CONTROL STRUCTURES AND THEORIES OF INTERACTION IN SPEECH UNDERSTANDING SYSTEMS

E.J. Briscoe and B.K. Boguraev
University of Cambridge, Computer Laboratory
Corn Exchange Street, Cambridge CB2 3QG, England

ABSTRACT

In this paper, we approach the problem of organisation and control in automatic speech understanding systems firstly, by presenting a theory of the non-serial interactions necessary between two processors in the system; namely, the morphosyntactic and the prosodic, and secondly, by showing how, when generalised, this theory allows one to specify a highly efficient architecture for a speech understanding system with a simple control structure and genuinely independent components. The theory of non-serial interactions we present predicts that speech is temporally organised in a very specific way; that is, the system would not function effectively if the temporal distribution of various types of information in speech were different. The architecture we propose is developed from a study of the task of speech understanding and, furthermore, is specific to this task. Consequently, the paper argues that general problem solving methods are unnecessary for speech understanding.

I INTRODUCTION

It is generally accepted that the control structures of speech understanding systems (SUSs) must allow for non-serial interactions between different knowledge sources or components within the system. By **non-serial interaction** (NSI) we refer to communication which extends beyond the normal, serial, flow of information entailed by the tasks undertaken by each component. For example, the output of the word recognition system will provide the input to morphosyntactic analysis, almost by definition; however, the operation of the morphosyntactic analyser should be constrained on some occasions by prosodic cues: say, that *her* is accented and followed by a "pause", whilst *dog* is not, in

(1) *Max gave her dog biscuits.*

Similarly, the output of the morphosyntactic analyser will provide the input to semantic analysis, but on occasion, the operation of the morphosyntactic analyser will be more efficient if it has access to information about the discourse: say, that *the horse* has no unique referent in

(2) *The horse raced past the barn fell,*

because this information will facilitate the reduced relative interpretation (see Crain & Steedman, in press). Thus, NSIs will be required between components which occur both before and after the morphosyntactic analyser in the serial chain of processors which constitute the complete SUS.

NSIs can be captured in a strictly serial, hierarchical model, in which the flow of information is always "upwards", by computing every possibility compatible with the input at each level of processing. However, this will involve much unnecessary computation within each separate component which could be avoided by utilising information already temporally available in the signal or context of utterance, but not part of the input to that level. An alternative architecture is the heterarchical system; this avoids such inefficiency, in principle, by allowing each component to communicate with all other components in the system. However, controlling the flow of information and specifying the interfaces between components in such systems has proved very difficult (Reddy & Erman, 1975). The most sophisticated SUS architecture to date is the blackboard model (Erman at al., 1980). The model provides a means for common representation and a global database for communication between components and allows control of the system to be centralised and relatively independent of individual components. The four essential elements of the model - blackboard entries, knowledge sources, the blackboard and an intelligent control mechanism - interact to emulate a problem solving style that is characteristically incremental and opportunistic. NSIs are thus allowed to occur, in principle, when they will be of greatest value for preventing unnecessary computation.

What is striking about these system architectures is that they place no limits on the kinds of interaction which occur between components, that is, none of them are based on any **theory** of what kind of interactions and communication will be needed in a SUS. The designers of the Hearsay-II system were explicit about this, arguing that what was required was an architecture capable of supporting **any** form of interaction, but which was still relatively efficient (Erman & Lesser, 1975:484).

There appear to be at least two problems with such an approach. Firstly, the designer of an individual component must still take into account which other components should be activated by its outputs, as well as who provides its inputs, precisely because no principles of interaction are provided by the model. This entails, even within the loosely structured aggregation hierarchy of the blackboard, some commitment to decisions about inter-component traffic in information - rational answers to these decisions cannot be provided without a theory of interaction between individual components in a SUS.

Secondly, a considerable amount of effort has gone into specifying global scheduling heuristics for maintaining an agenda of knowledge source activation records in blackboard systems, and this has sometimes led to treating the control problem as a distinct issue independent of the domain under consideration, localising it on a separate, scheduling, blackboard (Balzer, Erman and London, 1980; Hayes-Roth, 1983a). Once again, this is because the blackboard framework, as it is defined, provides no inherent constraints on interactions (Hayes-Roth, 1983b). While this means that the model is powerful enough to replicate control strategies used in qualitatively different AI systems, as well as generalise to problem-solving in multiple domains (Hayes-Roth, 1983a), the blackboard method of control still fails to provide a complete answer to the scheduling problem. It is intended predominantly for solving problems whose solution depends on heuristics which must cope with large volumes of noisy data.

