Actas do Encontro Científico
3º Festival Nacional de Robótica - ROBOTICA2003
Lisboa, 9 de Maio de 2003.

RECENT DEVELOPMENTS ON THE SPOKEN LANGUAGE HUMAN-ROBOT INTERFACE OF THE ROBOT CARL

António Teixeira [**,*,1] **Luís Seabra Lopes** [**,*]
Liliana Ferreira [*] **Pedro Soares** [*] **Mário Rodrigues** [*]

[*] *Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA)*
[**] *Departamento de Electrónica e Telecomunicações, Universidade de Aveiro, 3810 193 Aveiro, Portugal*

Abstract: The development of robots capable of accepting instructions in terms of familiar concepts to the user is still a challenge. For these robots to emerge it's essential the development of natural language interfaces, since this is regarded as the only interface acceptable for a machine which intents to have a high level of interactivity with Man. Our group has been involved for several years in the development of a mobile intelligent robot, named Carl, capable of interacting with humans using spoken language. In this paper we present the recent evolution of several parts of the spoken language interface: speech recognition, natural language understanding and natural language generation. The new speech recognition module and part of the natural language understanding were already integrated in Carl. The other experiments reported must be regarded as exploratory, aiming the development of a robust spoken language interface.

Keywords: intelligent robots, spoken language, speech recognition, natural language generation, natural language understanding, memory-based learning

1. INTRODUCTION

Project CARL - "Communication, Action, Reasoning and Learning in robotics" -, started in 1999 as an FCT project and now being continued as a separate transversal research area at the Aveiro Electronics and Telematics Engineering Institute (IEETA), aims to develop a robot capable of understand, using a friendly interface, instructions expressed in a way familiar to the human user.

Carl is the name of the robot of the CARL project. Carl, shown in Fig. 1, is a prototype of an intelligent service robot, designed having in mind such tasks as serving food in a reception (Seabra Lopes, 2001) or acting as a host in an organization (Seabra Lopes, 2002). The approach that has been followed in the design of Carl is based on an explicit concern with the integration of the major dimensions of intelligence, namely Communication, Action, Reasoning and Learning (CARL). Although different communities have thoroughly studied these dimensions in the past, their integration has seldom been attempted in a systematic way.

Carl is based on a Pioneer 2-DX platform from ActivMedia Robotics, with two drive wheels and a caster. It includes an onboard Pentium based computer running Linux, wheel encoders, front and rear bumpers rings, front and rear sonar rings,

[1] Email for contacts: `ajst@det.ua.pt`.

Fig. 1. Carl photo at its recent participation in the “International Cleaning Robots Contest”, Lausanne, Switzerland, October 2002

an audio I/O card and a Sony EVI D31 pan-tilt-zoom camera.

On top of this mobile platform, we have added a fiberglass structure that makes Carl approximately 1.10 cm high (see Fig. 1). This fiber structure carries a Voice Tracker directional microphone from Acoustic Magic and a speaker. In a normal stand-up position near the robot, the mouth of a person is at a distance of 1 m from the microphone array. This is enough for enabling speech recognition in a quiet environment. This was, actually, the main motivation for adding the fiber structure. For robust navigation, a set of 10 IR sensors was added to the fiber structure. With this platform, we are developing an autonomous robot capable, not only of wandering around, but also of taking decisions, executing tasks and learning. Recently a laptop was added to support tactile interaction, text input and display of a computer animated synthetic face synchronized with the speech synthesis process (Seabra Lopes *et al.*, 2003).

The control and deliberation architecture of Carl (Fig. 2) reflects the goals of our project. Human-robot communication is achieved through spoken language dialog (Seabra Lopes and Teixeira, 2000*b*; Seabra Lopes and Teixeira, 2000*a*). A set of Linux processes take care of speech recognition, natural language parsing and speech synthesis. More processes deal with other kinds of information that comes from the outside through vision, sonars, infrared and collision sensors. Another Linux process handles general perception and action, including navigation. High-level reasoning, including inductive and deductive inference, is mostly based on the Prolog inference engine (we use a freeware implementation with a good C-language interface, SWI Prolog). Another module of the architecture provides Carl with learning capabilities. A central manager coordinates the activities at the high level (Seabra Lopes, 2002, for more details).

