Proceedings Paper:
Atwell, ES and Demetriou, G (1994) Semantics in speech recognition and understanding: a survey. In: Evett, L and Rose, T, (eds.) Proceedings of the 1994 AISB Workshop on Computational Linguistics for Speech and Handwriting Recognition. 1994 AISB Workshop on Computational Linguistics for Speech and Handwriting Recognition, 12 April 1994, University of Leeds, UK. AISB.

Semantics in Speech Recognition and Understanding: A Survey

George C. Demetriou and Eric S. Atwell
Artificial Intelligence Division
School of Computer Studies
University of Leeds
e-mail: `george@scs.leeds.ac.uk` `eric@scs.leeds.ac.uk`

Abstract

This survey paper aims at summarizing the state of the art of computational semantic methods in speech recognition and understanding research. A taxonomy classifying the approaches adopted in the literature is divided into six main categories: semantic networks, semantic grammars, caseframes, statistical, unification-based and neural networks. For each approach, an overview of the variety of uses and relative strengths and weaknesses is given.

1 Introduction

In the literature, many methods and techniques have been devised to provide semantic constraints for speech recognition. A thorough classification is rather difficult since most systems were

developed and tuned under different task domains and vocabularies. We tried to survey the entries found in the literature in order to come up to a convergence for a useful classification. The different approaches are classified in six main categories: semantic network approaches, case-frame approaches, unification-based approaches, statistical-modelled approaches and connectionist approaches. Unfortunately due to space limitations, only a summary of each approach and a brief assessment of its advantages and disadvantages can be given. For a more thorough and extended survey and a more complete set of references, the reader is advised to look at Demetriou and Atwell (1994).

2 Semantic network approaches.

The most common axes for structuring knowledge in semantic networks as pointed out by Sagerer and Kummert (1988) are:

• *classification:* a real world object is associated with its generic type so that a concept can be distinguished from its instances (i.e 'instance-of' relation).
• *aggregation:* concepts or instances are related to other concepts or instances respectively of which they may be parts (i.e. 'part' relation).
• *generalization:* a concept is related to more generic ones (i.e. 'isa' relation). In this way, a hierarchy between concepts in the network is defined.

The main advantage of using semantic networks is that restrictions are facilitated in the semantic hierarchy so that generality in role fillers can be acquired by inheritance. They also allow for the specification and checks of relations between concepts and their instances in a sentence (nominal, adjectival, attribute-value, etc.). For speech recognition, semantic networks have been used for the construction of sentence hypotheses guided by concept relation judgements of content words (Sagerer and Kummert 1988), to propose additional words that might have occured in the original utterance but be missing due to poor match quality (Nash-Webber 1975a) or the reduction of syntactically well-formed but semantically inadmissable structures (Niedermair et al 1990).

For speech understanding, the semantic network of a recognized sentence can be linked with a preconstructed network for the input story (as in Shigenaga et al 1986) or for anaphora resolution by attaching semantic categories to semantically empty pronouns (Niedermair et al 1990). Semantic networks have also been used in Brietzmann and Ehrlich (1986) for the EVAR system (to represent a tree-structured inheritance between semantic features), and Hataoka et al (1990) (where 'concept' networks are proposed to represent the meanings of single words as well as three kinds of relational links - 'isa', generic and instance relations).

The main argument against the use of semantic networks is the inability to introduce world knowledge for more general applications and large vocabularies. Expanding the network to new tasks, search space increases drammatically due to the large numbers of concepts and relations affecting the overall system's efficiency. Even if hardware computer power increases to cope with heavy processing, the knowledge capture problem remains: how to create a general purpose, wide-coverage network[1].

3 Semantic grammar variants

Semantic grammars use semantic conditions closely integrated with the syntactic rules of the grammar. To construct semantic rules, syntactic entities are split into meaningful categories in the task domain. For example in (Hayes-Roth 1980), the non-terminal $AUTHOR can represent a main category so that a rule like

'If $AUTHORS=word(1),..., word(n) then form a set of sequence of requested authors including each word as an instance of $AUTHOR'

searches a small subset of its syntactic category instead of the whole class. These grammars are usually represented as transition networks (for example augmented transition networks as in Wolf and Woods 1980 or finite state networks as in Pieraccini and Lee 1991) where the transitions between states in the language environment are between conceptual categories rather than words. The rules can be used for semantic acceptability tests (by associating meaning components directly with the syntactic units) as well as for predicting adjacent constituents and confirming compatible hypotheses by postdiction[2], before, during or after syntactic processing.

