validation has not been a component in the development of AI models of comprehension.

In this presentation we will describe a model of comprehension whose validity will be partly established through its ability to behave "aphasically" when it is artificially lesioned. This is of interest to clinicians because clinical evidence will help to test some assumptions about language made by the model, and will help to dictate which aspect(s) of the model's knowledge or performance should be lesioned and how it should perform under such conditions. Of greater interest is the hope that the ability to examine in detail the moment-to-moment processing events and interactions which occur in the computer model under lesion conditions will help to generate or to test hypotheses about the unobservable processing problems and interactions which exist in the aphasic person. To the extent that the model provides novel insights into the nature of aphasic processing deficits, it also has potential for generating new or more highly specified hypotheses about assessment and treatment.

THE MODEL

Before discussing the role of clinical activity in developing the model, it is necessary to know something about the model itself. The model is called HOPE (Gigley, 1982a and b, to appear). It has several characteristics which represent attempts to be consistent with neurological, linguistic, and psychological data and theory about language comprehension.

Within the model there are aspects of knowledge ("spaces") which have been called Phonetic, Phon-cat-mean (Phonetic category meaning), Grammar, and Pragmatic. The Phonetic space contains the phonetic representation of words—at this level we can say that words are perceived. HOPE does not deal with how this happens and assumes that words are perceived correctly. Phon-cat-mean is the space in which the phonetic representation of a word is linked to all of its possible meanings. A major assumption of the model is that during the initial stages of processing all meanings of a word are accessed simultaneously (for data related to this assumption, see Marslem-Wilson and Tyler, 1980; Seidenberg and Tannenhaus, 1980; Swinney, to appear). For example, if the word "building" enters the system, the noun and verb meanings of "building" become active, only to be disambiguated during subsequent processing.

The Grammar in HOPE is a categorial grammar. It contains rules for linking syntactic categories together as they are identified during processing, and it ultimately coordinates syntactic form with semantic interpretation.

The Pragmatic space in the model is empty at the beginning of processing for any sentence. It is basically the location to which the interpretation of meaning within a sentence is sent once that interpretation is made. For example, in the sentence "the dog barks," once the model has decided that "the" belongs with the common noun "dog," and that "the dog" can be interpretated as a term, that interpretation would be housed in Pragmatic and await other decisions which build further the meaning of the sentence.

In terms of processing, HOPE exhibits bottom-up and top-down characteristics. For example, a word perceived in Phonetic influences the meanings triggered in Phon-cat-mean, which influences the grammatical processing in Grammar—this is bottom-up activity. But the ongoing processing in Grammar will continuously influence (in top-down fashion) the likelihood that one of a word's meanings in Phon-cat-mean will ultimately be selected as <u>the</u> meaning in the context of the sentence.

The model also exhibits serial and parallel processing characteristics. Serial processes include the sequential flow of a word from Phonetic to Phon-cat-mean to Grammar, and the sequential flow of words in the sentence. Parallel processing is illustrated by the fact that processing at each level of knowledge in the model goes on simultaneously. For example, the third word in a sentence may just be entering the system in Phonetic, while at the same time multiple meanings of the second word are being triggered in Phon-cat-mean, while at the same time the Grammar is predicting the grammatical category of the second word based on its processing of the first word.

The model also has neurologically-analogous thresholding mechanisms. For example, a word's multiple meanings have variable competing values or potentials over the course of sentence processing. When the Grammar disambiguates meaning, one of the multiple meanings will "fire" (be chosen) while competing meanings will be inhibited.

Information in the model decays over time if it is not reinforced or inhibited; this is analogous to short term memory limits. For example, words at the phonetic level may no longer be available by the time the meaning of the sentence has been interpreted. The analogous psycholinguistic observation is our ability to paraphrase some sentences' meanings without being able to repeat them verbatim.

The model also has other neurologically analogous timing properties which include a refractory period following thresholds, in which the entity reaching threshold cannot be affected by other events for a set period of time. There is also a post-refractory state in which decay of information is not as rapid as it is in short term memory decay.

Finally, two important characteristics of the model are that its knowledge is degradable and its timing characteristics are modifiable. These characteristics will be made clear later when the performance of the model under lesion conditions is described.

