Technical Report No 98/309, Département Informatique, EPFL, Lausanne

INTEGRATION OF SYNTACTIC CONSTRAINTS WITHIN A SPEECH RECOGNITION SYSTEM

COUPLING A SPEECH RECOGNIZER AND A CONTEXT FREE PARSER

Ramon Aragues Peleato, Martin Rajman, Jean-Cedric Chappelier

February 23, 1999.

LIA-EPFL

Table of contents

1 - Introduction

Automatic speech recognition is one of the important research domains in the field of man-machine interaction. Although the quality of the different existing speech recognizers is continuously increasing there is still room for further accuracy improvements. One of the promising approaches is the integration, in the acoustic modules, of higher level linguistic knowledge, such as, for example, syntactic constraints.

In this context, the present report focuses on the study of how to efficiently couple an existing speech recognizer and a syntactic analyzer. The method considered here corresponds to the sequential integration of the two modules: the output of the speech recognizer is given as input to the syntactic parser, which filters out the hypotheses of the speech recognizer that are not syntactically correct (c.f. Figure 1.1).

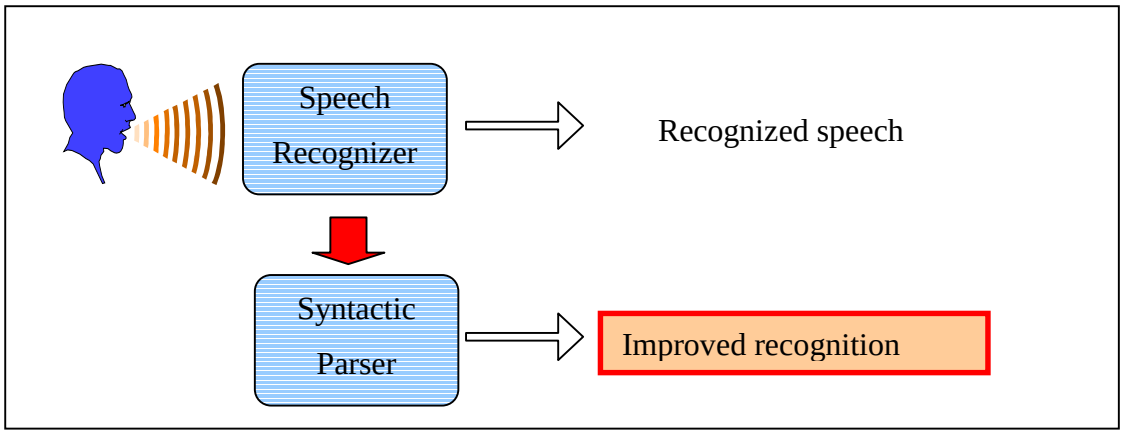

Figure 1.1 – The objective of the coupling: the syntactic parser, processes the results of the speech recognizer filtering out syntactically incorrect solutions.

This report analyzes the results of the speech recognizer and then compares them with the results obtained after the coupling. The analysis is done by comparing the 1-best sentences, the N-best lists and the time cost of the whole recognition system, and will establish the gains and the losses of the coupling. For a more detailed description of the work described in this report refer to [Arag98].

2 - The speech recognizer

2.1 – General description

The speech recognizer used in our experiments (STRUT + Noway; appendix A) produces as output word lattices (c.f. Figure 2.1) corresponding to thousands of hypotheses for the spoken sentence. Each path in the lattice is probabilistically weighted and contains one of the hypothesis of the speech recognizer. The probabilities are calculated on the basis of the acoustic likelihood of the words contained in the sequence and the probabilities provided by the language model (word bigrams or trigrams) used in the process.