Semantic grammars are seen as providing stronger constraints than pure syntactic grammars. They had been popular among the systems developed in the 70s during the first ARPA SUR project like the Hearsay II, HWIM, and Harpy systems (for an overview see Klatt 1977). Nevertheless, later work (Mergel or Paeseler 1987, Hauptmann et al 1988 among others) recognized their practical efficiencies for certain task domains. The main strength of semantic

---

[1]Demetriou and Atwell (1994a) (this Proceedings) propose to build a general-purpose network from a lexical database (LDOCE).

[2]prediction after the event

grammars lies in their ability to balance the satisfiability and diagnosticity[3] of grammatical constraints in order to optimize the computational cost for the particular task. For example (Hayes-Roth 1980), the testing of the category hypotheses $ARTICLE instead of $NOUN is far more diagnostic but not at the same degree computationally expensive. For other approaches in the same spectrum in the literature, the reader can refer to Klovstad and Levinson and Shipley (1980) and Matsunaga et al (1990).

Apart from the inherent difficulties in hand-coding and expanding the grammar for a new application, another limitation of a semantic grammar is the inability to express linguistic generalisations. By this is meant the fact that a syntactic category has to be repeated for every semantic class. This results in the enlargement of syntax and also causes usability problems since it leaves users feeling uncertain about the real linguistic coverage of the system (Thurmair 1988).

4 Caseframe approaches

Semantic constraints expressed in the form of caseframes (Fillmore 1968) have been adopted by a significant number of systems in the literature. The central idea is that of a head concept (generally the main verb or predicate of VP or the head noun of NP) modified by a set of cases (noun or adverbial phrases in VP or modifiers in NP) that play some related role and may in turn correspond to other caseframes. From the recognition point of view, frame-based approaches have been used for the production of sentence hypotheses from a word lattice and the choice of the most likely one (as in Brietzmann and Ehrlich 1986, Poesio and Rullent 1987, Bigorgne et al 1988, Fissore et al 1988, Young et al 1988), for filling gaps of missing words (Hayes et al 1986), for correcting errors in the recognized message (Young 1991) as well as for making word predictions during recognition (Niedermair 1986). Systems that used this technique to verify hypotheses proposed by N-best interfaces also exist (Norton et al 1991, Seneff et al 1991).

For understanding, this approach has been used at a post-recognition stage to disambiguate the recognized utterance and find its meaning representation in order to respond in a dialogue process (Luzzatti 1987, Jackson et al 1991, Rudnicky et al 1991, Ward 1991).

Caseframes are popular because they can combine acoustic reliability with semantic relevance. Unlike network-based techniques, parsing is able to start its interpretation from the most significant parts of the utterance and to extend these islands to the less reliable segments. This is very important for processing both well-formed and ill-formed input. Nevertheless, the recognition method may vary and be adapted to the particular use. This is because caseframes encode semantic information at a more abstract level than ATNs and constraints can be applied in multiple ways (Hayes et al 1986). Another advantage is that once an acceptable caseframe combination is derived, the semantic representation of the utterance is directly derivable.

For dialogue and back-end operations, interesting attempts are presented by Young et al (1988) for the VODIS II (Voice Operated Database Inquiry Systems) system (frames contain nested ranked alternatives of the input speech and knowledge of the task domain i.e. requests through telephone and are used to select a suitable response) and Seneff et al (1991) (where TINA's

[3] Satisfiability is described as a measure of the expected frequency for a test to yield positive results. Diagnosticity, on the other hand, measures the amount of information a constraint adduces (Hayes-Roth 1980).

outputs are transformed into frame representations and integrated with available frames from the history for text and SQL generation).

The main shortcomings of caseframes for speech can be summarized as follows:

• Caseframes rely heavily on finding head words (usually verbs, nouns or adjectives) which are easily distinguishable among others by having a high acoustic score. While for longer words this may be possible, for shorter ones may not (for example, 'rent', 'hat', 'gap', etc.). In addition, there is no efficient method of exploiting word scores in a way that can help the analysis during parsing. This strategy is therefore dependent on the accuracy of the recognizer in assigning the best scores to islands that correspond to frame headers.