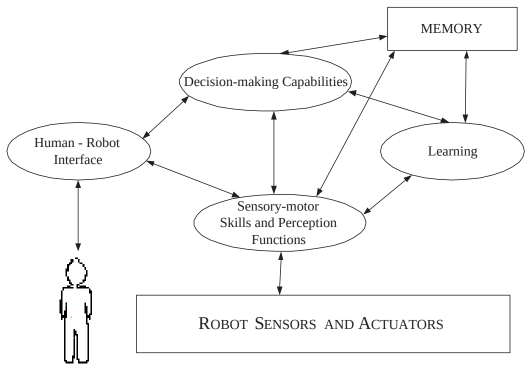

Fig. 2. Integration scheme of Carl’s several domains.

2. SPOKEN LANGUAGE INTERFACE

An interface function is to allow the communication between Man and machine. As Man’s main way of communication is the spoken language, the voice interfaces using natural language are the most intuitive for humans.

We can divide the architecture of an application by voice in the following main blocks: Speech Recognition; Natural Language Understanding; Dialogue Management; Natural Language Generation; and Speech Synthesis.

The recognition block converts the acoustic signal into a sequence of phonemes that form each word and afterwards the sentence. For an adequate interpretation of the recognized sentence the extraction of it’s syntactic structure is necessary. To do so there’s the module of Natural Language Understanding. This module structures the recognized sentence in a way that the Dialogue Manager can interpret it. The Dialogue Manager activity consists on interpreting the information of the user and react according to it. Therefore, the Natural Language Generation block, after having received the message that the robot wants to transmit to the user, builds a sentence. Because the message must be transmitted by voice, the Speech Synthesis Block generates an acoustic signal based on the text.

In this paper we present the last improvements made on the module of Speech Recognition (sec. 3), Natural Language Understanding (sec. 4) and Natural Language Generation (sec. 5).

3. SPEECH RECOGNITION

When someone imagines the faculty of speaking to a computer the first image that appears is usually the speech recognition that is the conversion of an acoustic signal into a flow of words. The recognition block is critical due to the fact of being affected by a relevant set of external conditions, as it is the case of the environment noise, parallel conversations and the noise of the robot's own engines. The sudden changes of context by the speaker and the quality of the microphone used are also relevant conditionings for the quality of the recognition.

The speech recognition module, existing in Carl since 1999/2000, was first implemented using Entropic's grapHvite and later using ViaVoice for Linux from IBM, using the grammar mode (Seabra Lopes and Teixeira, 2000*a*). The use of grammars has several disadvantages: limits the possible sentences, the non-recognition of one word may imply the non-recognition of the whole sentence, etc. There was, consequently, the need of evolution of this module so that it could deal with more sentences, have a better recognition rate and allow the training/adaptation of the models. Also the ViaVoice system for Linux was discontinued in the mean time, motivating our option for the development of a new module.

3.1 New Speech Recognition module based on Nuance 8.0

The Nuance system (Nuance, n.d.) uses Hidden Markov Models (Jurafsky and Martin, 2000), as most of current recognition systems do, as acoustic models in the association between the sound wave and its respective phoneme sequence.

Nuance recognizer is based on a client/server architecture. The recognition client is responsible for the acquisition and pre-processing of the audio input, the recognition server is responsible for the transformation of the sound wave into text, that is, for the recognition of the speech itself. This last one receives as entrance the voice signal and uses three specific components to accomplish the recognition: the acoustic module provided by the Nuance system, the dictionary files that contain the description of the phonetic pronunciations of the words defined in the grammar and that are also included in the system and finally the recognition grammars which define the set of the sentences that can be recognized. Efficiency of the acoustic processing is very important.

Another important characteristic of the Nuance's voice recognition system is its ability to allow the use of language models to restrict what the user can say; but allowing it to express itself in a natural way. Thus, a bigram language model (Jurafsky and Martin, 2000) was created, based on a representative group of sentences, sentences that are susceptible to be said in the context of CARL project. The sentences were generated using the grammar used previously for the ViaVoice in dictated mode.