In the context of a blackboard-based SUS, where the assumption that the formation of the "correct" interpretation of an input signal will, inevitably, be accompanied by the generation of many competing (partial) interpretations is implicit in the redundancy encoded in the individual knowledge sources, the only real and practical answer to the control problem remains the development of global strategies to keep unnecessary computation within practical limits. These strategies are developed by tuning the system on the basis of performance criteria; this tuning appears to limit interactions to just those optimal cases which are likely to yield successful analyses. However, insofar as the final system might claim to embody a theory about which interactions are useful, this will never be represented in an explicit form in the loosely structured system components, but only implicitly in the the run-time behaviour of the whole system; and therefore is unlikely to be recoverable (see the analogous criticism in Hayes-Roth, 1983a:55).

I INTERACTIVE DETERMINISM: A THEORY OF NON-SERIAL INTERACTION

In this section, we concentrate on the study of NSI between morphosyntactic and prosodic information in speech, largely from the perspective of morphosyntactic analysis. This interaction occurs between two of the better understood components of a SUS and therefore seems an appropriate starting point for the development of a theory of NSIs.

Lea (1980) argues that prosodic information will be of use for morphosyntactic processing. This discussion is based on the observation (see Cooper & Paccia-Cooper, 1980; Cooper & Sorenson, 1981), that there is a strong correlation between some syntactic boundaries and prosodic effects such as lengthening, step up in fundamental frequency, changes of amplitude and, sometimes, pausing. However, many of these effects are probably irrelevant to morphosyntactic analysis, being, for example, side effects of production, such as planning, hesitation, afterthoughts, false starts, and so forth. If prosody is to be utilised effectively to facilitate morphosyntactic analysis, then we require a theory capable of indicating when an ambiguous prosodic cue such as lengthening is a consequence of syntactic environment and, therefore, **relevant** to morphosyntactic analysis. None of Lea's proposals make this distinction.

In order to develop such a theory, we require a precise account of morphosyntactic analysis embedded in a model of a SUS which specifies the nature of the NSIs available to the morphosyntactic analyser. Consider a simple modular architecture of a SUS in which most information flows upwards through each level of processing, as in the serial, hierarchical model. This information is passed without delay, so any operation performed by a processor will be passed up to its successor in the chain of processors immediately (see Fig.1).

Furthermore, we constrain the model as follows: at least from the point of word recognition upwards, only one interpretation is computed at each level. That is, word recognition returns a series of unique, correct words, then morphosyntactic analysis provides the unique, correct grammatical description of these words, and so forth. In order to implement such a constraint on the processing, the model includes, in addition to the primary flow of information, secondary channels of communication which provide for the NSIs (represented by single arrows in the diagram). These interactive channels are bidirectional, allowing one component to request certain highly restricted kinds of information from another component and, in principle, can connect any pair of processors in a SUS

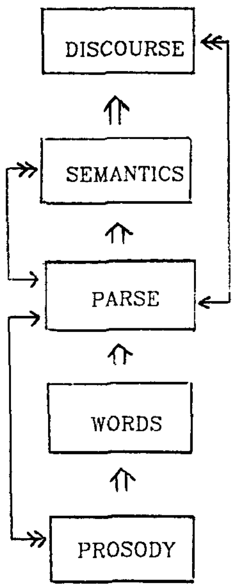

Fig.1*

Imagine a morphosyntactic analyser which builds a unique structure without backtracking and employs no, or very little, look-ahead Such a parser will face a choice point, irresolvable morphosyntactically, almost every time it encounters a structural ambiguity, whether local or global. Further, suppose that this parser seeks to apply some general strategies to resolve such choices, that is, to select a particular grammatical interpretation when faced with ambiguity. If such a parser is to be able to operate deterministically, and still return the correct analysis without error, in cases when a general strategy would yield the wrong analysis, then it will require interactive channels for transmitting a signal capable of blocking the application of the strategy and forcing the correct analysis. These are the secondary channels of communication posited in the model of the SUS above.