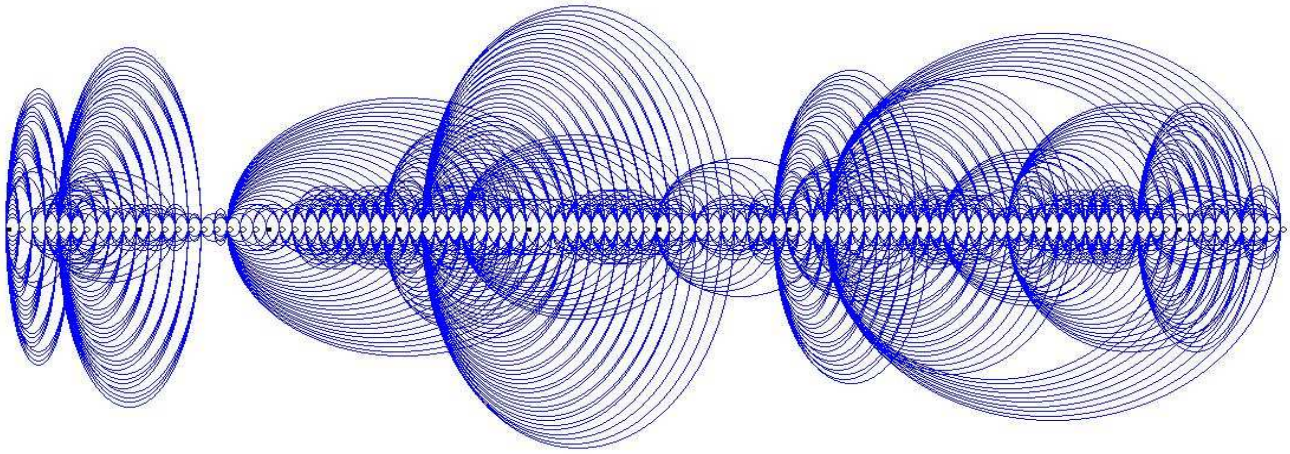

Figure 2.1 – An example of the word lattice the speech recognizer produces for the sentence "Un paradis terrestre possède un fleuve". Nodes are time instants and arcs are probabilistically weighted word hypotheses.

The number of paths in the word lattice (c.f. Table 2.1) represents the number of different word sequences that the syntactic parser will analyze to determine whether they are correct or not.

Table 2.1 - Number of paths inside the lattices obtained from experiment with $\alpha = 0.20$. The description of the experiments and the sample sentences can be found in appendix A. Due to the high computational cost of the procedure, we did not track more than 50 000 000 000 paths.

	Sentence 5	*Sentence 10*	*Other sentences*
Number of paths	3 261 227 208	6 009 538 238	> 50 000 000 000

The experiments (appendix A) were realized over 10 sample sentences from the Swiss French Polyphone 1.0a database [Polyphone]. The dictionary used for recognizing these sentences contained 11 064 words, and the language model used had 1 186 664 mono-grams and 28 861 312 bi-grams. All the experiments used the same values for the different parameters of the speech recognizer, except for the scaling factor (hereafter called α) which varied from 0.15 to 0.70. The experiments were realized on an Ultra Sparc 1 Sun Workstation running under Solaris 2.5.

2.2 – 1-best results

Table 2.3 presents the word error rate in the 1-best hypothesis for each of the experiments in appendix A. The error percentage has been calculated globally for the ten, therefore representing the word rate error of the concatenation of the ten sentences.

Table 2.3 - Word error rate with the speech recognizer for our experiments . Further details about these experiments my be found in appendix A.

Sentence	Number of Words	Word error rate in experiments						
		$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
1	7	4	5	5	5	5	5	6
2	10	**0**	1	1	1	1	2	4
3	6	3	3	3	3	3	4	8
4	14	15	13	16	12	12	13	19
5	7	7	7	6	4	4	4	8
6	5	2	3	2	1	1	1	7
7	4	4	3	2	2	2	5	9
8	7	7	7	6	6	7	7	9
9	6	4	4	2	2	2	2	2
10	3	3	**0**	3	2	2	2	3
Total	69	49	46	46	38	39	45	75
Error %		**71%**	**66%**	**66%**	**55%**	**56%**	**65%**	**108%**

The high word error rates obtained confirm the importance of finding alternative methods for speech recognition. In these experiments, only 2 out of the 70 cases correspond to a perfect recognition (i.e. zero values). However, it must be noted that these results do not necessarily represent the speech recognizer best performance because no specific tuning of the speech recognizer parameters was performed and therefore the recognition conditions were not optimal.

As the purpose of this work is not to find the optimal tuning of the speech recognizer, but to determine the effect of the coupling over the recognition process, the experiments have been performed with the actual system conditions.

