ISCA Archive
http://www.isca-speech.org/archive

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

LEARNING HOW TO UNDERSTAND LANGUAGE

Roberto Pieraccini, Esther Levin

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill NJ 07974 USA

Enrique Vidal

Departamento de Sistemas Informáticos y Computación
Universidad Politeécnica de Valencia
Camino de vera s/n 46071 Valencia, Spain

Abstract

In this paper we discuss learning paradigms for the problem of understanding spoken language. The basic idea consists in redefining the language understanding problem in terms of translation between a natural language and a formal language that represents the meaning of sentences. Within this framework, with the assumption that input and output sentences can be put into sequential correspondence, understanding can be seen as a problem of sequential transduction. In this case several techniques exist for learning the corresponding trasnducers, some of which can be properly stated in terms of Hidden Markov modeling (conceptual HMMs). If the sequential assumption does not hold, there are new algorithms that also seem able to solve the learning problem. This view of a language understanding system opens new perspectives in the field of automatic learning of language models.

10.21437/Eurospeech.1993-13

1 Introduction - The problem of understanding language

Generally the label *understanding system* is granted to those pieces of software that prove some abilities to successfully interact with a *trained* human being using a restricted subset of natural language. The most populated category of those systems, also called *natural language interfaces*, consists of machines that are designed to retrieve information when provided with a question in natural language. How to build a natural language interface is quite well understood [5] assuming we have enough knowledge about the characteristics of the specific subset of the language used in the application. Those characteristics are specified through grammars, built in general by hand, that strongly depend on the application task. Unfortunately for any natural language we may consider there is not such a thing like a general grammar that covers all the possible applications. It is quite clear that even though we had a general grammar of a language, it needs a sort of adaptation for dealing with a particular task. The problem is even more serious when we deal with spontaneous speech rather than with written language since spontaneous speech is often ungrammatical and idiomatic. Additionally, in spoken language there are phenomena like false starts and broken sentences that do not appear in written language. The idea of building a machine that is able of learning how to understand is a rather appealing one, but cannot be implemented without a substantial training corpus of proper examples of sentences and dialogues. The DARPA ATIS corpus [8] that was designed for the development of speech understanding systems is a good set of data for developing and testing some of the learning theories, and although it does not contain any explicit representation of the meaning of sentences, it contains other useful related informations that can be exploited within the framework of a learning strategy.

2 A speech understanding corpus

A relatively large corpus expressively designed for speech understanding (but not for learning) is being developed within the DARPA ATIS project [8]. ATIS stands for Air Travel Information System and the task is built around a subset of the OAG (Official Airline Guide) database. A corpus of spontaneous sentences is being collected and annotated by different sites [14]. The corpus is collected through a Wizard of Oz paradigm. Each subject is given a scenario and a travel planning problem to solve. The subjects are requested to solve the problem by interacting with a machine (that is actually a human *wizard*). The partial and the final responses of the machine are presented to the subjects via a display or a speech synthesizer. The sentences uttered by the subjects are recorded, transcribed and annotated carefully. Although the ATIS corpus may not be the best corpus for testing a semantic learning paradigm, it is readily available and it includes some kinds of annotation that can be indirectly used for our purpose.
Apart from the ATIS corpus many other efforts are currently being devoted to the acquisition of semantically motivated database [12, 25].

2.1 The problem of performance assessment

Assessing the performance of a language understanding system is still an open problem mainly because the concept of *correct answer* is generally ambiguous and must be based on defined conventions that are not task independent. The DARPA community agreed upon scoring answers, within the ATIS task, by comparison with given reference answers that are produced for each valid sentence of the corpus. Of course the problem of the definition of a *correct* answer still remains. For instance, for a question like