• Keeping syntax separate from semantics is not always feasible. This is basically for efficiency reasons since caseframes cause serious computational and memory problems for even simple task domains. This contributes to the loss of syntactic power once the syntax is embedded in the code, and it is difficult maintain each knowledge type separately.

• From the linguistic viewpoint, caseframes have limitations in expressing relations other than 'verb' plus 'sentence function' or 'noun' plus 'attribute' (Niedermair et al 1990). Thus, they are adequate only when they are implicitly present in the speaker's mind (for example, when spacial-temporal relationships are well defined - Luzzatti 1987). It is therefore difficult to apply them to general information dialogue systems. Communication processes built on these are at present possible for well-defined (usually small vocabulary) defined tasks.

5 Statistical approaches

Statistical language modelling has long been advocated by researchers at IBM (Baker 1975, Jelinek 1990) and elsewhere (Atwell 1983). The general idea is to assign a probability to any word string appearing in the lattice. The recognized sentence is the one that maximizes this probability. Probabilistic semantic constraints are expressed in terms of bigram or trigram[4] representations of lexical or semantic classes of words rather than individual words (as in the pure statistical approach). If $U=u_1, u_2,...,u_m$ corresponds to the 'observed' utterance, $W=w_1, w_2,...,w_l$ corresponds to words and for every w_i there is an associated semantic tag s_i, then the meaning can be represented as $S= s_1, s_2,...,s_l$. The system tries to maximize the conidtional probability $P(W,S/A)$ given the acoustic 'observation' A and the maximum aposteriori criterion. The probabilities for these sequences are obtained via training in text corpora. Generally, if the equivalence class of the past string is S_n, the probability of a word w will be estimated by the relative frequency $f(w/S_n)$ in a corpus.

In the literature, most systems have applied this strategy for the disambiguation of recognition lattices (Stern et al 1987, Fissore et al 1989, Paeseler and Ney 1989 and others). Others used it for the correction of recognition errors and the filling of missing words (Ward et al 1988), segmentation of sentences into phrases with semantic relations (Pieraccini and Levin 1992) and reordering sentences hypotheses in N-best lists (Kubala et al 1991).

The main advantages for using statistical-based speech recognition methods are the simplicity

[4]That is, the probability of the next word been uttered depends on the previous word or the two previous words respectively.

and effectiveness for real-time tasks. Other pros are the reliability in ordering multiple sentence hypotheses and the automatic training of parameters from text, thus avoiding the trouble of writing complex grammar rules. The reason for building a language model around semantic classes is that it requires less memory, storage and computational time than simple N-word models.

Rose and Evett (1992), Rose (1993) provide semantic support for a large vocabulary handwriting system based on word similarities from MRDs and text corpora. Their system operates upon well-formed syntactic alternatives. For text recognition, constraints or preferences as expressed by word overlap between sense definitions in MRDs (the Collins English Dictionary and The Oxford Advanced Learner's Dictionary) are used within a fixed-size window in order to determine semantically related content words in the candidate phrase and produce a score according to that overlap. Accordingly, word collocations (which indicate the co-occurence of words into meaningful fragments) as found in the LOB corpus are explored in a similar manner. That is, to discriminate between alternative sequences of candidate words, the program compares the collocational information for each candidate and those around it.

Statistical N-gram language models for speech present clear limitations in expressing semantic constraints since only local context is taken into account (it is difficult to use a large N due to computational reasons). This may result in ungrammatical sentences and unexpected answers especially when there are missing words in the lattice and no appropriate action has been taken. Training sets are, in most cases, found to be small and inadequate to provide statistic coverage for large vocabularies. Training also has the problem that statistical modelling may depend on the domain of the text corpus and applying a model to a new domain requires new training corpora for this domain. Furthermore, as far as semantic restrictions are concerned, it is not always straightforward to specify the optimal number of semantic classes in which words should be grouped for the particular task. Words can be better distinguished if grouped in more classes. However, this is computational expensive (it needs bigger corpora for tagging and training) and not appropriate with noisy input and missing segments (where more general, looser constraints should be used). With fewer classes, the grammar is more robust (a fairly small corpus can provide enough statistical information), but also less accurate. Although there are several tagged corpora to train syntax (tagged LOB corpus, Brown corpus, etc.), there are no available large-scale semantically-tagged corpora to act as training data[5].