2.3 – N-best results

The high word error rate obtained (section 2.2) is not significant when establishing the quality of a speech recognizer that couples with a syntactic parser. Therefore, we introduce the measure that finds the position of the spoken utterance in the N-best list (c.f. Table 2.4). For positions bigger than 500 the test was not realized due to the high time and memory spent, and it is only showed the distinction between those lattices that do not contain the sentence and those that may contain it.

Table 2.4 – Position of the spoken sentence in the 500-best list produced by the speech recognizer.

?? Not found in the 500-best hypotheses
--- Not in the lattice

Sentence	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$
1	---	---	---
2	**21**	**46**	??
3	---	---	---
4	---	---	---
5	---	---	---
6	??	??	??
7	---	??	??
8	---	---	---
9	**12**	**18**	**23**
10	**1**	**2**	---

Only 3 sentences out of 10 were found in the 500-best list. As explained in section 2.2 this could however be improved with a deeper tuning of the speech recognizer.

2.4 – Word occurrence results

A lattice that do not contain one or more of the words said by the speaker will never achieve a good recognition with the coupling, because the syntactic parser performs a filtering of the sequences of words existing in the lattice and does not create new sequences. Therefore, a third measure is introduced, which finds the number of words said by the speaker that are not contained in the lattice. Further details about these results can be found in appendix C.

Table 2.5 – Number of words that do not appear in the lattice produced with the different experiments (appendix A). The results used to create this table are delivered as appendix C.

*** Lattices that contain all the words said by the speaker.**

Sentence	Number of Words	Number of words not appearing in the lattices							
		$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$	
1	7	1	1	1	1	1	1	2	
2	8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	*****
3	5	1	1	1	1	1	1	1	
4	12	5	4	4	4	3	3	4	
5	7	3	2	3	2	3	3	3	
6	5	**0**	**0**	**0**	**0**	**0**	**0**	1	*****
7	4	2	1	**0**	**0**	**0**	1	1	*****
8	7	2	1	1	1	1	1	2	
9	6	**0**	**0**	**0**	**0**	**0**	**0**	1	*****
10	3	2	**0**	**0**	1	1	1	1	*****

The conclusion that can be derived from this new measure is that there is not one single best experiment to obtain good lattices. This is the reason why the use of several experiments in the recognition process is recommended, for instance the mixture of experts. In this table, 5 out of 10 sentences contain words that are never produced by the speech recognizer, which will make impossible for the syntactic parser to find the sentence said by the speaker. These results stress the importance of the objectives previously described in the report: as the use of acoustic knowledge and low-level language models do not yield optimal results, it should be tried to include higher linguistic knowledge in the recognition, with the aim of improving these results.

On the other hand, the speech recognizer must be tuned so that the sentence said by the speaker can be found among the different hypothesis of the lattices. Without this requirement the coupling will never bring recognition scores to a high level, even if it is capable of improving the recognition.

3 – Coupling the speech recognizer and the syntactic parser

3.1 – General description of the coupling

As explained in section 2, the speech recognition systems use acoustic information and low level language models to produce word lattices. In the coupling process, this word lattice is the input of a syntactic parser that filters out the syntactically incorrect hypotheses (c.f. Figure 3.1).