SHOW THE LATE EVENING FLIGHTS BETWEEN BOSTON AND DALLAS

the correctness of the answer depends upon the conventional definition of *late evening*. Then, once a time interval has been defined for late evening, it is still not clear what is the information to be listed. It could be the airline and flight number of each flight, but it could also include the departure time, the arrival time, the fare, and so on. Thus in any kind of evaluation of a language understanding system that involves the comparison with reference *correct* answers a set of interpretation rules must be defined. A special committee within the DARPA community agreed upon a certain number of rules, called *principles of interpretation* [14], that should rule the majority of cases for the ATIS task.
Another way for assessing the accuracy of a language understanding system consists in making the comparison at the representation level rather than at the answer level. This involves storing, for each test sentence, the *correct* meaning representation and comparing it, element by element, with the one obtained by the system that is being evaluated. This would allow for a more accurate evaluation than the answer-based assessment, for instance in terms of substitution deletion and insertion rates of units of meaning. However there are at least two problems connected with this kind of evaluation. First, a *standard* meaning representation must be agreed upon in the community, and this seems not to be an easy task. Second, although a set of interpretation rules may not be needed since the concepts can be referred to by a conventional symbol[1], all the symbols for a given task must be specified before the reference representations are generated.

3 Understanding as a translation process

A natural language understanding system is a machine that produces an action as the result of an input sentence (speech or text). Recently, several researchers proposed to look at the understanding process as to a translation (or transduction) process (see Fig. 1) composed of two functional blocks. The first, called *semantic translator*, analyzes the input sentence in natural language *(N-L)* and generates a representation of its meaning in a formal semantic language *(S-L)*. The *action transducer* deterministically converts the meaning representation into one or more statements of a given computer language *(C-L)* (e.g. an SQL statement) for executing the required action.

[1] e.g. the time interval for *late evening* does not have to be defined, but it could be represented by the symbol late_evening

The input natural language is given, and very little can be done for adapting it to our system except imposing restrictions on the lexicon and on the grammar. Instead we have the choice of designing *(S-L)* in order to make the implementation of the semantic translator and the action transducer a feasible task. However the boundary between the first and the second module is quite arbitrary. In [20], for instance, an automatic system was designed for translating English sentences into a representation that is very close to *(C-L)*, and in [9] a paradigm is described for translating natural language into actions without any intermediate representation of the meaning. It is clear that the closer we move the definition of *(S-L)* to *(N-L)*, the more complicate becomes the design of the *action transducer*, reaching in the limit the complexity of a complete understanding system. Conversely, when we move the definition of *(S-L)* closer to *C-L*, we may find that learning the parameters of the *semantic translator* becomes quite a difficult problem when the application entails a rather complex semantics.

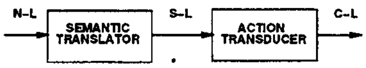

Figure 1: Understanding as a translation process

(S-L) must be powerful enough to allow the representation of the most complex sentences, for instance those including relative and embedded clauses. A semantic language with these characteristics must have, at least, the representational power of a context free grammar, thus allowing to represent the meaning of a sentence for instance with frames [1] or semantic networks [4]. However, the complexity of the algorithms required for dealing with context free grammars, like for instance the *inside-outside* algorithm [3, 15] and the lack of training data suggest the use of a simpler approach. In real applications, like for instance ATIS [8], using a very simple *(S-L)* seems not to create severe limitations on the language. Most of the sentences collected in a spontaneous fashion in information retrieval applications have a very simple structure. They are made up of sequences of phrases, each one of them expressing a distinct concept, with generally no relative clauses. There are a few prototypical learning systems [10, 13, 11] that are based on this observation. Those systems make the assumption that the meaning of a sentence can be expressed by a *sequence of meaning units*, and this sequence can be put in sequential correspondence with the phrases of the sentence. Of course this is a very strong assumption and there are lots of examples that violate it. However, in limited but realistically defined semantic contexts this assumption holds for the majority of sentences. An example of a simple *sequential* representation is shown in Table 1. In the examples the meaning of a sentence is represented by a sequence of *keyword/value* pairs $m_j = (k_j, v_j)$, where k_j, is a conceptual category (i.e. a *concept* like for instance *origin of a flight, destination, meal*) and v_j is the