6 Unification-based approaches

In unification grammar formalisms (Shieber 1986) linguistic knowledge is structured with featural constraints at the levels of morphology, syntax and semantics that all occur in the same expression. In such grammars, rules are made up of category elements that are not atomic symbols, but complex structures consisting of a category label and attribute specifications that are assigned values of a more general category type. For example (from Chow and Roukos 1989), the rule

```
(S:mood) —> (NP: person: number)
             (VP: person: number: mood)
```

[5]But see also Jost and Atwell (1993), Wilson and Rayson (1993) for research on semantic tagging of corpora.

enforces agreement between the NP and VP phrases in the values of ': person' and ': number' features ('person' can be first, second or third, and number can be singular or plural). It also requires that S and VP have the same mood. Unification refers to the operation used for building and combining feature structures. In unification based parsing, the interpretations of constraints (in a conjuctive or disjunctive logical connection) are used for variable matching and substitution in order to satisfy agreement that yields a sentential feature structure.
The contribution of semantics is by the semantic role constraints. The parsing process integrates syntactic unification with semantic restriction checks and unification of feature structure succeeds when meeting these restrictions.

Arguments for using unification-based parsing include its declarativeness and better integration of richer linguistic information (syntactic and semantic) to eliminate sentence hypotheses. It also offers global structure synthesis capabilities and flexibility in handling several kinds of argument variations for which other approaches are costly. For example, the processing of *'fly from Denver to Boston'* and *'fly to Boston from Denver'* is better handled by unification-based algorithms than caseframes for which multiple frames are needed. Moreover, they can be designed to handle complex logical constraints involving conjunction, disjunction, implication and negation.

Parsing recognition lattices using a unification grammar has been practised by a number of references in the literature (Tomabechi and Tomita 1988, Chow and Roukos 1989, Bobrow et al 1991, Chien et al 1991, Andry et al 1992). Unification based-algorithms have also been used by Kasper and Hovy (1990) (they combine unification based parsing with classification-based knowledge representation) and Moore and Dowding (1991) (who divide the categories in the unification grammar into context-dependent and context-independent ones in order to deal with gaps in the utterance).

Despite its increasing popularity in computational linguistics, unification-based processing is a matter of controversial discussion. For speech, its main disadvantages are the complexity in designing and maintaining the grammar and its poor computational efficiency. Complexity in the exact specification of features affects the expansion of the grammar into larger domains. Thus, unification algorithms work fairly well for small grammars but are unsatisfactory for larger grammars. For large-scale implementations the designer should find the optimum level of analysis into feature structures and balance them against computational efficiency. Inherent deficiencies of unification-based parsing may affect time performances (see Kasper and Hovy 1990 and Ingria 1990 for discussions). These are associated with making new copies of feature structures in order to guarantee correct unification whenever a description of a sentence is to be built. Since there is no way for a unification to use the results of prior unifications (because the results of computations are not saved), sub-expression computations have to be repeated very often. Furthermore, unification is unable to determine whether any dependencies between structures occur without unifying them. This results in compatibility checks between structures that have no features in common.

7 Connectionist approaches

Artificial neural network modelling is seen as having a great potential in speech recognition (Lippmann 1989), since it exhibits properties like parallelism in processing and learning capability that resemble human-like characteristics. In neural networks, processing elements or nodes are connected by links with variable weights. Connection weights between nodes are adapted from training data and are continuously modified during use. Initially the system knows nothing about the associations between the words and their syntax and semantics. After a pattern and a true label for that pattern are input to the system, the classifier produces an error signal which indicates the distortion measure between the input and the true pattern. This error plus the true label are fed back to modify the system's internal parameters. By this way the system learns by receiving feedback as a response to its action. As words are fed to the network, activation patterns across the feature units which represent the syntactic and semantic properties of the words are produced. The output is typically the distributed representation of the sentence's composition of syntactic and semantic features.

The way semantics is utilized is the mapping from words to their semantic information. This mapping is reflected in the connections between words units and feature units. These constructions are governed by a semantic error signal to control the feedback learning process.