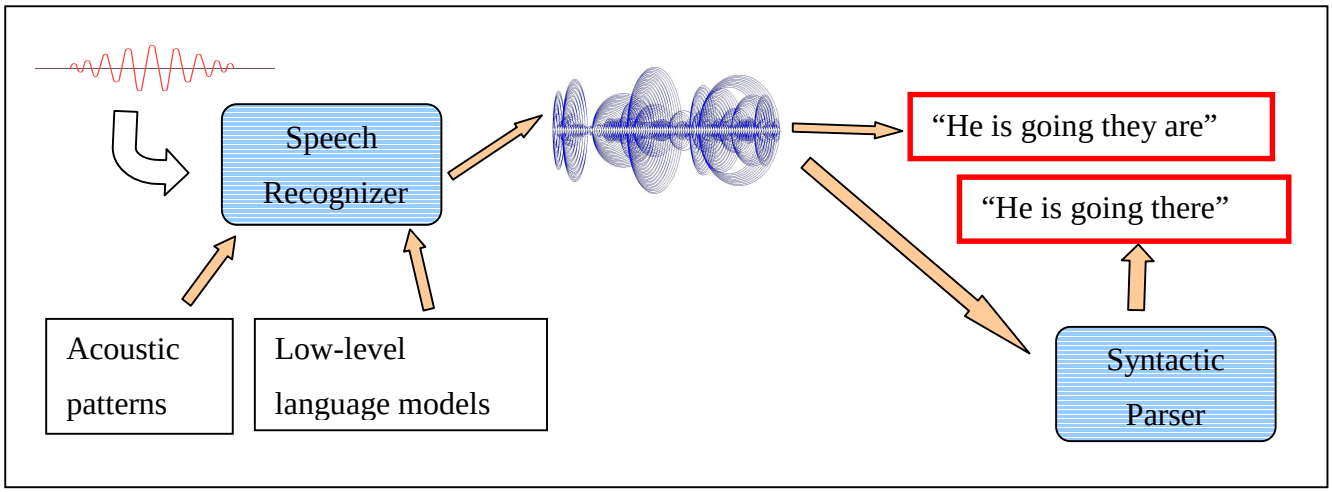

Figure 3.1 – The concept of the coupling. The word lattice is the input of the syntactic parser that filters out the syntactically incorrect hypotheses.

However, to allow the coupling between the two modules two operations must be realized. First of all, it will be necessary to remove the silences that are represented in the lattice, which are meaningless from the syntactic point of view. Once the silences are removed, a data transformation from the word lattice to a representation compatible with the syntactic parser is required.

The syntactic parser used for the experiments was the one included in the Syntactical language processing toolkit (SLP toolkit). This toolkit has been developed at the EPFL's "Laboratoire d'Intelligence artificielle" [ChRaj98a] [ChRaj98b].

The syntactic parser uses two main resources: a lexicon and a grammar. The lexicon used in the experiments had 14 076 entries corresponding to the words contained in the dictionary used by the speech recognizer. A toy grammar with 140 rules was created to perform the analysis of the 10 sentences used in our experiments. Future experiments on a longer scale will have to consider the use of a complete grammar of the considered domain.

3.2 - The results of the coupling

3.2.1 - Comparison of the number of solutions

The number of possible sentences produced by the speech recognizer and the syntactic parser (Table 3.1) will be fundamental in contexts where it is necessary to have access to all the possible solutions. A small number of possible solutions is desired, so that the result is not useless because of its large size. Further details about the number of interpretations for each experiment may be found in appendix D3.

Table 3.1 – Comparison of the number of solutions obtained in experiment with $\alpha = 0.20$ and the analysis performed by the syntactic parser

Sentence	**Number of hypotheses in the lattice**	**Number of syntactic interpretations**	**Reduction**
1	>50 000 000 000	521 234	
2	>50 000 000 000	862 334 592	
3	>50 000 000 000	4 594 041	
4	>50 000 000 000	3 326 421 643	
5	3 261 227 208	984	$3.3 * 10^6$
6	>50 000 000 000	290 658	
7	>50 000 000 000	5 873	
8	>50 000 000 000	92 322	
9	>50 000 000 000	26 740	
10	6 009 538 238	273	$22 * 10^6$

This table confirms that the syntactic parser drastically reduces the number of solutions present in the lattice. The reduction obtained is very representative of the number of solutions of the speech recognizer that were syntactically incorrect, and were filtered out by the parser.

3.2.2 – Comparison of the recognition with and without the coupling

A definitive comparison of the results with and without the coupling can not be done at this moment, because more work is needed in order to solve the following problems:

- The grammar used in our experiments by the syntactic parser is only a toy grammar dedicated to the 10 sentences considered.

- The coupling considered here was the simplest possible: the syntactic parser does not make use itself of the acoustic information contained in the lattice. The analysis of the word sequences provided by the speech recognizer is made as if all the words had the same acoustic probability.

Even with the actual conditions, excellent results have been obtained in some cases with this simple coupling (Table 3.2). Since the results of the syntactic parser are only based in syntactical probabilities, it is feasible to think that introducing the acoustic probabilities into the process should have desirable effects: it should give the most syntactically and acoustically probable sentence, which will be probably better than the results shown in Table 3.2.