Its also relevant to mention that this recognizer is capable of supplying alternatives to the most probable sentence, as well as the definition of the level of trust starting from which all the sentences are rejected. It was decided that five alternatives, maximum, would be kept and that any sentence associated to a level of trust inferior to 45% would be rejected. This percentage was chosen, by trial and error, in such a way that the sentences badly recognized due to interferences, weren't accepted, but at the same time that they don't cause a rejection to the minimum interference.

3.2 Evaluation

Two similar tests were made to the recognizer. The difference between the tests consists on the fact that one was developed on a silent environment (an investigation laboratory with several computers working) and the other tries to simulate the real environment of a robot demonstration through the existence of background noise (similar to the silent environment, with the only difference of the presence of several people talking at a reasonable sound level).

Table 1 presents the results of the test in the environment with more noise. The numbers in brackets ahead of the percentages are the respective number of words.

The tests weren't exhaustive for two reasons: first, the usual lack of time; second, what interests us most is the performance of the system as a whole; we weren't interested on concentrating on the optimization of one of the modules only. Results were satisfactory, enabling us to proceed with the replacement of ViaVoice by our new Nuance based Speech Recognition module.

We have verified that the most frequent mistake, either on the test with noise or in the test without

Table 1. Nuance Evaluation Results in a noisy environment

	1st choice		2nd choice		5th choice	
total sentences	67		63		33	
% sentences correct	47.76	(32)	7.94	(5)	0.0	(0)
total words (T)	323		308		185	
% correct words	79.57		68.18		72.43	
% replaced words (R)	13.00	(42)	24.03	(74)	19.46	(36)
% inserted words (I)	1.55	(5)	2.92	(9)	5.41	(10)
% deleted words (D)	7.43	(24)	7.79	(24)	8.11	(15)
WER	20.43		31.82		27.57	
Accuracy = (T-I-D-R)/T	78.02		65.26		67.03	

noise, corresponds to the substitution of words. The most frequent replacement corresponds to the substitution of the word *is* for the word *you*. The justification for this substitution is on the frequent bad pronunciation of the word by Portuguese speakers. This problem could be solved through a training of the models. Other word frequently substituted is *the*, replaced for *a* and vice-versa. Despite originating a badly recognized sentence, this substitution, doesn't influence the meaning of that one, diminishing its relevance.

The reason for the insertion of words is possibly different: it can be forced either by restrictions imposed by the language model in use, or by the existence of noise at the time the speaker says the word to be recognized. For example, in following situation when the speaker says: *"Where are you"* and the robot recognizes: *"Where are you eating"*. The insertion of the word eating in the recognized sentence was probably caused by the existence of noise at the time the test was taking place. It's also important to notice that the recognized sentences are still subjected to a level of trust and because of that, a sentence like the one given in the example would only be presented if the level of trust of the alternatives were inferior to the level of trust of that sentence. That justifies the fact that the percentage of insertions in the tests was reduced; once again we can conclude, by the fine performance of the recognizer (even in the test were there is noise).

Last, it's important to analyze the possible causes of the removed words. These, similar to words inserted, can have their origin in restrictions imposed by the language model. The problem can also emerge from the alignment of the words, that is, the possible inexistence of a correct identification of where one word ends and the other one begins.

4. NATURAL LANGUAGE UNDERSTANDING

Sequences of words coming from the speech recognizer are presently analyzed to extract semantic information using ALE (Attribute Logic Engine) (Seabra Lopes *et al.*, 2003). However, the voice recognition made by the recognizer isn't always interpretable nor, most of the times are the sentences that it returns grammatically correct. Supposing that the ALE manages to withdraw information from all the grammatically correct sentences, it becomes necessary to find another way of detecting and interpreting incomplete and badly recognized sentences, because each of the sentences that is not analyzed will be information that won't be available for Carl. The aim is to improve the semantic interpretation, namely at the level of the mutilated sentence processing, or without a grammatically correct structure. The creation of a module that would take care of these problems was then thought.

The Memory-Based Learning (MBL) (Aha, 1997) approach was adopted. The option is justified by the possibility of using training examples, created to serve our needs. The classification program TiMBL (Daelemans *et al.*, 2001) was used. Two tools were also used with the intention of increasing the information concerning each sentence. The first tool makes the attribution of a part-of-speech (POS) tag to each word (Jurafsky and Martin, 2000, p. 298) whereas the second tool makes syntactic segmentation, that is, the division of the sentence in the so-called chunks.