Neural nets have been used both for language acquisition modelling for speech understanding in dialogue tasks (Gorin et al 1990, 1991, Wang and Waibel 1991) and for connectionist parsing to confirm or verify sentence hypotheses (Jain and Waibel 1990). Apart from the intrinsic parallelism capability, connectionism offers several advantages over more conventional approaches. It can combine symbolic and non-symbolic information effectively and can generalize from examples. This is more attractive than constructing complex formal grammars for spoken language domains. Moreover, neural networks, by acquiring the parsing behaviour by themselves during training, tend to be more tolerant to noisy speech input and more efficient in processing loose structures.

The objection against the use of neural networks for speech recognition, is the need for training procedures. Till now, connectionist approaches have been tested on small constrained tasks. Extending them for larger domains results in many thousands of nodes and millions of connections making the networks impractical to train both in terms of computability[6] and learnability. Futhermore, standard evaluation measures on the accuracy of learnt information (like coverage and perplexity) cannot be used to assess the generalization capabilities. Many of the drawbacks in of statistical approaches also apply on neural net approaches. There are no large domain independent training sets, i.e. semantically tagged corpora; and if there were, the computation problem would get even worse. Atwell (1993) compares a Markov n-gram parser trained on the LOB corpus and an analogous back propagation neural parser; he shows that the connectionist approach is much slower and/or less accurate.

8 Conclusion

We have seen the main approaches of applying semantic constraints for spech recognition. To conclude, consider the argument that there is no adequate theory for semantic constraints for speech recognition. Current trends in speech processing move to continuous recognition, large vocabularies and speaker independence. Yet, no system at present has demonstrated large vocabulary continuous speech to text transcription. As vocabularies increase, the confusability and ambiguity between words also increase and there is a clear need for efficient use of all kinds of available knowledge to control search of the resulting huge spaces. Semantics' contribution could improve the performance of such systems considerably. Each of the

[6]Although weights can be automatically trained the number of units and connections have to be predetermined by experiments.

approaches presented above has its own advantages and disadvantages, but nearly all are developed for small vocabularies (up to 1000 words) and specific task domains.

To deal with large vocabulary recognition three problems should be considered:

(i) How to acquire semantic knowledge

Obviously, hand-coding solutions developed for small vocabularies and exploited in the form of semantic grammars, unification-based grammars, case-frames or semantic nets are not viable for vocabularies of 10-20000 words or more. Similarly, extracting statistical semantic measures from text corpora (when this is possible) has proved to have coverage problems. In best cases the training vocabulary cannot exceed 5-7000 words and taking into account that such a size represents approximately 95% of the actual vocabulary used in a real situation, then we assume 5% error rate for acoustic word recognition. But, from the point of view of semantics and speech understanding, the problem gets bigger. The 'missing' words will probably be content words (and most probably 'rare' words used for giving emphasis, strange jargon words, technical words, etc.) so that the lack of their 'strong' semantic content will cause inconsistencies in the semantic language model. For practical reasons, a connectionist approach is not efficient for large vocabulary semantic language acquisition either, due to computational inefficiencies in training a neural net for large vocabularies.

(ii) How to specify, represent and express the semantic constraints.

A task domain may be characterized by a variety of different types of semantic relationships. In general, these specify the requirements for membership in conceptual categories and for participation in meaningful domain relationships (Hayes-Roth 1980). Most systems to date have developed their semantic modules for specific task domains; for example, by assuming that a semantic grammar could adequately represent semantic constraints between lexical items of a command control language (i.e. 'Open spreadsheet, expand it to fill the screen'). Obviously, designing such a grammar must take into account the close, tight relationships for the particular task between 'open, 'spreadsheet', etc., to reduce the number of possible alternatives and constrain search. The question is, could the same grammar be used in another application in which conceptual relationships of words can be of different nature, possibly not realized within the syntactic classes of the words, as for example in machine dictation (in which the list of candidates following 'open' would probably be longer, specifying more abstract and general constraints)? Is such an approach adequate for spontaneous speech applications possibly full of ungrammaticalities?

(iii) How to express computationally the semantic constraints

The main requirement, as far as recognition is concerned, is real time processing of speech input. There is no way at present to make efficient use of semantic constraints for top-down control of search. Presumably the high number of possible candidates at each point of the utterance rules out the possibility of making predictions and proposing hypotheses to control search in large vocabulary recognition. Simply relying on bottom-up recognition and applying semantic restrictions to verify acoustic hypotheses may not prove optimally efficient. The extension of semantic constraints beyond sentence boundaries and the use of discourse information may sometimes be necessary. In addition, no system to date simulates human integrative behaviour of semantics to assist recognition. Human recognition is assumed to be highly distributed in nature. A technique that could constrain acoustic hypotheses in real time, by using semantic information as early as possible, would be the ultimate solution for optimal efficiency.