Table 3.2 – Comparison between 1-best hypothesis of the speech recognizer and the 1-best interpretation of the syntactic parser. Highlighted words in the input utterance indicate words not contained in the lattice produced by the speech recognizer. The lattices used in the experiment are those produced with $\alpha = 0.20$. The "*comparison*" column indicates if the coupling has improved (+), worsen (-) or had not effect (=) over the solution

Sentence said by the speaker	**1-best hypothesis of the speech recognizer with $\alpha = 0.20$**	**1-best interpretation of the syntactic parser using experiment with $\alpha = 0.20$**	**Comparison**
La Suisse recule dans tous les **domaines**	La Suisse recul entouré de main	L'suisse recul entoure le demain	=
Le président du parti est aussi président du comité national	Le président du parti tous aussi président du comité national	L'président du parti est aussi résident du produit national	-
Un paradis terrestre possède un fleuve	Un paradis terrestre aussi de fruits	Un parti terrestre possède un fruit	+
Le plastique a renvoyé l'**acier dans** les **limbes** de l'âge de **fer**	Le parti cas en vol à six mille vingt l'âge de l'heure où	Une plastique à envoyé a finalement large du hereux de nous	+
Je n'**ai jamais** supporté le monde	Si nous allons supporter l'homme	Ils avons supporté l'âme	=
Ils s'installeront en France	Il s'installe en France	Ils installent en France	=
Les problèmes restent nombreuses	I problèmes restent nombreux six	Un dix populaire est nombreux de ce	-
Les causes de cet engouement sont **variées**	Le camp de ces temps de maçons paris	Les causes de cette tendance ont parti	+
Je me sens bien avec eux	Je ne sont bien avec eux	Je est cent bien avec eux	=
La roue tourne	La route ou deux	La roue tourne	+

Even with the limited conditions used, the coupling improves the recognition in 4 sentences out of 10. There are some problems with the actual coupling that will be easily solved when changing some of its characteristics. For example, in the second sentence, the syntactic parser is giving as correct *"... résident du ..."*. The word said by the speaker was *"président"*, but *"résident"* and *"président"* have the similar lexical probabilities, and the parser gave the wrong sentence because it does not look at the acoustic probability of each of the two words (same problem for "*produit*" and "*comité*"). Some techniques to be applied in the future to solve this problem are discussed later in this section and in section 4.1.

Another way to improve the results consist of the tuning of the grammar. Different tests have been made to analyze its effect, with the objective of determining if future experiments using more appropiate grammars will obtain better results. The results of these experiments indicate a slight general improvement in most of the sentences when the grammar is improved, and a general decrease of the word error rate.

Table 3.3 shows the results of the comparison with and without the coupling for the different experiments. All the 1-best interpretations of the syntactic parser are delivered as Appendix D1.

Table 3.3 – Comparison between the 1-best hypothesis of the speech recognizer and the 1-best interpretation of the syntactic parser in the different experiments. Further details about the results may be found in appendix D1. For each experiment it is shown if the coupling has improved (+), worsen (-) or had not effect (=) over the solution

Sentence	Comparing the recognition with and without coupling					
	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$
1	=	=	=	=	=	-
2	-	-	-	=	=	-
3	-	-	+	+	+	+
4	=	+	+	+	+	-
5	=	-	=	-	-	-
6	-	=	=	=	=	=
7	=	-	-	-	-	-
8	+	+	+	+	+	+
9	=	+	=	+	+	-
10	=	-	+	+	+	+
Total +	1	3	4	5	5	3
Total =	6	2	4	3	3	1
Total -	3	5	2	2	2	6

This table shows that the best results of the speech recognizer (those with $\alpha = 0.20 \ldots 0.30$) are the ones that are more improved. For example, for experiment where $\alpha = 0.25$, 5 sentences out of 10 have been improved with the coupling, and only two were made worse.

The experiment of using the acoustic information contained in the lattice for the syntactic process gave the results shown in Table 3.4. In this experiment, the lexical probabilities of the words were multiplied by the acoustic probabilities contained in the lattice. As done in Table 3.3, a qualitative measure is used, were the 1-best sentences with and without the acoustic information are compared.