SHOW ME THE FLIGHTS TO BOSTON	(question,display) (subject,flight) (destin,BBOS)
HOW MUCH IS THE PRICE OF THE FLIGHT FROM ATLANTA	(question,display) (subject,fare) (destin,MATL)
IS BREAKFAST SERVED ON THE FLIGHT?	(question,yes-no) (subject,breakfast)

Table 1: Example of keyword/pair representations of simple phrases within the ATIS domain.

value with which k_j is instantiated in the actual sentence (e.g. *Boston, San Francisco, breakfast*).

4 The stochastic approach: sequential transduction

The stochastic approach to language understanding (see [10, 17, 16, 18]) is based on the *noisy channel* paradigm that was introduced for formalizing the general speech recognition problem in [2] and that constitutes today the theoretical basis of most of the current working speech recognizers. A version of this paradigm was recently proposed for formalizing the problem of the translation between two natural languages [6]. The same paradigm can be used for the formalization of the speech/text understanding problem [10]. The first assumption we make is that the meaning of a sentence can be expressed by a sequence of basic units $\mathbf{M} = m_1, m_2, \ldots, m_{N_M}$ with $m_i \in \Gamma = \{\gamma_1, \gamma_2, \ldots \gamma_N\}$ (i.e. the dictionary of meaning units or *concepts*) and that there is a *sequential correspondence* between each m_j and a subsequence of the acoustic observation $\mathbf{A} = a_1, a_2 \ldots a_{N_A}$, so that we could actually segment the acoustic signal into consecutive portions, each one of them corresponding to a phrase that expresses a particular m_i. The second assumption consists in thinking of the acoustic representation of an utterance as a version of the original sequence of meaning units corrupted by a noisy channel whose characteristics are generally unknown. Thus, the problem of understanding a sentence can be expressed in this terms: given that we observed a sequence of acoustic measurements $\mathbf{A}$ we want to find which semantic message $\mathbf{M}$ most likely produced it, namely the one for which the a posteriori probability $P(\mathbf{M} \mid \mathbf{A})$ is maximum. Hence the problem of understanding a sentence is reduced to that of maximum a posteriori probability decoding (MAP). The most likely words and meaning units can be obtained from the maximization of the product of three factors, namely:

$$\max_{\mathbf{W},\mathbf{M}} P(\mathbf{A} \mid \mathbf{W}) P(\mathbf{W} \mid \mathbf{M}) P(\mathbf{M}) \tag{1}$$

The first factor, the acoustic model $P(\mathbf{A} \mid \mathbf{W})$, is the probability of a sequence of acoustic observations given a sequence of words. Models for the maximization of this probability are well understood, and are generally implemented in the form of acoustic Hidden Markov Models [2, 21]. The second term, the *syntactic model* $P(\mathbf{W} \mid \mathbf{M})$, is the probability of a sequence of words $\mathbf{W} = w_1, \ldots, w_{N_W}$ given a sequence of meaning units $\mathbf{M}$. Finally, the semantic term $P(\mathbf{M})$ expresses the probability of a sequence of meaning units. The syntactic and the semantic terms can be combined in a single model that, with some additional regularity assumptions, takes the form of a non-recursive hierarchical stochastic Finite State Network (FSN). The higher level states of the FSN correspond to conceptual labels and each one of these states contains a lower level stochastic FSN representing an appropriate *state dependent* language model for the sub-sequences of words that convey the meaning associated to the conceptual label of this state. Then, given a sequence of words, the sequence of conceptual labels can be easily recovered by means of Viterbi decoding.