REFERENCES

For such an extensive survey, errors are, perhaps, unavoidable. Any errors in the findings presented above are to the responsibility of the authors. We would like to apologise in advance for any omissions in important work related to the subject and we invite readers to notify us of any such ommissions.

1. F. Andry, N. Fraser, S. McGlashan, S. Thornton and Nick Youd (1992), Making DATR Work for Speech: Lexicon Compilation in SUNDIAL. In: Computational Linguistics vol. 18, no. 3, 245-267.
2. E. S. Atwell (1983), Constituent Likelihood Grammar. ICAME Journal 7, 34-67.
3. E. S. Atwell (1993), Corpus-based statistical modelling of English grammar. In: Corpus-Based Computational Linguistics, C. Souter and E. Atwell (eds), Rodopi Press, 195-214.
4. J.K. Baker (1976), Stochastic Models for Speech Understanding. In: Speech recognition: Invited papers of the IEEE Symposium, D.R. Reddy (ed), New York: Academic Press, 521-542.
5. D. Bigorne, A. Cozannet, M. Guyomard, G. Mercier and L. Miclet (1988), A versatile speaker-dependent continuous speech understanding system. In: IEEE Proceedings of ICASSP-88, 303-306.
6. R. Bobrow, R. Ingria and D. Stallard (1991), The Mapping Unit Approach to Subcategorisation. In: Proceedings of DARPA Speech and Natural Language Workshop, 185-189.
7. A. Brietzmann and U. Ehrlich (1986), The Role of Semantic Processing in an Automatic Speech Understanding System. In: Proceedings of COLING-86, 596-598.
8. Lee-Feng Chien, K. J. Chen and Lin-Shan Lee (1991), A Preference-first Language Processor Integrating the Unification Grammar and Markov Language Model for Speech Recognition Applications. In: Proceedings of ACL-91, 293-298.
9. M. De Mattia and E.P. Giachin (1989), Experimental results on Large Vocabulary Continuous Speech Understanding. In: Proceedings of the ICASSP-89, 691-694.
10. G. Demetriou and E. Atwell (1994), Semantics in Speech Recognition: A Tutorial Survey. Submitted to Artificial Intelligence Review.
11. G. Demetriou and E. Atwell (1994a), A Large Vocabulary Semantic Network for Computerised Speech Recognition. In: AISB Workshop on Computational Linguistics for Speech and Handwriting Recognition (this proceedings).
12. U. Ehrlich and H. Niemann (1988), Using semantic and pragmatic knowledge for the interpretation of syntactic constituents. In: Recent Advances in Speech Understanding, H. Niemann, M. Lang and G. Sa-gerer (eds), NATO ASI Series, vol. F46, Springer Verlag, 485-490.
13. C. Fillmore (1968), The case for case. In: Universals in Linguistic Theory, E. Bach and R. Harms (eds), New York: Holt, Rinehart and Winston, 1-88.
14. L. Fissore, E. Giachin, P. Laface, G. Micca, R. Pieraccini and C. Rullent (1988), Experimental results on large-vocabulary continuous speech recognition and understanding. In: IEEE Proceedings of ICASSP-88, 414-417.
15. L. Fissore, P. Laface, G. Micca and R. Pieraccini (1989), A Word Hypothe-sizer for a Large Vocabulary Continuous Speech Understanding System. In: Proceedings of the International Conference of ASSP, 453-456.
16. W. A. Gale and K. W. Church (1990), Poor Estimates of Context are Worse than None. In: Proceedings of DARPA Speech and Natural Language Workshop, 283-287.