A theory of NSIs should specify **when,** in terms of the operation of any individual processor, interaction will be necessary; interactive channels for this parser must be capable of providing this information at the **onset** of any given morphosyntactic ambiguity, which is defined as the point at which the parser will have to apply its resolution strategy. In order to make the concept of onset of ambiguity precise a model of the morphosyntactic component of a SUS was designed and implemented. This analyser (henceforth the LEXical-CATegorial parser - because it employs an Extended Categorial Grammar (eg. Ades & Steedman, 1982) representing morphosyntactic information as an extension of the lexicon) makes specific predictions about the temporal availability of **non-morphosyntactic information crucial to the theory of NSIs** presented here. LEXICAT's strategy for resolution of ambiguities is approximately a combination of late closure (Frazier, 1979) and right association (Kimball, 1973). LEXICAT is a species of shift-reduce parser which employs the same stack for the storage and analysis of input and inspects the top three cells of the stack before each parsing operation. Reduction, however, never involves more than two cells, so the top cell of the stack acts as a very restricted one word look-ahead buffer. In general, LEXICAT reduces the items in cells two and three provided that reduction between cells one and two is not grammatically possible*.

When LEXICAT encounters ambiguity, in the majority of situations this surfaces as a choice between shifting and reducing. When a shift-reduce choice arises between either cells one and two or two and three, reduction will be preferred **by default;** although, of course, a set of interactive requests will be generated at the point when this choice arises, and these may provide information which blocks the preferred strategy. The approximate effect of the preference for reduction is that incoming material is attached to the constituent currently under analysis which is "lowest" in the phrase structure tree. LEXICAT is similar to recent proposals by Church (1980), Pereira (in press) and Shieber (1983), in that it employs general strategies, stated in terms of the parser's basic operations, in order to parse deterministically with an ambiguous grammar.

A theory of NSIs should also specify **how** interaction occurs. When LEXICAT recognises a choice point, it makes a request for non-morphosyntactic information relevant to this choice on all of the interactive channels to which it is connected; if any of these channels returns a positive response, the default interpretation is overridden. The parser is therefore agnostic concerning which channel might provide the relevant information; for example, analysing

(3) *Before the King* ***rides*** *his horse*
it's usually groomed.

The onset of this morphosyntactic ambiguity arises when ***the horse*** has been analysed as a noun phrase. LEXICAT must decide at this point whether ***rides*** is to be treated as transitive or intransitive; the transitive

----

* This diagram is not intended to be complete and is only included to illustrate the two different types of communication proposed in this paper.

----

* This is not completely accurate; see Briscoe 1984:Ch3 for a full description of LEXICAT.

reading is preferred given the resolution strategy outlined above. Therefore, an interactive request will be generated requesting information concerning the relationship between these two constituents. A simple yes/no response is all that is needed along this interactive channel: "yes" to prevent application of the strategy, "no" if the processor concerned finds nothing relevant to the decision. In relation to this example, consider the channel to the prosodic analyser which monitors for prosodic "breaks" (defined in terms of vowel lengthening, change of fundamental frequency and so forth): when the request is received the prosodic analyser returns a positive response if such a break is present in the appropriate part of the speech signal. In (3) none of these cues is likely to occur since the relevant boundary is syntactically weak (see Cooper & Paccia-Cooper, 1980), so the interactive request will not result in a positive response, the default resolution strategy will apply and ***his horse*** will be interpreted as direct object of ***rides***. In

(4) *Before the King rides his horse is usually groomed*,

on the other hand, an interactive request will be generated at the same point, but the interactive channel between the prosodic and morphosyntactic components is likely to produce a positive response since the boundary between ***rides*** and ***his horse*** is syntactically stronger. Thus, attachment will be blocked, closing the subordinate clause, and thereby forcing the correct interpretation.