4.1 Results

Type of sentence - For a first adaptation to TiMBL and Brill POS Tagger (Brill, 2000), an experiment was made on the automatic determination of the sentence type. The sentence could be of five different types: declarative, imperative, interrogative, yes or no question, or false (in case it couldn't considered a sentence) (Jurafsky and Martin, 2000).

The training file consisted of information vectors containing the tags of the words in the sentence. To each sentence corresponds one information vector. Each vector had 13 positions or features. The features were only filled with the tags: feature 1 was filled with the tag of word 1, feature 2 was filled with the tag of word 2 and in this way successively. The features that remained empty

Table 2. Type of sentence results, showing percentage of sentences correctly classified. First row indicates percentage of sentences used for training.

%	1	5	10	20	40	60	80
Train	152	762	1.5k	3k	6k	9k	12k
Test							
1	60.4	80.4	90.0	92.8	94.6	94.9	95.0
2	44.3	84.7	90.1	92.6	94.1	94.6	94.8
3	39.9	87.1	91.0	92.2	93.8	94.3	95.2
4	76.7	89.2	89.7	92.5	93.7	94.2	94.7
5	47.4	89.6	92.1	92.3	93.7	94.3	94.7
6	67.9	88.8	91.6	93.1	93.8	93.9	94.4
7	48.3	87.5	87.4	92.8	93.5	94.4	94.0
8	78.3	82.8	92.4	92.3	94.0	94.3	94.8
9	47.5	86.2	90.6	92.7	93.7	94.5	94.7
10	59.3	86.6	90.6	92.9	93.8	94.5	94.8
mean	57.0	86.3	90.6	92.6	93.9	94.4	94.7
std	13.7	2.9	1.4	0.3	0.3	0.3	0.3

Table 3. Results, as percentage of correct decisions, of sentence/non-sentence tests. First row indicates percentage of sentences used for training.

%	1	5	10	20	40	60	80
Train	152	762	1.5k	3k	6k	9k	12k
Test							
1	85.4	91.5	93.8	95.0	95.9	96.2	96.4
2	81.7	93.7	92.7	94.9	95.5	96.2	96.1
3	83.9	91.5	94.0	94.9	95.5	95.4	96.5
4	92.0	93.5	94.0	95.1	95.4	95.6	96.5
5	79.4	94.2	94.7	94.7	95.7	96.0	95.8
6	81.3	92.7	94.9	95.3	95.6	95.5	95.6
7	52.5	91.7	93.3	95.0	95.3	96.0	95.4
8	83.8	91.0	94.1	94.9	95.5	96.1	96.0
9	83.8	94.0	94.3	95.0	95.5	96.3	96.4
10	72.1	92.7	92.9	94.8	95.6	95.7	96.2
mean	79.6	92.7	93.9	95.0	95.5	95.9	96.1
std	10.7	1.2	0.7	0.2	0.2	0.3	0.4

were filled with "-". The 13th feature was the classification to give to the sentence.

We started by creating a training file with 20 sentences and tested on 100 new sentences. The result was very good: 98% of correct decisions. Only two sentences have been incorrectly classified. After the problem was analyzed, we came to the conclusion that the mistake was due to a bad classification of the tagger, which classified "move" as a name. The results obtained allowed the creation of excellent perspectives as far as the future use of these tools is concerned. Later, with availability of a reasonable sized corpus, this experiment was repeated, now using as features not only the tags but also sentence words each followed by its POS tag. Results are presented in Table 2. For a training set of more than 1500 sentences, results are above 90 %, raising to around 95 % for training sets of 12000 sentences.

Detection of non-sentences - The objective was to prepare a program to filter the sentences that come out of the voice recognizer. For that reason, among the hypothesis given by the recognizer is the need of identification of the grammatically correct sentences. To solve this problem two programs were created. The difference between them lays on the tools used to withdraw information from the sentence. The first uses only the tagger; the second uses the tagger and the chunker. The training file that TiMBL uses is identical in both cases. The formation of the training file information vectors is the following: the two first features are the first word of the sentence and respective tag; the third and fourth positions are the second word of the sentence and respective tag, and in this way successively. The 19th position is the sentence classification. Results are presented in Table 3. With 762 examples in the training set correct decisions are above 90 %. 3000 examples are enough to obtain 95 % correct decisions. Results are very similar for the 10 different runs, resulting in low standard deviation of the results.