Table 3.3 – Comparison between the 1-best sentences obtained with and without the acoustic information. Further details about the results may be found in appendix D2. For each experiment it is shown if the acoustic information has improved (+), worsen (-) or had not effect (=) over the solution of the coupling.

Sentence	Comparing the recognition with and without the acoustic information					
	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$
1	=	-	=	=	-	=
2	=	+	+	=	=	=
3	=	+	=	=	=	=
4	+	-	+	=	=	=
5	=	=	=	+	=	=
6	+	+	+	+	+	+
7	+	+	=	=	=	=
8	=	+	+	=	=	=
9	=	=	=	=	=	=
10	=	=	-	=	=	=
Total +	3	5	4	2	1	1
Total =	7	3	5	8	8	9
Total -	0	2	1	0	1	0

This table shows that the use of the acoustic information has a positive effect over the results. It is improving of leaving the same solution for most sentences, and the negative solutions are isolated cases.

The comparison between the N-best list of the lattice and the sentences found by the syntactic parser did not produce remarkable results. The impossibility of constraining the number of interpretations of the syntactic parser to a given N >1 (algorithm still needs to be implemented) made not possible to find the position of the spoken sentence in the N-best list resulting from the coupling.

3.2.3 - Time Performance of the coupling

The time necessary for the speech recognizer to produce all the different hypothesis will be significant when validating the coupling. The influence of the time spent of including the syntactic parser in the process will have to be analyzed to determine the gains and losses of the coupling.

Table 3.5 shows the different timings obtained for the lattice creation and each of the stages of the recognition process: the speech recognizer, the data translation, the parsing of the sentences and the other operations involved in the coupling. Further details of the timings are in appendix D3.

Table 3.5 – Average times for the different steps of the recognition. Further details of the timings are in appendix D2.

Experiment	Average speech recognizer time (s)	Coupling total time		
		Average translation time (s)	Average parsing time (s)	Filtering + I/O (s)
$\alpha = 0.15$	219	0.077	0.54	0.1
$\alpha = 0.20$	209	0.083	0.67	0.1
$\alpha = 0.25$	81	0.079	0.63	0.1

Knowing the total time spent of the coupling, the analysis of its influence over the total recognition process can be realized. The results shown by Table 3.5 validate the coupling, as it does not significantly slow down the recognition process (Figure 3.2).