NSI then, is restricted to a set of yes/no responses* over the interactive channels at the explicit request of the processor connected to those channels, where a positive response on one interactive channel suffices to override the unmarked choice which would be made in the absence of such a signal. This highly restricted form of interaction is sufficient to guarantee that LEXICAT will produce the correct analysis even in cases of severe multiple ambiguity; for example, analysing the noun compound in

(5) *Boron epoxy rocket motor chambers*,

(from Marcus, 1980:253), there are fourteen+ licit morphosyntactic interpretations, assuming standard grammatical analyses (eg. Selkirk, 1983). However, if this example were spoken and we assume that it would have the prosodic structure predicted by Cooper & Paccia-Cooper's (1980) algorithm for deriving prosody from syntactic structure, LEXICAT could produce the correct analysis without error, just through interaction with the prosodic analyser. As each noun enters the analyser, reduction will be blocked by the general strategy but, because LEXICAT will recognise the existence of ambiguity, an interactive request will be generated before each shift. The prosodic break channel will then prevent reduction after ***epoxy*** and after ***motor***, forcing the correct analysis ***((boron epoxy) ((rocket motor) chambers))***, as opposed to the default right-branching structure.

Thus, NSI between the morphosyntactic and prosodic components can be captured by a bistable, bidirectional link capable of transmitting a request and signaling a binary reponse, either blocking or allowing the application of the relevant strategy according to the presence or absence of a prosodic break. Given the simplicity of this interaction, the prosodic analyser requires no more information from the parser than that a decision is requested concerning a particular boundary. Nor need the prosodic analyser decide, prior to an interactive request on this channel, whether a particular occurrence of, say lengthening, is signalling the presence of a prosodic break, rather than for instance stress, since the request itself will help resolve the interpretation of the cue. Moreover, we have a simple generalisation about when interactive requests will be made since this account of NSIs predicts that prosodic information will only be relevant to morphosyntactic analysis at the onset of a morphosyntactic ambiguity.

If we assume (boldly) that this account of NSI between the morphosyntactic and prosodic analysers will generalise to a complete model of SUS, then such a model makes a set of predictions concerning the temporal availability of interactive information in the speech signal and representation of the context of utterance. In effect, it claims that the SUS architecture simply presupposes that language is organised in the appropriate fashion since the model will not function if it is not. We call this strong prediction about the temporal organisation of the speech signal the Interactive Determinism (ID) Hypothesis since it is essentially an extension of Marcus' (1980) Determinism Hypothesis.

II TESTING THE INTERACTIVE DETERMINISM HYPOTHESIS

The ID hypothesis predicts that speech and the representation of context is organised in such a way that information will be available, when needed, via NSI to resolve a choice in any individual component at the point when that choice arises. Thus in the case of prosodic interaction with morphosyntactic analysis the theory predicts that a prosodic break should be present in speech at the onset of a morphosyntactic

---

* Possibly these responses should be represented as confidence ratings rather than a discrete choice. In this case levels of certainty concerning the presence/absence of relevant events could be represented. However, for the rest of this paper we assume binary channels will suffice.

\+ Corresponding to the Catalan numbers; see Martin et al. (1981).

ambiguity which requires a non-default interpretation **and which is not resolved by other non-morphosyntactic information.** This aspect of the ID hypothesis has been tested and corroborated by Paul Warren (1983; in prep; also see Briscoe, 1984:Ch4), who has undertaken a series of speech production experiments in which (typically) ten subjects read aloud a list of sentences. This list contains sets of pairs of locally ambiguous sentences, and some filler sentences so that the purpose of the experiment is not apparent to the subjects. Their productions are analysed acoustically and the results of this analysis are then checked statistically. The technique gives a good indication of whether the cues associated with a prosodic break are present at the appropriate points in the speech signal, and their consistency across different speakers.

Returning to examples (3) and (4) above, we noted that a prosodic break would be required in (4), but not (3), to prevent attachment of ***rides*** and ***his horse***. Warren found exactly this pattern of results; the duration of ***rides*** (and similar items in this position) is an average 51% longer in (4) and the fall in fundamental frequency is almost twice as great with a corresponding step up to ***horse***, as compared to a smooth declination across this boundary in (3). Similarly, analysing

(6) *The company awarded the contract [to/was] the highest bidder.*

LEXICAT prefers attachment of ***The company*** to ***awarded***, treating ***awarded*** as the main verb. In the case where ***awarded*** must be treated as the beginning of a reduced relative, Warren found that the duration of the final syllable of ***company*** is lengthened and that the same pattern of fall and step up in fundamental frequency occurs. Perhaps the most interesting cases are ambiguous constituent questions; Church (1980,117) argued that it is probably impossible to parse these deterministically by employing look-ahead:

> "The really hard problem with wh-movement is finding the "gap" where the wh-element originated. This is not particularly difficult for a non-deterministic competence theory, but it is (probably) impossible for a deterministic processing model."