Other exploring experiments - The AI module, at this time, only accepts ALE information. When none of the alternatives presented by the recognizer is correct, it is necessary to try the reconstruction of the sentence. For that, a function, which tries to perform sentence reconstruction, was added to the program that filters the sentences coming from the recognizer. A common error was detected in the recognized sentences: the words *in*, *is* an *you*, are often mixed up. So the first step of the reconstruction function is to search for these words in the sentence, and create three new sentences, substituting one word for the others.

The next step is the creation of new sentences with combinations of the initial sentence words. The ALE has a function that allows recovering incorrect sentences. This function tries to fill any blank spaces that the sentence may have, in a way that it can be correctly restructured. Thus, a program, which sends to ALE sentences with wholes, was created. At the time of writing this program has not been tested yet.

5. NATURAL LANGUAGE GENERATION AND SPEECH SYNTHESIS

The representation of information enclosed in a computer application contains meaning for itself, but it can be difficult for a normal person to understand it. Consequently, the work of a generator starts with the initial intention of communicating, choosing the contents to be transmitted, selecting the words and organizing ideas. It concludes by creating a grammatically correct sentence or forming texts in natural language. The tasks are divided in three main processes: text planning, phrase planning and surface realization (Becker, 1999).

The most divulged contemporary text generators are the so-called Canned Text Systems, systems using pre-defined sentences: the system simply generates sentences without any change (error messages, warnings, etc.). The Template-Based Systems are the next level of sophistication. They use permanent text structures where only small alterations are made. Another type of systems are the Phrase-Based Systems that use what can be seen as generalized structures, whether at the sentence level (where the structures are similar to sentence structural grammatical rules), or at the speech level (where they're often denominated as text plans). The Feature-Based Systems represent the state-of-the-art. The main disadvantage of this type of system is in the complexity of the necessary information entrance. It's in this class of generators that the more advanced systems can be found, as an example we have the PENMAN/KPML and the FUF (Jurafsky and Martin, 2000, chapter 20).

5.1 NLG for Carl

The functioning of NLG module in Carl will be the following: the artificial intelligence module (AI) will supply the generator the idea to be transmitted. The idea will already be divided in information blocks that represent sentences; the generator will create one or more sentences in natural language that can transmit the idea; the NLG module will supply the sentences to be converted in speech signal to the synthesis module. Hence, the generation module necessary to Carl project will only aim the making of a surface execution, having three intrinsic components: the syntactic component, the morphologic and the orthographic components, this last one being of insignificant relevance.

We choose to use the ASTROGEN (Dalianis, 1999), developed at the University of Stockholm, as a tool for the development of our sentence generation module.

One of the objectives that we intended to achieve was to make Carl capable of answering questions about IEETA (the investigation institute where Carl is being developed) and about himself. A database was also created containing the necessary information for him to be up to these situations. This sentence generation module was also responsible for the access to data base information.

The database was built in Prolog to facilitate the interaction with other programs. The information that it holds is stored under the form of predicates, providing information regarding entities attributes (laboratories, projects, professors and Carl), relations between the several entities and IEETA internal map. The type of all entities is also defined.

The main predicates are:

- type (entity, type)
- relation (relation, entity1, entity2)
- attribute (entity, type of attribute, attribute)
- cabinet (entity, place)
- place (cabinet, localization)

It's using these attributes that the AI module specifies which is the information to transmit. The way to call the sentence generation is: *Generate (M,L)*. M is the information list to transmit, and the L variable will receive the synthesized sentence. The M list can have the following forms:

- a combination of relation predicates, cabinet and type;
- [researchers (laboratory or project)];
- [labs];
- [projects];
- [responsible (laboratory or project)].

5.2 Results

Table 4 represents some examples of sentences generated by the NLG module. The results are almost related to the Carl project, however, identical information, related to other projects or laboratories, is processed the same way. Always generating sentences of the same type on the presence of information with the same format is a problem that is pointed out to the Astrogen. Therefore, if it's necessary to give information to the same person about Carl project investigators, in two consecutive answers, the result will be:

- professor seabra is a researcher of the carl project.
- Professor antónio teixeira is a researcher of the carl project

This is one of the elements to be corrected. Also bugs and lack of robustness prevented integration with other Carl modules.