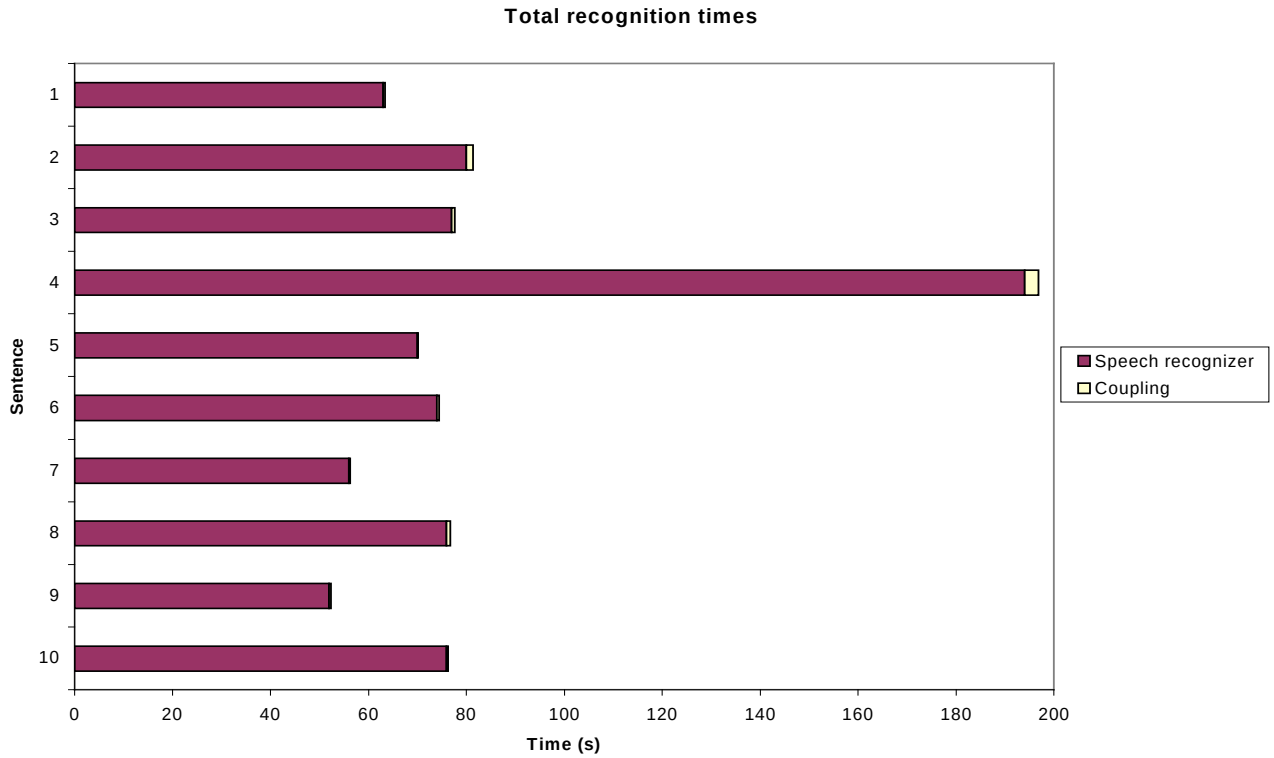

Figure 3.2 – Division of the recognition times between the speech recognizer ($\alpha = 0.25$) and the coupling.

Figure 3.3 represents the different stages of the coupling, showing that the main process in the coupling is the analysis of the different hypotheses of the speech recognizer. Figure 3.3 is an extension of Figure 3.2, where only the times concerning syntactic analysis are shown.