Several schemes can be adapted for the representation of the different models involved under this general framework. For instance in [11, 22] automatically learned stochastic regular grammars of different classes are successfully used in tasks such as "understanding" Spanish spoken numbers and semantically decoding spontaneous spoken queries in Spanish. Similarly a better understood and perhaps more convenient scheme is one in which *n-grams* are used both for the state dependent word language modes and for the higher level model of conceptual unit concatenation [10]. In this case, assuming $n = 2$ for the conceptual unit concatenation model, the whole FSN takes the form of a Hidden Markov Model (*conceptual* HMM) whose states correspond to conceptual labels and whose observations are sequences of words modeled, for each state, with an *n-gram* language model. Formally the conceptual model is defined by a set of states $\Gamma = \{\gamma_1, \gamma_2, \ldots \gamma_N\}$, a set of *concept conditional n-grams* $P(w_i \mid w_{i-n+1}, \ldots, w_{i-1}, m_i = \gamma_j)$ and the *concept transition probabilities* $P(m_i = \gamma_j \mid m_{i-1} = \gamma_k)$. The performance of a system based on this paradigm tested on the ATIS task range around 68% of correct answers from text input.

4.1 Training the conceptual model

The concept conditional n-grams and the concept transition probabilities can be estimated from a corpus of training examples. Each training example consists of a sentence $\mathbf{W}$ and the sequence $\mathbf{M}$ of associated conceptual labels. Unfortunately, the conceptual labels must be provided manually for each sentence along with the corresponding semantic segmentation of the sentence $\mathbf{W}$. The cost of such handlabeling of all the sentences in a big corpus can be comparable and even greater than that of writing a grammar for the application language. Thus, it becomes important to develop strategies for reducing the cost of conceptual annotation of sentences in a big corpus. The annotation cost can be reduced either devising a strategy that takes advantage of all the possible semantically related information that is already in the corpus, or by making the annotation very simple and performable by non specialized people.

For instance, in the ATIS corpus, the meaning of sentences is not available in a declarative form. Instead, each sentence is associated with the *action* resulting from the *interpretation* of the meaning, namely the correct answer. One way of using this information for avoiding the handlabeling and segmentation of all the sentences in the corpus consists in creating a training loop in which the provided correct answer serves the purpose of a feedback signal. In practice, all the available sentences are analyzed by the understanding system obtained with an initial estimate of the conceptual model parameters. The answers are then compared to the reference answers and the sentences are divided into two classes. The *correct* sentences, for which we assume that the conceptual segmentation obtained with the current model is correct, and the *problem sentences.* Then the segmentation of the correct sentences is used for reestimating the model parameters, and the procedure is repeated again. The procedure can be repeated until it converges to a stable number of correct answers. Eventually the remaining *problem sentences* are corrected by hand and included in the set of correct sentences for a final iteration of the training algorithm. This procedure can be used effectively for reducing the amount of handlabeling [17] necessary for a big corpus. Of course the training loop cannot solve completely the annotation problems. We noticed that more than 20% of the sentences in a corpus must be manually inspected and conceptually annotated after the training loop. However the annotation operation can be made very simple making the semantic language *(S-L)* flexible and easy to use. The methodology used for annotating the ATIS corpus [14] constitutes a good example of this idea. In fact the annotators rephrase each valid sentence in an artificial language that is a restricted and unambiguous form of English. This *pseudo-English* rephrasing (called *win* or *wizard input*) constitute the input of a parser, called NL-parse [7], that generates the SQL query. For instance, for a sentence like:

> *I'D LIKE TO FIND THE CHEAPEST FLIGHT FROM WASHINGTON D C TO ATLANTA*

The *win* rephrasing is:

> *List cheapest one direction flights from Washington and to Atlanta*

NL-parse then produces the corresponding SQL statement. Both the SQL query and the *win* sentence can be considered semantic representations of the original sentence. In fact the SQL query is the final target of the understanding system and can be unambiguously obtained from the *win* sentence. If the sequential correspondence assumption holds between the pseudo-English *win* sentence and the original message, like in the previous example, the pseudo-English language can be thought of as an alternate candidate for *(S-L).* Using *win* for representing the meaning may lead to two different solutions. In the first we can think of developing a system that learns how to translate natural language sentences into pseudo-English sentences and then use the existing parser for generating the SQL queries. In the second solution each *win* sentence in the corpus can be translated in the corresponding keyword/value representation. This translation is unambiguous (*win* is an unambiguous artificial language by definition). A parser can be easily designed for performing the translation or the conceptual model itself can be trained for that.