LEXICAT predicts that in a sentence such as

(7) *Who did you want to give the presents to Sue?*

the potential point of attachment of ***Who*** as direct object of ***want*** will be ignored by default in preference for the immediate attachment of ***to give***. Thus there is a prediction that the sentence, when spoken, should contain a prosodic break at this point. Warren has found some evidence for this prediction, i.e. ***want*** is lengthened as compared to examples where this is not the correct point of attachment of the preposed phrase, such as

(8) *Who did you want to give the presents to?*

but the prosodic cues, although consistent, are comparatively weak, and it is not clear that listeners are utilising them in the manner predicted by the theory (see Briscoe, 1984:Ch4).

A different kind of support is provided by sentences such as

(9) *Before the King rides a servant grooms his horse.*

which exhibit the same local ambiguity as (3) and (4) but where the semantic interpretation of the noun phrase makes the direct object reading implausible. In this case it is likely that an interactive channel between the semantic and morphosyntactic analysers would block the incorrect interpretation. So there is a prediction that the functional load on prosodic information will decrease and, therefore, that the prosodic cues to the break may be less marked. This prediction was again corroborated by Warren who found that the prosodic break in examples such as (9) was significantly less marked acoustically than for examples such as (4)*. In general then, these experimental results support the ID hypothesis.

III CONTROL STRUCTURE AND ORGANISATION

In a SUS based on the ID model, the main flow of information will be defined by the tasks of each component, and their medium of communication, will be a natural consequence of these tasks; as for the serial, hierarchical model. However, in the ID model, unlike the hierarchical model, there are less overheads because unnecessary computation at any level of processing will be eliminated by the NSIs between components. These interactions will, of course, require a large number of interactive channels; but these do not imply a common representation language because the information which passes along them is representation-independent and restricted to a minimal request and a binary response. Each channel in the full SUS will be dedicated to a specific interaction between components; so the morphosyntactic component will require a prosodic break channel and a unique referent channel (see example (1)), and so forth. Thus, a complete model of SUS will implement a theory of the types of NSI required between all components. Finally, the ID model will not require that any individual processor has knowledge of the nature of the operations of another processor; that is, the

----

* Note that this result is inexplicable for theories which attempt to derive the prosodic structure of a sentence directly from its syntactic structure; see Cooper & Paccia-Cooper (1980:181f).

morphosyntactic analyser need not know what is being computed at the other end of the prosodic break channel, or how: nor need the prosodic analyser know why it is computing the presence or absence of a prosodic break. Rather, the knowledge that this information is potentially important is expressed by the existence of this particular interactive channel.

The control structure of this model is straightforward; after each separate operation of each individual component the results of this operation will be passed to the next component in the serial chain of processors. An interactive request will be made by any component only when faced with an indeterminism irresolvable in terms of the input available to it. No further scheduling or centralised control of processing will be required. Furthermore, although each individual component determines when NSIs will occur, because of the restricted nature of this interaction each component can still be developed as a completely independent knowledge source.

The deterministic nature of the individual components of this SUS eliminates the need for any global heuristics to be brought into the analysis of the speech signal. Thus we have dispensed neatly with the requirement for an over-powerful and over-general problem-solving framework, such as the blackboard, and replaced it with a theory specific to the domain under consideration; namely, language. The theory of NSIs offers a satisfactory specific method for speech understanding which allows the separate specialist component procedures of a SUS to be "algorithmetized" and compiled. As Erman et al. (1980:246) suggest: "In such a case the flexibility of a system like Hearsay-II may no longer be needed".

The restrictions on the nature and directionality of NSI channels in a SUS, and the situations in which they need to be activated, allows a modular system whose control structure is not much more complex than that of the hierarchical model, and yet, via the network of interactive channels, achieves the efficiency sought by the heterarchical and blackboard models, without the concomitant problems of common knowledge representations and complex communications protocols between separate knowledge sources. Thus, the ID model dispenses with the overhead costs of data-directed activation of knowledge sources and the need for opportunistic scheduling or a complex focus-of-control mechanism.