5.3 Speech Synthesis

Conversion from text to speech continues to be made by using the IBM ViaVoice TTS system. An exploratory experiment with limited domain synthesis (Black and Lenzo, 2000) using Festival framework was done recently. Due to the limited vocabulary used by Carl to convey information to the user, it was possible with a small amount of recording and processing to develop a new voice with a more natural quality.

Table 4. Some examples of sentences generated by the NLG module.

List M	Sentence Generated
[responsible(carl_proj)]	professor seabra is the coordinator.
[projects]	projects are: carl project and f. c. portugal.
[researchers]	researchers are: seabra lopes antonio teixeira.
[relation(researcher,carl_proj,X)]	professor antonio teixeira is a researcher of the carl project.
[cabinet('antonio teixeira',X)]	professor antonio teixeira is in the second floor in the cabinet 210.
[type('seabra lopes',X)]	professor seabra lopes is a researcher.
[relation(researcher,carl_proj,Y), relation(researcher,carl_proj,X)]	professor antonio teixeira and professor seabra lopes are researchers of the carl project.
[cabinet('antonio teixeira',X), cabinet('seabra lopes',Y)]	professor antonio teixeira is in the second floor in the cabinet 210 and professor seabra lopes is in the first floor in the cabinet 111.
[type('seabra lopes',X), type('antonio teixeira',Y)]	professor antonio teixeira and professor seabra lopes are researchers.
[relation(researcher,carl_proj,Y), cabinet(Y,X)]	professor seabra lopes is in the first floor in the cabinet 111 and professor seabra lopes is a researcher of the carl project.
[type('carl',X), relation(researcher,carl_proj,Y)]	carl is a robot and professor seabra lopes is a researcher of the carl project.

6. CARL'S PARTICIPATION IN A RECENT ROBOT CONTEST

Carl participated recently on the first edition of the "International Cleaning Robots Contest", which took place in Lausanne, in Switzerland. This event was integrated in the "2002 IEEE/RSJ International Conference on Intelligent Robot Systems (IROS'2002)", one of the most important conferences in the intelligent robotic branch, where a communication about Carl was presented.

The contest includes three competition modalities: "floor cleaning", "window cleaning" and "housekeeping ideas". Carl played the butler role in the third of these modalities, demonstrating advanced human-robot capacities of interaction.

In Lausanne, Carl demonstrated his capacities during IROS'2002 opening reception raising an enormous curiosity among the conference participants. The only rival it had in the "housekeeping" modality was a robot that watered plants. Both have been awarded.

In this contest Carl used the new version of the recognizer and natural language understanding module that identifies sentences and non-sentences.

7. DISCUSSION

7.1 Speech Recognition

The work presented had as main objective the implementation of a new voice recognition system.

Analyzing the results obtained from the Nuance voice recognition system, we observe that despite the imposition of a limitation of the language used through a language module, the users don't need to restrict themselves to one limited set of commands; continuing to obtain satisfactory results. We should notice, however, that when imposing a language model of this type the vocabulary that can be recognized is restricted, that is, limitations are imposed in the quantity of words liable to be recognized.

The problems that appeared with the use of the Nuance were connected, with relative frequency, to the substitution of words. Such as referred previously, these problems can possibly be reduced through training of the acoustic models in a way that they can be adapted to the Portuguese users.

Therefore, the future of the voice recognition module development passes by the implementation of adaptation of the models to the users and to the environment that surrounds the robot.

7.2 Natural Language Understanding

The introduction of new language processing techniques in CARL project will have, in short term, an important role to play. The great advantage of this approach is the freedom and robustness that it gives us, unthinkable with the use of grammars.

Results for type of sentence and sentence/non-sentence distinction were very good. Even with a small amount of training data results were around 95 %. Sentence/non-sentence was integrated in Carl and used in a recent demonstration at a robot contest with good results.

Continuation of the exploratory experiment on the application of MBL to the recognition of the subject and object of a sentence, after the creation of training files, will eventually result in new module capable of performing some syntactic and semantic analysis on non-sentences, leaving for ALE only the sequences classified as sentence. This approach gains strength when the reconstruction of sentences is necessary. If we have a program that allows an understanding based only in a few words of the sentence, it would be easier to get around the mistakes, and withdraw information.

