ON THE LIMITATIONS OF STOCHASTIC CONCEPTUAL FINITE-STATE LANGUAGE MODELS FOR SPEECH UNDERSTANDING

J. Colás, J. Ferreiros, J. M. Montero, J. Pastor, A. Gallardo, J. M. Pardo

Grupo de Tecnología del Habla - Departamento de Ingeniería Electrónica
E.T.S.I. Telecomunicación – Universidad Politécnica de Madrid
Ciudad Universitaria s/n, 28040 Madrid Spain

ABSTRACT

In a limited domain task (e.g. airline reservation, database retrieval, etc) many robust understanding systems, designed for both speech and text input, have been implemented [1][3][5][6][10] based on the ***Stochastic Conceptual Finite-State*** paradigm (*Semantic Network*) or ***CHRONUS*** paradigm [11] (Conceptual Hidden Representation of Natural Unconstrained Speech), which establishes relations between conceptual entities through a probabilistic graph-like structure. The use of this kind of grammar to model semantic information presents limitations, which have been analysed during the implementation of a flexible architecture for a robust information retrieval system, based on the same paradigm [1]. We have tried to solve some of them by integrating a set of conceptual probabilistic and non-probabilistic grammars, which allow certain complexity in the functionality of the application, such as applying non-SQL functions to the results of SQL queries in order to retrieve information not explicitly included in the database, translating certain natural spoken sentences (that would produce difficult embedded queries and therefore more natural queries) without so many restrictions in the relative position of inter-concept relationship.

1. INTRODUCTION

Many robust understanding systems, designed for both speech and text input, have been implemented [1][3][5][6][10][11] based on the ***Stochastic Conceptual Finite-State*** paradigm, which establishes relations between conceptual entities through a probabilistic graph structure. These *concept relations*, or *linguistic cases*, can be used to label the phrases of a sentence and obtain an intermediate representation useful for its interpretation. In a limited domain task (e.g. airline reservation, database retrieval, etc) the number of different concepts can be assumed to be finite and directly deducible from the knowledge of the task itself. The output of this module is called conceptual segmentation and it is the input for a second module, the ***Template Generator***, which translates the initial segmentation into a template conforming to a more abstract formalism. Finally, an ***SQL Translator*** generates SQL query for extracting the requested information from a database. When processing spoken queries, some systems [7] [10] [11] classify them before being translated, using ***Supervised Stochastic Classifiers*** based on neural networks, dynamic programming, etc. These classifiers need a great amount of labelled data and a set of complicated rules to fill the slots of the semantic frames recognised by the classification. This approach is less flexible when adding further functionality is needed. Other approaches use integrated syntactic and semantic grammars to obtain structural information before applying the translation rules [3] [4].

In this paper, we present some conclusions about limitations of these systems, obtained during the implementation of a flexible architecture for a robust information retrieval system based on the same paradigm [1], and our solutions for some of them. Besides, we present an alternative classifier based on both Structure Analyser (**SA**) and Structure Transformer (**ST**) modules. Compared to stochastic classifiers [7][10][11], this architecture is more flexible when adding new functionality without the need of collecting and labelling a large amount of additional data. Nevertheless, the **PFSN** (probabilistic finite state network) must be re-trained and new **CM** (conceptual mapping), **SA** (structure analysis) and **ST** (structure transformation) rules could have to be written.

2. LIMITATIONS OF CONCEPTUAL FINITE-STATE LANGUAGE MODELS

Although the Probabilistic Conceptual Finite-State paradigm has proved to be a good solution to implement robust and trainable speech understanding systems, it shows some limitations derived from itself. To develop a robust speech understanding module using a small labeled-data corpus we need to relax the linguistic restrictions, smoothing the semantic (conceptual) grammar used to decode the speech input into a concept sequence from the application semantic domain. Some systems [1][5][10] use an ergodic conceptual automata (no restrictions between concepts) to allow many concept sequences not trained. A relaxed semantic grammar (ergodic) with not modeled long-term inter-concept relationships (Finite-State Network) and the use of a "garbage" semantic category (to model words not included in the semantically categorized lexicon) produce "**Semantic Ambiguity"** problems (e.g. dates, numbers, ambiguous concepts in the domain, etc) (see *Example 1*).