It is extremely difficult and time-consuming to illustrate what goes on in the model during sentence processing but its activity can be broadly summarized as follows: the normal model processes information in series and in parallel, with bottom-up and top-down interactions between levels of processing, and with reinforcement, inhibition, and decay of information over time until the sentence's meaning is built to an extent that an unambiguous interpretation of its meaning is made. (For details of the model, see Gigley, 1982a, 1982b, to appear).

TESTING THE MODEL'S ASSUMPTIONS

Where do clinical studies come into the picture? One place is at the level of testing the model's assumptions as they relate to aphasia. For example, the ability of aphasic patients to identify multiple meanings of words is a question raised by the model's assumption that multiple meanings of words are accessed during comprehension. This question has received little or no attention clinically and it illustrates a salutary effect of the modelling process; its demands for explicit hypotheses about processing raise questions with potential clinical relevance which otherwise may have been overlooked.

To answer this question in a preliminary way, the ability of a group of 10 aphasic (PICA Overall Mean Percentile = 63; range = 40-93) and eight control subjects to identify multiple meanings of picturable words was examined. The multiple meaning test consisted of 20 words whose meanings were pictured among six response choices. For each item, the examiner said the word and the subject pointed to all pictures among the six choices that could represent the word. The number of correct meanings pictured ranged from zero to six and averaged three.

The primary findings of the study are summarized in Table 1. They indicate that the aphasic group identified multiple meanings on fewer items with multiple meanings and identified fewer of the total meanings on the test than did the control subjects. Individually, all aphasic patients showed some ability to identify multiple meanings, with half performing within two standard deviations of normal. (For the details of this clinical study, see Gigley, 1982a.)

Table 1. Summary of performance of control and aphasic groups for number of items with multiple meanings on which multiple meanings were identified, and for the total number of meanings represented in the test which were identified.

	Number of Items with Multiple Responses*	Total Meanings Identified**
Control (N=8)		
$\overline{X}$	17.5 (91%)	52.8 (88%)
SD	1.1	4.8
Range	15 - 18	44 - 58
Aphasic (N=10)		
$\overline{X}$	15.4 (81%)***	43.2 (72%)***
SD	3.2	9.7
Range	8 - 18	26 - 55

* Maximum possible score = 19
** Maximum possible score = 60
*** Significantly different from control group ($p < .05$)

These results suggest two possibilities for lesion experiments with the model. In one, multiple meaning access should be left intact, with sentence comprehension deficits induced by lesions in other aspects of the model. In the other, multiple meaning access should be restricted to varying degrees and the effect of such restriction on sentence comprehension observed. The model's performance under these two lesion conditions would then be validated by comparing its performance to the sentence comprehension performance of aphasic patients with and without problems with multiple meaning access.

LESION SIMULATION

"Determiner Lesion"

What follows is an example of the model's performance on a lesion simulation and its relationship to clinical activity.

Based on clinical evidence that patients with so-called Broca's aphasia may have problems related to the comprehension of function words (Bradley, Garrett, and Zurif, 1980; Zurif, Caramazza, and Myerson, 1972; Zurif et al., 1976), the model's sentence comprehension was observed under conditions in which information about the determiner "the" was lesioned. Specifically, the model functioned normally in all respects with the exception that it did not know that a noun should follow a determiner, so whenever "the" was processed a common noun was not predicted.

The sentence "the boy saw the building" will be used to illustrate the model's performance. In the clinical setting if we suspected difficulty with determiners, logic might tell us that the sentence could be interpreted as "boy saw building," and that the patient might interpret "building" as a verb. A point-to comprehension task for this sentence would probably include a picture of the correct sentence representation, as well as a foil picturing a boy looking at a building activity, thus allowing for detection of a verb interpretation of "building." We might also include semantically related foils such as a girl looking at a building or a boy looking at a car.