IV CONCLUSION

In this paper we have proposed a very idealised model of a SUS with a simple organisation and control structure. Clearly, the ID model assumes a greater level of understanding of many aspects of speech processing than is current. For example, we have assumed that the word recognition component is capable of returning a series of unique, correct lexical items; even with interaction of the kind envisaged, it is doubtful that our current understanding of acoustic-phonetic analysis is good enough for it to be possible to build such a component now. Nevertheless, the experimental work reported by Marslen-Wilson & Tyler (1980) and Cole & Jakimik (1980), for example, suggests that listeners are capable of accessing a unique lexical item on the basis of the acoustic signal and interactive feedback from the developing analysis of the utterance and its context (often before the acoustic signal is complete). More seriously, from the perspective of interactive determinism, little has been said about the many other interactive channels which will be required for speech understanding and, in particular, whether these channels can be as restricted as the prosodic break channel. For example, consider the channel which will be required to capture the interaction in example (9); this will need to be sensitive to something like semantic "anomaly". However, semantic anomaly is an inherently vague concept, particularly by comparison with that of a prosodic break. Similarly, as we noted above, the morphosyntactic analyser will require an interactive channel to the discourse analyser which indicates whether a noun phrase followed by a potential relative clause, such as ***the horse*** in (2), has a unique referent. However, since this channel would only seem to be relevant to ambiguities involving relative clauses, it appears to cast doubt on the claim that interactive requests are generated automatically on every channel each time any type of ambiguity is encountered. This, in turn, suggests that the control structure proposed in the last section is oversimplified.

Nevertheless, by studying these tasks in terms of far more restricted and potentially more computationally efficient models, we are more likely to uncover restrictions on language which, once discovered, will take us a step closer to tractable solutions to the task of speech understanding. Thus, the work reported here suggests that language is organised in such a manner that morphosyntactic analysis can proceed deterministically on the basis of a very restricted parsing algorithm, because non-structural information necessary to resolve ambiguities will be available in the speech signal (or representation of the context of utterance) **at the point when the choice arises during morphosyntactic analysis.**

The account of morphosyntactic analysis that this constraint allows is more elegant, parsimonious and empirically adequate than employing look-ahead (Marcus, 1980). Firstly, an account based on look-ahead is forced to claim that local and global ambiguities are resolved by different mechanisms (since the latter, by definition, cannot be resolved by the use of morphosyntactic information further downstream in the signal), whilst the ID model requires only one mechanism. Secondly, restricted look-ahead fails to delimit accurately the class of so-called garden path sentences (Milne, 1982; Briscoe, 1983), whilst the ID account correctly predicts their "interactive" nature (Briscoe, 1982, 1984; Crain & Steedman, in press). Thirdly, look-ahead involves delaying decisions, a strategy which is made implausible, at least in the context of speech understanding, by the body of experimental results summarised by Tyler (1981), which suggest that morphosyntactic analysis is extremely rapid.

The generalisation of these results to a complete model of SUS represents commitment to a research programme which sets as its goal the discovery of constraints on language which allow the associated processing tasks to be implemented in an efficient and tractable manner. What is advocated here, therefore, is the development of a computational theory of language processing derived through the study of language from the perspective of these processing tasks, much in the same way in which Marr (1982) developed his computational theory of vision.

Acknowledgements: We would like to thank David Carter, Jane Robinson, Karen Sparck Jones and John Tait for their helpful comments. Mistakes remain our own.