Besides, in some languages as Spanish, the *phrase order* in the sentence has a degree of freedom, allowing a variety of conceptual sequences to express the same query, in the surface level of sentence. So, in many cases, attribute-value pairs (in a relational database sense) do not appear together, and the value-concept has to be decoded as a genera concept (number, date, etc) and is not linked to the attribute

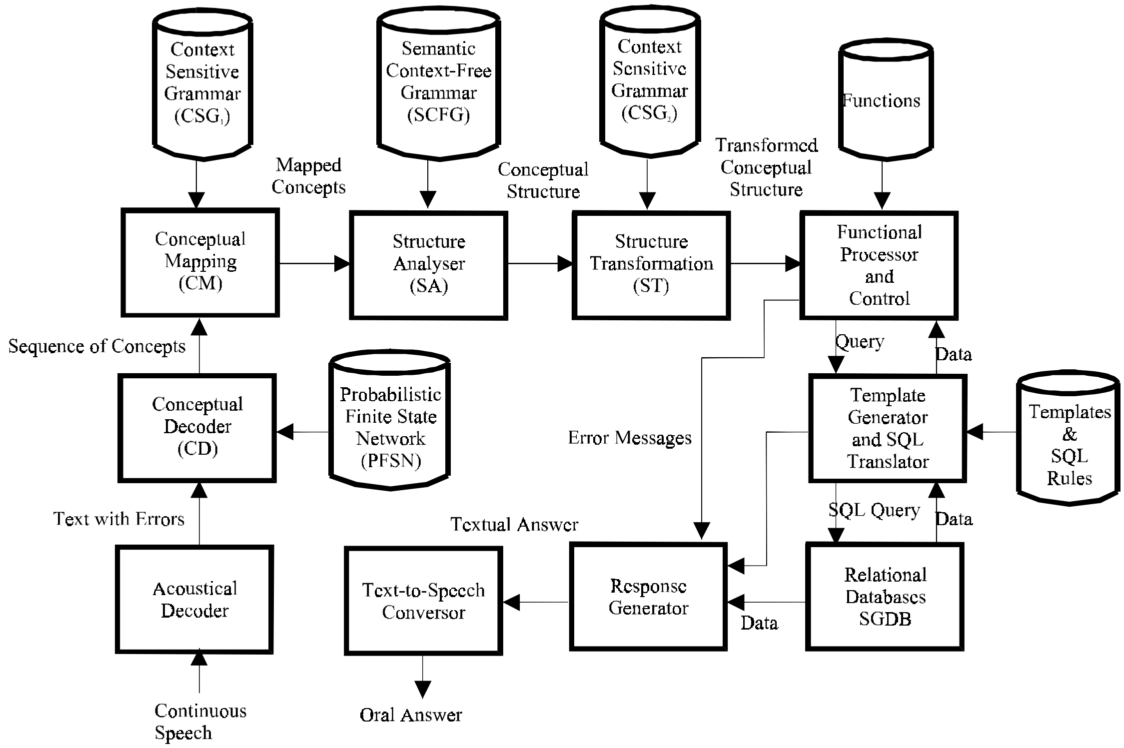


Figure 1. Modules Diagram of the proposed Architecture for the Robust Speech Understanding System

Concept (**"Bad-Formed Concepts"**) (see *Example 2*). This is a problem to understand correctly the sentence because a attribute-value pair must be linked in a single concept before translating it.To solve these problems we include a Conceptual Mapper module which uses context sensitive grammar and a set of semantic features to find inter-concepts relationships and to properly decode some general concept as dates, numbers, and some concepts with the same interior structure. Some previous works [8] have proposed the use of context-free grammars to analyze the interior of each concept but an inter-concept grammar is also needed to model long-term inter-concept relationships [9].

3. THE PROPOSED SPEECH UNDERSTANDING ARCHITECTURE

A non-integrated two level approach has been implemented, including an acoustic decoder and an understanding module. Our proposed architecture makes use of several kinds of grammars: a probabilistic finite state network (**PFSN**) for concept decoding (**CD**), a context sensitive grammar (**CSG$_1$**) for conceptual mapping (**CM**), a semantic context-free grammar (**SCFG**) for structure analysis (**SA**) and another context sensitive grammar (**CSG$_2$**) for structure transformation (**ST**). This combination *allows answering a complex query through the structured execution of multiple simple queries and functions*.

The understanding architecture (**Figure 1**) is composed of the following modules that run in a PC:

3.1. Acoustic Decoder

A One-Pass continuous speech recogniser based on semi-continuous HMM (SCHMM) with a smoothed 160 **POS** bigram is used [2] to decode the acoustic input from the user.

3.2. Understanding Module

It takes an input sentence from the acoustical decoder and produces an answer that, most of the times, is the result of the application of several available non-SQL functions on simple queries. In other cases it is the result of answering a simple query. The following modules compose this understanding system:

Probabilistic Conceptual Decoder

It is based on a dynamic programming algorithm and an ergodic regular grammar of concepts, related to the application domain, where each concept is modelled by a **PFSN** of semantic categories [3] [4] [6]. For the first version we have not trained these probabilities and they are uniform.