The model's performance did not precisely correspond to the logical prediction. In fact, the only clearly interpreted word in the sentence was "building" and it was interpreted as a verb. "Boy" was interpreted as a noun but was never identified as the subject of the sentence. Of greatest interest was the finding that all meanings of the word "saw" were still active at the end of the sentence; the multiple meanings of "saw" were never disambiguated. With the picture choices just mentioned, the model may have chosen the foil with the boy looking at the building activity because it interpreted "building" as a verb. But the model's (and perhaps the patient's) decision would have been considerably harder if foils also included: a boy building something; a picture of a saw, representing the noun meaning of "saw"; a boy sawing something, perhaps a building; or perhaps even a foil with a boy, a saw, and a building shown as separate entities.

If it is not already apparent, it should be pointed out that at this stage in the model's development it would be totally unjustified to claim that its interpretation of this sentence corresponds to clinical reality for any aphasic patient. The value of this lesion run is that it raises some interesting questions about aphasic performance that might otherwise have been overlooked, and it generates some ideas about clinical test construction that potentially represent diagnostic refinements.

"Slowed Propagation"

The previous example could be thought of as a lesion of knowledge or competence. This example is of a timing or performance deficit. Again, we'll use the sentence "the boy saw the building." This time, knowledge in the model was left intact but the rate at which each new word in the sentence entered the system was speeded up relative to the rate at which words were processed or propagated once they entered the system. This problem may roughly correspond to the pattern of auditory deficit described by Brookshire (1974) as information capacity deficit. Such patients are believed to have trouble receiving and processing information at the same time, and their performance subsequently appears alternately good and poor within a message.

When this slowed propagation was induced in the model, it performed as follows. As normally occurs, "the" was processed and predicted a common noun to follow. However, by the time that prediction was made, "boy" had already been processed but without the influence of the grammatical prediction. "Saw" entered the system and activation of its multiple meanings was influenced by the prediction of the common noun by "the." In effect, the model understood "the saw." "Boy," because it never interacted with grammar, simply decayed out of the system by the end of the sentence.

Whether this result is clinically "real" or not remains to be determined. The potential value of this lesion run is that it may be detailing the abnormal processing interactions which translate into clinical observations of alternately good and poor comprehension within sentences. It also has implications for the design of comprehension tasks, in that it makes explicit predictions about sentence interpretation which are difficult to make on the basis of clinical intuition alone. In addition, without repairing the lesion in the model, its performance can be "normalized" by slowing the rate at which new words are introduced so that they are in synchrony with the slow rate of word propagation through the system. This treatment of the model is highly similar to Brookshire's (1974) recommendation that patients with information capacity deficit be treated by inserting pauses within messages.

CONCLUSION

Before concluding, we want to acknowledge that the model we have described is in its infancy, that it does not presently account for many factors known to affect language comprehension, and that it undoubtedly will make numerous false predictions about normal and aphasic language performance. If it is to be validated in a way that contributes to the clinical understanding, assessment, and treatment of aphasia, constant interaction between the model and clinical observations will be necessary. Clinical observations will determine how the model should be lesioned and the model's performance under lesion conditions will generate new hypotheses about aphasic performance, which, in turn, will have to be verified in the clinic. Mismatches between the artificial aphasia and the clinical aphasia will be used to further modify the model.

Finally, as clinicians we should recognize that technological and theoretical advances in Artificial Intelligence are making computational neurolinguistics a reality (for relevant discussion, see Arbib and Caplan, 1979; Arbib, 1982; Marcus, 1982; Lavorel, 1982). In the future, numerous computational models of language _and_ aphasia will probably appear and claim, or be assumed to have, psychological validity. The danger of this—for those of us interested in clinical reality—is that we are potentially entering an age of modern-day "diagram-makers" (Head, 1926), whose versions of language and its pathology at best will have no implications for clinical practice, and at worst will be elegantly misleading. If artificial aphasiology is to contribute anything to Clinical Aphasiology, clinical intelligence will have to play an important role in the process.

REFERENCES

Arbib, M.A. From artificial intelligence to neurolinguistics. In M.A. Arbib, D. Caplan and J.C. Marshall (Eds.), _Neural Models of Language Processes_. New York: Academic Press, 1982.

Arbib, M.A. and Caplan, D. Neurolinguistics must be computational. _Behavioral and Brain Sciences_, 2, 449-483, 1979.

Bradley, D., Garret, and Zurif, E. Syntactic deficits in Broca's aphasia. In D. Caplan (Ed.), _Biological Studies of Mental Processes_. Cambridge: MIT Press, 1982.