17. A. L. Gorin, S. E. Levinson, A. N. Gertner, A. Ljolje and E. R. Goldman (1990), On Adaptive Acquisition of Language. In: IEEE Proceedings of ICASSP-90, 601-604.
18. A. L. Gorin, S. E. Levinson and A. N. Gertner (1991), Adaptive Acquisition of Spoken Language. In: IEEE Proceedings of ICASSP--91, 805-808.
19. N. Hataoka, A. Amano, T. Aritsuka and A. Ichikawa (1990), Large vocabulary speech recognition using neural-fuzzy and concept networks. In: IEEE Proceedings of ICASSP-90, 513-516.
20. A. G. Hauptman, S. R. Young and W. Ward (1988), Using Dialog-level Knowledge Sources to Improve Speech Recognition. In: Proceedings of AAAI-88.
21. F. Hayes-Roth (1980), Syntax, Semantics, and Pragmatics in Speech Understanding Systems. In: Trends in Speech Recognition, W. A. Lea (ed), 206-233.
22. P. J. Hayes, Alexander g. Hauptmann, J. G. Carbonell and M. Tomita (1986), Parsing Spoken Language: a Semantic Caseframe Approach. In: COLING-86, 588-592.
23. R. J. P. Ingria (1990). The Limits of Unification. In: Proceedings of DARPA Speech and Natural Language Workshop.
24. E. Jackson, D. Appelt, J. Bear, R. Moore and A. Podlozny (1991), A Template Matcher for Robust NL Interpretation. In: Proceedings of DARPA Speech and Natural Language Workshop (Feb 1991).
25. A. N. Jain and A. H. Waibel (1990), Robust Connectionist Parsing of Spoken Language. In: IEEE Proceedings of ICASSP-90, 593-596.
26. F. Jelinek (1990), Self-Organised Language Modeling for Speech Recognition. In: Readings in Speech Recognition, A. Waibel and Kai-Fu Lee (eds).
27. U. Jost and E. Atwell (1993), Deriving a probabilistic grammar of semantic markers from unrestricted English text. In Grammatical Inference: theory, applications, and alternatives: Colloquim Digest, Institution of Electrical Engineers, Essex, 9/1-9/7.
28. R. T. Kasper and E. H. Hovy (1990), Performing Integrated Syntactic and Semantic parsing Using Classification. In Proceedings of DARPA Speech and Natural Language Workshop, 54-59.
29. D. Klatt (1977), Review of the ARPA Speech Understanding Project. In: Journal of the Acoustical Society of America.
30. F. Kubala, S. Austin, C. Barry, J. Makhoul, P. Placeway and R. Schwartz (1991), BYBLOS Speech Recognition Benchmark Results. In: Proceedings of DARPA Speech and Natural Language Workshop, 77-82.
31. W. A. Lea (1980a) Speech Recognition: Past, Present, and Future. In: Trends in Speech Recognition, W. A. Lea (ed), 39-98.
32. S. E. Levinson and K. L. Shipley (1980), A Conversational-Mode Airline Information and Reservation System Using Speech Input and Output. In: The Bell System Technical Journal, vol 59, No. 1. pp. 119-137.
33. R. P. Lippmann (1989), Review of Neural Nets for Speech Recognition. In: Neural Computation 1, 1-38.
34. D. Luzzati (1987), ALORS: a skimming parser for spontaneous speech processing. In: Computer Speech and Language 22, 159-177.
35. S. Matsunaga, S. Sagayama, S. Homma and S. Furui (1990), A continuous speech recognition system based on a two-lvel grammar approach. In: IEEE Proceedings of ICASSP-90, 589-592.
36. D. Mergel and A. Paeseler (1987), Construction of Language Models for Spoken Database Queries. In: IEEE Proceedings of ICASSP-87, 844-847.
37. R. Moore and J. Dowding (1991), Efficient Bottom-Up Parsing.