Robustness and coverage are increased by means of a special "garbage" category. Stochastic concept automata mainly model the structural information of each concept, but an inter-concept grammar is also trained from labelled data.

Before **CD**, we use a semantically labelled dictionary to look up each word. Due to ambiguity, the application of the labeller on a sentence produces a directed category graph. A dynamic programming algorithm processes this graph to get the most probable sequence of concepts and categories, solving the ambiguity problem.

Conceptual Mapper

As the **CD** is unable to model non-contiguous inter-concept relationships, some specific concepts such as numbers or dates, are not correctly labelled. Considering the context, each of these ambiguous concepts can be correctly mapped or assigned to the right concept using a set of mapping and merging **CSG**$_1$ rules (for instance, a date can be an attribute of more than one entity in the Entity-Relationship scheme). As we can see in the example below, due to the presence of a Report concept, Date can be mapped onto Date-Report (*Example 1*)**:**

Was last casualty report regarding Zeus in November ?

Report **Ship** **Date** = **Report Date**

Some contiguous and non-contiguous concepts can be merged. In the example below, considering the Length concept in the sentence, Number concept can be incorporated to Length (attribute-value pair) (*Example 2*):

Is the length of the fastest ship 5 meters or more ?

Length **Speed** **Ship** **Number** = **Length**

Structure Analyser

Our proposed alternative to stochastic classifiers is based on a **SCFG** concept parser. The structure obtained this way allows determining the simple queries and the functions (distance, comparison, etc.) that process them. In order to get a reduced and more general set of grammar rules, a concept taxonomy (verb, function, entity, attribute) is used. To design these rules, the application functionality and certain general linguistic phenomena (co-ordination, relative clauses, negation, etc.) has been taken into account.

Structure Transformer

If the **SCFG** extracts conceptual structure, the **ST** rules complete subqueries, solving some problems such as concept ellipsis, relative clauses, and so on.

Functional Processor and Control

The role of this module is to process the result of the queries using, if required by the answering strategy, the non-SQL functions. Additional functions can be easily incorporated if necessary.

Template Generator and SQL Translator

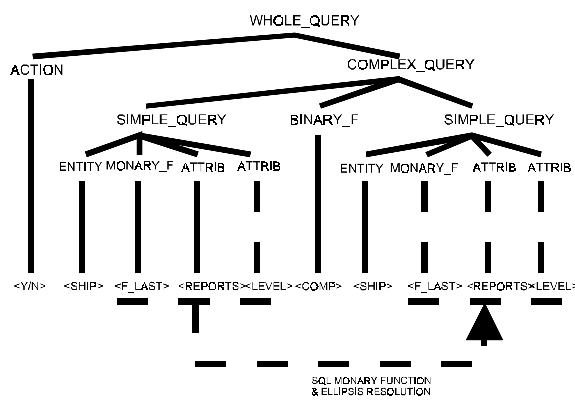
The simple queries detected and completed by previous modules are translated to SQL in two steps: templates are filled with key information from the sentence and, then, these templates are translated into SQL queries by means of a set of rules that include database information.

3.3 Examples

One example of complex query processing could be: “Is England´s latest casualty report rated worse than America´s?” It is decomposed into two simple queries (“England´s latest casualty report” and “America´s latest casualty report”) which are the arguments of the comparative function “worse than”. These simple queries are easily translated into SQL after the ST module solves the concept ellipsis problem in the second query. In this example, we include the following illustration that shows the concept structure and the transformation process. The output of CD and CM is:

<Y/N> : is
<SHIP> : England´s
<F_LAST> : latest
<REPORTS> : casualty report rated
<F_COMP> : worse than
<SHIP> : America´s

where concepts appear between brackets. As can be seen in the illustration below, each concept is assigned to one of the classes in the taxonomy (ACTION, ENTITY, MONARY_F, ATTRIB, BINARY_F, etc).



Another example: “Is Neptune closer to Titanic than to Zeus?” It is decomposed into three simple SQL queries, which obtain the geographic position of each ship. With this information two distances are calculated and compared by the control module.

4. TASK DESCRIPTION AND EVALUATION

An Information Retrieval System has been developed in order to get navy information, allowing not only simple queries but also more complex ones, including non SQL functions and multi-query questions.

The application has a restricted semantic domain. The vocabulary is about 1100 words. The first version of the system has been implemented using 600 Spanish sentences for training (both text and speech) and 400 for testing. 4 speakers (2 male and 2 female) have uttered the speech. To further evaluate the system we have collected up to 3000 sentences (plain text) from new users and several speakers will speak some of them, after the necessary review.