5 The sequential correspondence assumption

In section 4 we based our formalization of the speech understanding problem on the assumption that there is a sequential correspondence between the representation of a sentence (words or acoustic measurements) and the corresponding representation of meaning. Unfortunately, when using a simple annotation procedure like the *win* language, the sequential correspondence assumption will not hold for a good percentage of the sentences. A typical example is constituted by the following sentence:

> *COULD YOU PLEASE GIVE ME INFORMATION CONCERNING AMERICAN AIRLINES A FLIGHT FROM WASHINGTON D C TO PHILADELPHIA THE EARLIEST ONE IN THE MORNING AS POSSIBLE*

whose corresponding *win* annotation is:

> *List earliest morning flights from Washington and to Philadelphia and American.*

The problem of reordering the words of the *win* representation for aligning it with the original sentence is a complex problem that cannot be solved optimally. Suboptimal solutions with satisfactory performance can be developed based on heuristic search. We will not discuss the details of how the reordering can be put into practice. Rather we want to emphasize the fact that an iterative algorithm based on a model similar to the conceptual model of section 4 led to almost 91% correct alignments between English sentences and corresponding *win* representations on a corpus of 2863 sentences.

6 Subsequential transduction

Another different approach is to explicitly accept the non-sequential nature of the output (*win*) language and to try to translate directly the natural language sentences into the given output semantic representation. Once non-sequentiality is acknowledged, we should also recognize that increasing degrees of non-sequentiality exist and that increasingly complex devices are required to solve the corresponding translation problems. In this framework, it is interesting to mention the problem of *subsequential transduction* which can be solved by finite-state *subsequential transducers*. Subsequential transduction allows for a significantly higher degree of input-output unsequentialities than pure sequential transduction does. Moreover, this class of transducers has recently been shown to be learnable from training input-output examples using the so called *Onward Subsequential Transducer Inference Algorithm* (OSTIA) [19].

An example of application of OSTIA to (pseudo)natural language understanding is given in [23]. The task considered in this work consists of understanding the meaning of English sentences that describe simple visual scenes. These sentences are generated by a context-free grammar which is unknown to OSTIA, and each training sentence is accompained by its corresponding semantic transcription in terms of *first-order logic formulae*. Possible objects in a scene are represented by *variables* and *unary* or *binary predicates* are used to represent object attributes (such as shape shade and size) or the relative positions between objects (touch, above, below, right, left, far above, etc.) Some examples of English sentences along with their corresponding transcription are shown in Table 2. It is interesting to notice that although input-output sequentiality is clearly violated in this representation, many of the concepts (objects and attribute predicates) are in fact fairly sequential with the input. The results of using OSTIA in learning this language understanding task were very encouraging. Almost perfect and quite compact transducers (semantic error less than 1% with a FSN of about 60 states) were learned from relatively small training sets (about 10000 input-output examples).

Apart from these experiments with a rather artificial task, an additional experiment was carried out in a more real setting. In this experiment, a set of 1146 *simple* sentences was selected from the ATIS database (see [24] for details). 1000 of these sentences, acompained by their corresponding pseudo-English (*win*) transcription, were supplied to OSTIA for training. The resulting semantic transducer (which had about 40 states) was tested on the remaining 146 English sentences, 87 of which were perfectly translated into the correct *win* commands. Although these results were also judged as rather encouraging, an analysis of the errors yield the conclusion that an exceedingly large increase in training data should have been necessary to adequately learn a subsequential transducer that would effectively account for the kind of non-sequentialities involved in this restricted version of the ATIS task.