This approach is also interesting for a planned future version of Carl with a spoken language interface using Portuguese. Creation of a corpus for training the modules, despite labor intensive, can be done with less knowledge of the developer about syntax and semantics.

To conclude, work in this area was mainly a start, a trial of the capacity of this type of approach.

7.3 Generation

The NLG module processing depends a lot on the module that commands the AI module. The interaction between these two modules is fundamental to achieve the generation in a competent way. This was one of the problems found in the accomplishment of this task. The quantity of information that the Astrogen needs to make the generation is superior to the one made available by the AI module today. Although it gets a lot of the information that it uses from the database, it still needs the evolution of the AI module.

One other problem, possibly even more serious, was the difficulty that there was in the adaptation of the Astrogen to our needs. Astrogen is only prepared for lexis and grammar alterations, which won't permit the modification of original sentence schemes, what makes it too limitative. It was necessary to add new words, and consequently, new grammar rules, leading to a deficient Astrogen internal treatment of this new information. Despite all the problems, the NLG module created is adaptable to Carl, executing the required task.

8. ONGOING AND FUTURE WORK

As an ongoing project, Carl is continuously evolving. We are now working on the application of partial parsing to the natural language understanding module. A new syntactic analyzer is being developed using LCFlex (Rosé and Lavie, 2001). Planned for the near future are: improvements to the NLG module, improving connection between NLG and speech synthesis, and adaptation of the speech recognition acoustic models.

ACKNOWLEDGMENTS

Major part of the work reported in the paper was done as final year licentiate training post (*Estágio*) of University of Porto Department of Applied Mathematics *Licenciatura* in Mathematics Applied to Technology. Our thanks to Prof. Ana Paula Rocha and Teresa Mendonça.

REFERENCES

Aha, D. W., Ed.) (1997). *Lazy Learning*. Kluwer Academic Publihsers. reprinted from: Artificial Intelligence Review, **11**:1-5.

Becker, T. (1999). Natural language generation: Some examples.

Black, A. and K. Lenzo (2000). Limited domain synthesis. In: *Proc. ICSLP*. Beijing, China.

Brill, E. (2000). Part-of-speech tagging. In: *Handbook of Natural Language Processing* (Robert Dale, Hermann Moisl and Harold Somers, Eds.). Marcel Dekker.

Daelemans, W., J. Zavrel, K. van der Sloot and A. van den Bosh (2001). *TiMBL: Tilburg Memory Based Learner Reference Guide, v. 4.1.*

Dalianis, H. (1999). A general validation tool by natural language generation for the STEP/EXPRESS Standard.

Jurafsky, D. and J. H. Martin (2000). *Speech and Language Processing*. Prentice Hall.

Nuance (n.d.). *Nuance Speech Recognition System, version 8.0: Introduction to the Nuance System*. And the other manuals.

Rosé, C. Penstein and A. Lavie (2001). Balancing robustness and efficiency in unification-augmented context-free parsers for practical applications. In: *Robustness in Languages and Speech Technology* (J.-C. Junqua and G. van Nord, Eds.). Chap. 10. Kluwer Academic Publishers.

Seabra Lopes, L. (2001). Carl, a learning robot, serving food at the AAAI reception. In: *Proc. of the AAAI Mobile Robot Competition and Exhibition Workshop*. Seattle, WA, USA.

Seabra Lopes, L. (2002). Carl: from situated activity to language level interaction and learning. In: *Proc. IEEE IROS*. pp. 890–896.

Seabra Lopes, L., A. Teixeira, D. Gomes, C. Teixeira, J. Girão and N. Sénica (2003). A friendly high-level human-robot interface. In: *Encontro Científico do 3o. Festival Nacional de Robótica*.

Seabra Lopes, L. and A. Teixeira (2000*a*). Human-robot interaction through spoken language dialogue. In: *Proc. IEEE/RSJ IROS*.

Seabra Lopes, L. and A.J.S. Teixeira (2000*b*). Teaching behavioral and task knowledge to robots through spoken dialogues. In: *My Dinner with R2D2: Proc. AAAI Spring Symposium on Natural Dialogues with Practical Robotic Devices*.