The results on text corpora, manually comparing the system output to the reference queries, show that there are *92 % correctly generated queries for the training text corpus, and 89 % for the testing one.*

We have evaluated the spoken corpora (the same 87 test sentences for 4 speakers, a subset of the 400 sentences) by manually comparing the system output to the reference queries. So we have obtained conclusions about the limitations and advantages of using robust concept decoding (with garbage semantic categories) in restricted domain applications. The text input to the decoder has been obtained using Semicontinuous HMM and a 160 **POS** bigram trained from different text corpora without relation to the semantic domain application. The percentage of correct sentences is shown in Table I.

Speaker	**Correct Sentences**
FER (male)	53%
LEA (male)	69%
ROM (female)	70.5%
ENR (female)	65%
Average	**64.37%**

Table I: Percentage of Correct Sentences for Speech Understanding

5. FUTURE WORK

We are currently collecting 10 new speakers (5 men and 5 female), to evaluate the proposed system and to obtain conclusions about the limitations and advantages of using robust concept decoding (with garbage semantic categories) in restricted domain applications. To increase the performance of the system we will try to integrate structural information into conceptual segmentation of the sentence to improve the Conceptual Mapper. Also, we will increase the lexicon with new words to improve coberture.

6. CONCLUSIONS

We have presented a flexible bottom-up approach to deal with complex understanding tasks and to solve some limitations due to the use of relaxed stochastic conceptual grammars. To attain this, the architecture integrates robust conceptual decoding and semantic structure analysis to answer a complex query through the structured execution of multiple simple queries and functions. The results that we have obtained so far are encouraging. We have studied the advantages and limitations of using the ***Stochastic Conceptual Finite-State*** paradigm, and a partial solution to these limitations has been found (Conceptual Mapper). A more flexible and alternative approach (**SA** and **ST** modules) to ***Stochastic Classifiers*** to decode the complexity of sentences has been implemented and evaluated with encouraged performance.

7. ACKNOWLEDGEMENTS

This work has been supported by CICYT MEC (Ministry of Education and Science) TIC 94-0119 Sicovox II Continuous Speech Recognition Systems. The authors wish to thank in general to all Speech Technology Group members for their help and specially to Samir Kilany and Jorge de Teresa for their work on several modules of the speech understanding architecture.

8. REFERENCES

1. José Colás, Juan M. Montero, Javier Ferreiros, José M. Pardo , “An alternative and flexible approach in robust information retrieval system”, EUROSPEECH’ 97, ISSN 1018-4074, Rhodes, Greece
2. J. Ferreiros, “Contribution to Markov Models training methods for continuous speech recognition”. Ph D. Thesis. Universidad Politécnica de Madrid, 1996.
3. Nagai Akito, Ishikawa Yasushi, Nakajima Kunio, “Integration of Concept-Driven Semantic Interpretation with Speech Recognition”, Proc. ICASSP-96,pp.431,434. Atlanta, USA.
4. S.K. Bennacef, H. Bonneau-Maynard, J.L. Gauvian, L. Lamel, W. Minker, “A Spoken Language System For Information Retrieval”, Proc. ICSLP-94, pp. S22-8.1, 8.4, Yokohama, JAPAN
5. Egidio Giachin, Paolo Baggia, Giorgio Micca, “Language Models for Spontaneous Sppech Recognition: A Bootstrap Method for Learning Phrase Bigrams”.Proc. ICSLP-94, pp. 16-17.1, 16-17.4.
6. Enrique Vidal, Roberto Pieraccini, Esther Levin, “Learning Associations Between Grammars: A New Approach to Natural Language Understanding”, Proc. EUROSPEECH-93, pp.1187-1190.
7. P. Bianchini, P. Ferragina, M. Notturno Granieri, L. Tarricone, “New Techniques for Speech Understanding”. Proc. ICASSP-93, pp. II-127, II-13
8. Ying Cheng, Yves Normandin, Paul Fortier, “Integration of Neural Networks and Robust Parsers in Natural Language Understanding”. Proc. Eurospeech-93, pp.1311-1314.
9. Evelyne Millien and Roland Kuhn,”A Robust Analyzer for Spoken Language Understanding”, Proc. Eurospeech-93, pp.1331-1334.
10. Egidio P. Giachin, “Automatic Training of Stochastic Finite-State Language Models for Speech Understanding”. Proc. ICASSP-92, pp. I-173,I-176.
11. Roberto Pieraccini, Evelyne Tzoukermann, Zakhar Gorelov, Jean-Luc Gauvain, Esther Levin, Chin-Hui Lee, Jay G. Wilpon, “A Speech Understanding System Based on Statistical Representation of Semantics”. Proc. ICASSP-92, pp.I-193, I-196