In order to better account for this unsequential nature, a new approach has recently been introduced which does not assume or require any kind of input-output sequentiality. The order of input and output tokens is explicitly modeled by conventional stochastic finite state input and output language models, while relations between these tokens are indirectly represented by probabilistic associations between rules and/or non-terminals of these models. This approach, along with its application to ATIS data, is presented in [24] and is omited here for the sake of brevity.

7 Conclusions

A new framework for language understanding based on the concept of translation between a natural and a formal language was recently proposed along with some preliminary attempts to learn this function from examples. This paradigm is represented in Fig. 1; a first module translates a natural language sentence *(N-L)* into a semantic language *(S-L)* that represents the meaning. The natural language characteristics are generally unknown, while the semantic language is specifically designed to cover the semantics of the application. The second step consists in the translation of the sentence in *(S-L)* into computer language code *C-L* for performing the requested action. This second module can be generally (but not necessarily) designed to cover all the possible sentences in *(S-L)*, since both *(S-L)* and (*C-L*) are known.

Although the strategies described in this paper use a very simple intermediate semantic representation, they can successfully handle most of the sentences in a database query application like the ATIS task. When the sequential correspondence assumption holds between *(N-L)* and *(S-L)* the problem of *semantic translation* can be formalized as a transduction problem solved by parsing the input string with a stochastic FSN that can be inferred directly from the training examples. The parsing consists in Viterbi decoding. An interesting case is when the FSN takes the form of a Hidden Markov Model.

However there are cases in which it is not possible to maintain the sequential correspondence between a sentence and its meaning represented with a formal language. In these cases there is an algorithm (OSTIA) that can solve the learning problem in a subsequential transduction framework. Although OSTIA was shown to give encouraging results for learning a subsequential transductor within the ATIS task, increasing the performance requires a huge amount of training data that is not available at the moment. Another algorithm that does not make any assumption on the sequentiality between the *(N-L)* and *(S-L)* gave interesting results in the same task and is described in detail in another paper in these proceedings.

References

[1] Minski, M., "A Framework for Representing Knowledge," in P. Winston (ed.). *The Psychology of Computer Vision,* New York: McGraw-Hill, 211-277, 1975.

[2] Jelinek, F., "Continuous Speech Recognition by Statistical Methods," *Proceedings of IEEE,* vol. 64, no. 4, pp. 532-556, 1976.

a small triangle touches a medium light circle and a large square SM(x) & TR(x) & MD(z) & LI(z) & CI(z) & LR(w) & SQ(w) & TO(x,z) & TO(x,w)	
a small square and a large dark triangle are far above a small dark square SM(x) & SQ(x) & LR(y) & DK(y) & TR(y) & SM(z) & DK(z) & SQ(z) & FAB(x,z) & FAB(y,z)	

Table 2: Examples of English sentences describing simple visual scenes and their corresponding semantic transcription in terms of logical formulae.

[3] Baker, J., "Trainable Grammars for Speech Recognition," in Wolf, J. J., Klatt, D. H., editors, *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America,* MIT, Cambridge, MA, June 1979.

[4] Sowa, J., *Conceptual Structures.* Reading, MA:Addison-Wesley, 1984.

[5] Allen, J, "Natural Language Understanding," The Benjamin/Cummings Publishing Company, Inc., 1987.

[6] Brown, P. F, Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., Mercer, R. L., Roosin, P. S., " A Statistical Approach to Machine Translation," *Computational Linguistics*, Volume 16, Number 2, June 1990.

[7] Hemphill, C. T., Godfrey, J. J., Doddington, G. R., "The ATIS Spoken Language Systems, pilot Corpus," *Proc. of 3rd DARPA Workshop on Speech and Natural Language*, pp. 102-108, Hidden Valley (PA), June 1990.

[8] Price, P. J., "Evaluation of Spoken Language Systems: the ATIS Domain," *Proc. of 3rd DARPA Workshop on Speech and Natural Language*, pp. 91-95, Hidden Valley (PA), June 1990.