In: Proceedings of DARPA Speech and Natural Language Workshop (Feb 1991), 200-203.
38. B. Nash-Webber (1975), Semantic Support for a Speech Understanding System. In: IEEE Transactions on ASSP, vol 23, no. 1, 124-129.
39. G. T. Niedermair (1986), Divided and Valency-oriented Parsing in Speech Understanding. In: Proceedings of COLING-86, 593-595.
40. G. T. Niedermair (1988), Merging Acoustics and Linguistics in Speech-Understanding. In: Recent Advances in Speech Understanding, H. Niemann, M. Lang and G. Sagerer (eds), NATO ASI Series, vol. F46, Springer Verlag, 479-484.
41. G. T. Niedermair, M. Streit and H. Tropf (1990), Linguistic Processing Related to Speech Understanding in SPICOS II. In: Speech Communication 9, 565-585.
42. L. Norton, M. Linebarger, D. Dahl and N. Nguyen (1991), Augmented Role Filling Capabilities for Semantic Interpretattion of Spoken Language. In: Proceedings of DARPA Speech and Natural Language Workshop, 125-133.
43. A. Paeseler and H. Ney (1989), Continuous-Speech Recognition Using a Stochastic Language Model. In: IEEE Proceedings of ICASSP-89, 719-722.
44. R. Pieraccini and C. H. Lee (1991), Factorization of Language Constraints in Speech Recognition. In: Proceedings of ACL-91, 299-306.
45. R. Pieraccini and E. Levin (1992), Stochastic representation for semantic structure for speech understanding. In: Speech Communication 11:283-288.
46. M. Poesio and C. Rullent (1987), Modified Caseframe Parsing for Speech Understanding Systems. In: Proceedings of the 10th International Joint Conference on Artificial Intelligence, vol.2, 622-625.
47. D. R. Reddy (1976), Speech recognition by machine: a review. In: Speech recognition: Invited papers of the IEEE Symposium, D.R. Reddy(ed), New York: Academic Press.
48. T. G. Rose and L.J. Evett (1992), A Large Vocabulary Semantic Analyzer for Handwriting Recognition. In AISBQ-80, 34-39.
49. T. G. Rose (1993), Large Vocabulary Analysis for Text Recognition, PhD Thesis, Nottingham Trent University.
50. A. I. Rudnicky, J-M. Lunati and A. M. Franz (1991), Spoken language recognition in an office management domain. In: IEEE Proceedings of ICASSP-91, 829-836.
51. G. Sagerer and F. Kummert (1988), Knowledge-Based Systems for Speech Understanding.In: Recent Advances in Speech Understanding, H. Niemann, M. Lang and G. Sagerer (eds), NATO ASI Series, vol. F46, Springer Verlag, 421-459.
52. S. Seneff, L. Hirschman and V. W. Zue (1991), Interactive Problem Solving and Dialogue in the ATIS Domain. In: Proceedings of DARPA Speech and Natural Language Workshop, 354-359.
53. S. M. Shieber (1986), An Introduction to Unification-Based Approaches to Grammar. CSLI, Stanford, CA.
54. M. Shigenaga, Y. Sekigughi, T. Yagisawa and K. Kato (1986), A Speech Recognition System for Continuously Japanese Sentences - SPEECH YAMANASHI. In: Transactions of the IECE of Japan, vol. E69, no.5, 675-683.
55. R. M. Stern, W. H. Ward, A. G. Hauptmann and Juan Leon (1987), Sentence parsing with Weak Grammatical Constraints. In: IEEE Proceedings of ICASSP-87, 380-383.
56. G. Thurmair (1988), Semantic Processing in Speech Understanding. In: Recent Advances in Speech Understanding, H. Niemann, M. Lang and G. Sagerer (eds), NATO ASI Series, vol. F46, Springer Verlag, 397-419.
57. H. Tomabechi and M. Tomita (1988), The Integration of Unification-based Syntax/Semantics and Memory-based Pragmatics for Real-Time Understanding of Noisy Continuous Speech Input. In AAAI-88, vol 2, pp. 724-728.
58. Ye-Yi Wang and A. Waibel (1991), A Connectionist Model for Dialog Processing. In: IEEE Proceedings of ICASSP-91, 785-788.
59. W. H. Ward, A. G. Hauptmann, R. M. Stern and T. Chanak (1988), Parsing Spoken Phrases despite Missing Words. In: IEEE Proceedings of ICASSP-88, 275-278.
60. W. Ward (1991a), Understanding Spontaneous Speech: The Phoenix System. In: ICASPP-91, 365-367.
61. A. Wilson and P. Rayson (1993), Automatic content analysis of spoken discourse. In: Corpus-Based Computational Linguistics, C. Souter and E. Atwell (eds), Rodopi Press, 215-222.
62. J. J. Wolf and W. A. Woods (1980), The HWIM Speech Understanding System. In: Trends in Speech Recognition, W. A. Lea (ed), 316-339.
63. S. J. Young, N. H. Russell and J. H. S. Thornton (1988), Speech Recognition in VODIS II. In: IEEE Proceedings of ICASSP-88, 441-444.
64. S. Young (1991), Using Semantics to Correct Parser Output for ATIS Utterances. In: Proceedings of DARPA Speech and Natural Language Workshop, 106-111.