Brookshire, R.H. Differences in responding to auditory verbal materials among aphasia patients. _Acta Symbolica_, 1-18, 1974.

Erman, L., Hayes-Roth, E., Lesser, V., and Reddy, D.R. The HEARSAY-II speech-understanding system: Integrating knowledge to resolve uncertainty. Computing Surveys, 12, 213-253, 1980.

Gigley, H.M. Neurolinguistically constrained simulation of sentence comprehension: Integrating artificial intelligence and brain theory. Ph.D. Dissertation, University of Massachusetts, 1982a.

Gigley, H.M. A computational neurolinguistic approach to processing models of sentence comprehension. (COINS Tech. Report 82-9). Amherst: University of Massachusetts, March, 1982b.

Gigley, H.M. Artificial intelligence meets brain theory: An integrated approach to simulation modelling of natural language processing. In H. Trappl (Ed.), Proceedings of the sixth European meeting on cybernetics and systems research. North-Holland, to appear.

Head, H. Aphasia and Kindred Disorders of Speech. New York: MacMillan, 1926.

Katz, R.C. and Tong Nagy, V. Development of a computerized treatment system for chronic aphasic patients. Paper presented at the twelfth annual Clinical Aphasiology Conference, Oshkosh, Wisconsin, June, 1982.

Lavorel, P.M. Production strategies: A systems approach to Wernicke's aphasia. In M.A. Arbib, D. Caplan, and J.C. Marshall (Eds.), Neural Models of Language Processes. New York: Academic Press, 1982.

Marcus, M. A theory of syntactic recognition for natural language. Ph.D. Dissertation, M.I.T, 1978.

Marcus, M.P. Consequences of functional deficits in a parsing model: Implications for Broca's aphasia. In M.A. Arbib, D. Caplan, and J.D. Marshall (Eds.), Neural Models of Language Processes. New York: Academic Press, 1982.

Marslen-Wilson, W. and Tyler, L. The temporal structure of spoken language understanding. Cognition, 8, 1-71, 1980.

Mills, R.H. Microcomputerized auditory comprehension training. Paper presented at the twelfth annual Clinical Aphasiology Conference, Oshkosh, Wisconsin, June, 1982.

Newell, A. Harpy, production systems, and human cognition. CMU Technical Report, Carnegie-Mellon University, 1978.

Rieger, C. and Small, S. Toward a theory of distributed word expert natural language parsing. IEEE Transactions on Systems, Man, and Cybernetics, 11, 43-51, 1981.

Seidenberg, M. and Tannehaus, M. Chronometric studies of lexical ambiguity resolution. Proceedings of the Conference of the ACL, 155-157, 1980.

Simon, H.A. Studying human intelligence by creating artificial intelligence. American Scientist, 69, 300-309, 1981.

Small, S. Word expert parsing: A theory of distributed word-based natural language understanding. Ph.D. Dissertation, University of Maryland, 1980.

Swinney, D.A. Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T.F. Myers, J. Laver, and J. Anderson (Eds.), The Cognitive Representation of Speech. New York: North-Holland, in press.

Wilks, Y. Parsing English II. In Charniak and Wilks (Eds.), Computational Semantics. New York: North-Holland, 1976.

Winograd, T. Understanding Natural Language. New York: Academic Press, 1972.

Woods, W.A. WHIM: A speech understanding system on a computer. In M.A. Arbib, D. Caplan, and J.C. Marshall (Eds.), Neural Models of Language Processes. New York: Academic Press, 1982.

Woods, W.A., Bates, M., Brown, G., Bruce, B., Cook, C., Klovstad, J., Makhoul, J., Nash-Webber, B., Schwartz, R., Wolf, J., and Zue, V. WHIM, Speech Understanding Systems: Final report. (BBN Report No. 3438) Cambridge, Mass., December, 1976.

Zurif, E.B., Caramazza, A., and Myerson, R. Grammatical judgments of agrammatic aphasics. Neuropsychologia, 10, 405-417, 1972.

Zurif, E.B., Green, E., Caramazza, A., and Goodenough, C. Grammatical intuitions of aphasic patients: Sensitivity to functors. Cortex, 12, 183-186, 1976.