[9] Gorin, A. L., Levinson, S. E., Gertner, A., "Adaptive Acquisition of Language," *Computer Speech and Language*, **5**, pp. 101-132, 1991.

[10] Pieraccini, R., Levin, E., Lee, C.-H., "Stochastic Representation of Conceptual Structure in the ATIS task," *Proc. of 4th DARPA Workshop on Speech and Natural Language*, Pacific Grove, (CA), February 1991.

[11] Prieto, N., Vidal, E., "Learning Language Models through the ECGI Method," *Proc. of EUROSPEECH 91*, Genova, Italy, September 1991.

[12] Zue, V., Glass, J., Goodine, D., Leung, H., Phillips, M., Polifroni, J., Seneff, S., "Integration of Speech Recognition and Natural Language Processing in the MIT VOYAGER System," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing,* ICASSP-91, Toronto, Canada, May 1991.

[13] Giachin, E. P, "Automatic Training of Stochastic Finite-State Language Models for Speech Understanding", *Proc. of International Conference on Acoustics, Speech and Signal Processing,* ICASSP-92, San Francisco, CA, March 1992.

[14] *MADCOW*, "Multi-Site Data Collection for a Spoken Language Corpus,", *Proc. of Fifth Darpa Workshop on Speech and Natural Language*, Harriman, NY, Feb 1992.

[15] Pereira, F., Schabes, Y., "Inside-Outside Reestimation from Partially Bracketed Corpora," *Proc. of Fifth Darpa Workshop on Speech and Natural Language*, Harriman, NY, Feb 1992.

[16] Pieraccini, R., Tzoukermann, E., Gorelov, Z., Levin, E., Lee, C.-H, Gauvain, J.-L, "Progress Report on the Chronus System: ATIS Benchmark Results," *Proc. of Fifth Darpa Workshop on Speech and Natural Language,* Harriman, NY, Feb 1992.

[17] Pieraccini, R., Gorelov, Z., Levin. E., Tzoukermann, E., "Automatic Learning in Spoken Language Understanding," *Proc. of 1992 International Conference on Spoken Language Signal Processing,* ICSLP 92, Banff, Alberta, Canada, October 1992.

[18] Tzoukermann, E., Pieraccini, R., Gorelov, Z., "Natural Language Processing in the CHRONUS System," *Proc. of 1992 International Conference on Spoken Language Signal Processing,* ICSLP 92, Banff, Alberta, Canada, October 1992.

[19] Oncina, J., García, P., Vidal, E., "Learning Subsequential Transducers for Pattern Recognition Interpretation Tasks," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 15, No. 5, pp. 448-458, May 1993.

[20] Kuhn, R., De Mori, R., "Learning Speech Semantics with Keyword Classification Trees," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing,* ICASSP-93, Minneapolis, Minnesota, April 27-30, 1993,

[21] Rabiner, L. R., Juang, B. H., "Fundamentals of Speech Recognition," Prentice Hall Signal Processing Series, 1993.

[22] Segarra, E., "Una Approximacion Inductiva a la Comprension del Language," PhD dissertation, Universita de Valencia, Valencia, Spain, June 1993.

[23] Castellanos, A., Vidal, E., Oncina, J., "Language Understanding and Subsequential Transducer Learning," First Int. Colloquium on Grammatical Inference: Theory, Applications and Alternatives, Proc. Univ. of Essex, April 1993.

[24] Vidal, E., Pieraccini, R., Levin, E., "Learning Associations Between Grammars, a New Approach to Natural Language Understanding," Proc. of EUROSPEECH 1993.

[25] Diaz, J. E., Rubio, A. J., Peinado, A. M., Segarra, E., Prieto, N., Casacuberta, F., "Development of Task Oriented Spanish Speech Corpora," Proc. of Eurospeech 93, Berlin, Germany, September 1993.