V REFERENCES

Ades,A. and Steedman,M.(1982) 'On the Order of Words', *Linguistics and Philosophy, vol.5*, 320-362

Balzer,R., Erman,L., London,P. and Williams,C.(1980) 'HEARSAY-III: A Domain-Independent Framework for Expert Systems', *Proceedings of the AAAI(1)*, Stanford, CA, pp.108-110

Briscoe,E.(1982) 'Garden Path Sentences or Garden Path Utterances?', *Cambridge Papers in Phonetics and Experimental Linguistics, vol.I*, 1-9

Briscoe,E.(1983) 'Determinism and its Implementation in Parsifal' in Sparck Jones,K. and Wilks,Y.(eds.), *Automatic Natural Language Parsing*, Ellis Horwood, Chichester, pp.61-68

Briscoe,E.(1984) *Towards an Understanding of Spoken Sentence Comprehension: The Interactive Determinism Hypothesis*, Doctoral Thesis, Cambridge University

Church,K.(1980) *On Memory Limitations in Natural Language Processing*, MIT/LCS/TR-245

Cole,R and Jakimek,J.(1980) 'A Model of Speech Perception' in Cole,R.(eds), *Perception and Production of Fluent Speech*, Lawrence Erlbaum, New Jersey

Cooper,W. and Paccia-Cooper,J.(1980) *Syntax and Speech*, Harvard University Press, Cambridge, Mass

Cooper,W. and Sorenson,J.(1981) *Fundamental Frequency in Sentence Production*, Springer Verlag, New York

Crain,S. and Steedman,M.(In press) 'On Not Being Led Up the Garden Path: the Use of Context by the Psychological Parser' in Dowty,D., Karttunen,L. and Zwicky,A.(eds.), *Natural Language Processing*, Cambridge University Press, Cambridge

Erman,L., Hayes-Roth,F., Lesser,V. and Reddy,R.(1980) 'The Hearsay-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty', *Computing Surveys, vol.12*, 213-253

Erman,L. and Lesser,V.(1975) 'A Multi-Level Organisation for Problem Solving Using Many, Diverse, Cooperating Sources of Knowledge', *Proceedings of the 4th IJCAI*, Tbilisi, Georgia, pp.483-490

Frazier,L.(1979) *On Comprehending Sentences: Syntactic Parsing Strategies*, IULC, Bloomington, Indiana

Hayes-Roth,B.(1983a) *A Blackboard Model of Control*, Report No.HPP-83-38, Department of Computer Science, Stanford University

Hayes-Roth,B.(1983b) *The Blackboard Architecture: A General Framework for Problem Solving?*, Report No.HPP-83-30, Department of Computer Science, Stanford University

Kimball,J.(1973) 'Seven Principles of Surface Structure Parsing in Natural Language', *Cognition, vol.2*, 15-47

Lea,W.(1980) 'Prosodic Aids to Speech Recognition' in W.Lea(eds.), *Trends in Speech Recognition*, Prentice Hall, New Jersey, pp 166-205

Marcus,M.(1980) *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, Mass.

Marr,D.(1982) *Vision*, W.H.Freeman and Co., San Francisco

Marslen-Wilson,W. and Tyler,L.(1980) 'The Temporal Structure of Spoken Language Understanding: the Perception of Sentences and Words in Sentences', *Cognition, vol 8*, 1-74

Martin,W., Church,K. and Patil,R.(1982) *Preliminary Analysis of a Breadth-First Parsing Algorithm: Theoretical and Experimental Results*, MIT/LCS/TR-261

Milne,R.(1982) 'Predicting Garden Path Sentences', *Cognitive Science, vol.6*, 349-373

Pereira,F.(In press) 'A New Characterization of Attachment Preferences' in Dowty,D., Karttunen,L. and Zwicky,A.(eds.), *Natural Language Processing*, Cambridge University Press, Cambridge

Selkirk,E.(1983) *The Syntax of Words*, MIT Press, Cambridge Mass.

Shieber,S.(1983) 'Sentence Disambiguation by a Shift-

Reduce Parsing Technique', *Proceedings of the 21st Annual Meeting of ACL*, Cambridge, Mass, pp 113-118

Reddy,R. and Erman,L.(1975) 'Tutorial on System Organisation for Speech Understanding' in R.Reddy(eds.), *Speech Recognition: Invited Papers of the IEEE Symposium*, Academic Press, New York, pp.457-479

Tyler,L.(1981) 'Serial and Interactive-Parallel Theories of Sentence Processing', *Theoretical Linguistics*, *vol.8*, 29-65

Warren,P.(1983) 'Temporal and Non-Temporal Cues to Sentence Structure', *Cambridge Papers in Phonetics and Experimental Linguistics, vol.II*

Warren,P.(In prep) *Durational Factors in Speech Processing*, Doctoral Thesis, Cambridge University