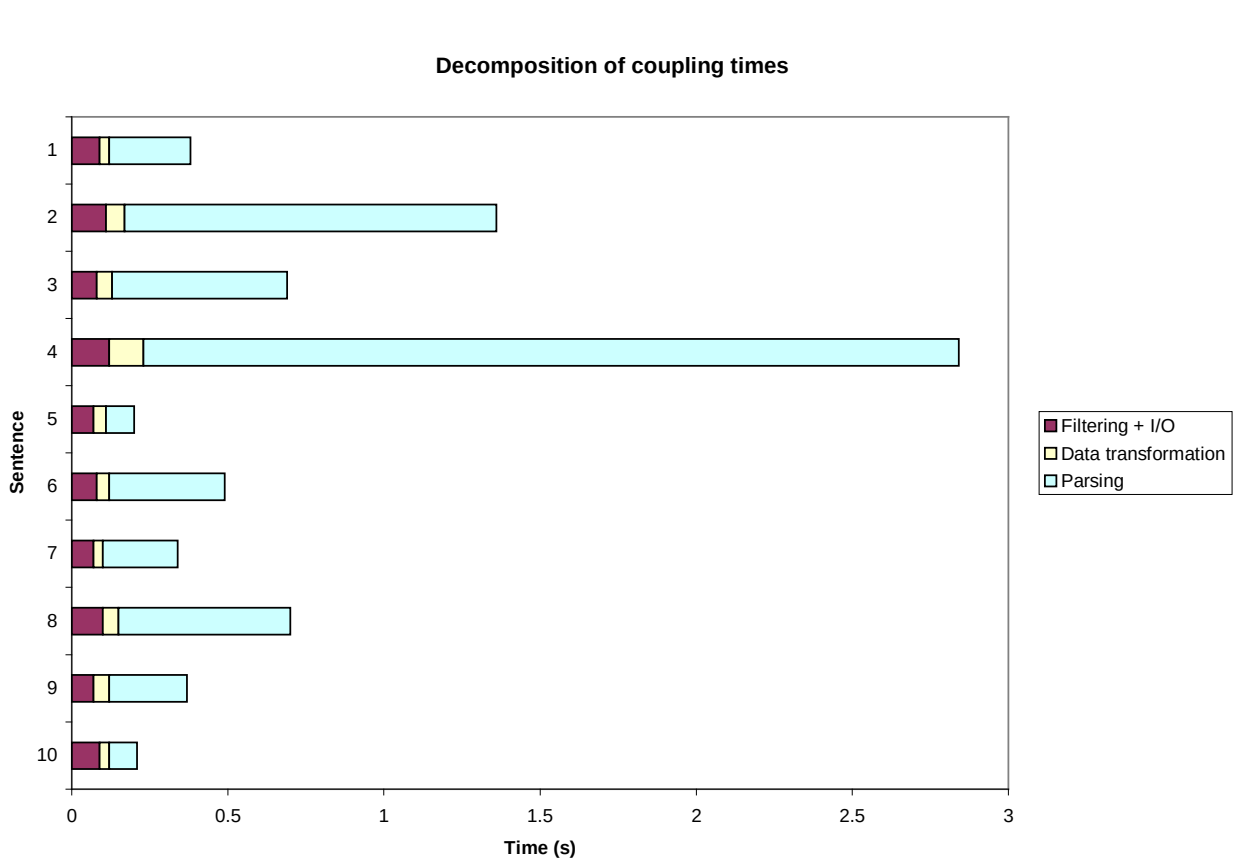

Figure 3.3 – Decomposition of coupling times in the different stages. Experiment realized using $\alpha = 0.25$.

Removing the silences from the lattice and reading the different files involved in the coupling are included under the name "*Filtering + I/O*". The "*data transformation*" refers to the translation of the lattice into a structure compatible with the syntactic parser, and the "*parsing*" refers to the analysis of the hypotheses of the speech recognizer.

Our final conclusion on the results of the coupling is that even though the conditions used for this very first try were not optimal, the improvements obtained seems promising. We envision that the implementation of several optimizations, as discussed further in section 4.1, will produce better results.

4 – Conclusions

4.1 - Future directions

Although the results we have obtained do not fully validate the coupling, they have shown the deficiencies of the actual system that must be corrected. The work described in this report opens many promising paths that can be followed in order to solve those problems.

The approach to the coupling in this work described in the report is sequential, where the input of the syntactic parser is the output of speech recognizer, without any further feedback. If the sequential approach demonstrates promising performance of the coupling, other approaches will have to be implemented, as for example the tight coupling [Gil97].

Concerning the sequential coupling there are however several enhancements that are still possible:

1 - Combining acoustical and syntactical probabilities

The first problem in the actual coupling is the way that the input of the syntactic parser is using the acoustic information contained in the lattice. The method of assigning to each word the probability obtained from multiplying the lexical and the acoustic probabilities is not optimal and new paths will have to be explored.

Combining the acoustic and syntactical probabilities in the process of recognition has several problems in keeping the consistency of the grammar and in the way to weight each of the probabilities. The simple approach consisting of making the hypothesis of independence (section 3.2.2) slightly improved the results. However, future work on this point will have to investigate how to optimally integrate the acoustic probabilities of the words within the syntactic process.

One approach to the combination of both probabilities into the recognition process will be a two-pass method (lattice re-scoring). First, the coupling described in this document will be realized. Then, for each of the interpretations of the syntactic parser the total acoustic probability will be calculated, and will be used to find the sentence most probable among the syntactically correct interpretations.

2 - Using several lattices in the syntactic analysis

The most important problem realizing the coupling is that the lattice given as input to the syntactic parser might not contain the sentence said by the speaker. If this sequence is not in the lattice, it will not be possible to find it with the syntactic module, which in the considered coupling acts only as a filter and does not create new sentences.

The different experiments realized with the speech recognizer have shown that many lattices do not contain the sequence of words said by the speaker. But one important observation is that realizing experiments with different parameters over the same input spoken sentence it is possible to produce different lattices that contain all the words of the sentence said by the speaker.

This fact establishes the necessity of using different experiments over the input sentences, which carried to the coupling means that instead of using one lattice, two or more lattices of the same spoken sentence are given as input to the syntactic parser.

This alternative approach to the considered coupling can be the solution to the fact that lattices produced with specific parameters can be adequate or inadequate depending on the situation. There will be parameters that optimize the answer in noisy situations but decrease the quality of the lattice in isolated conditions. On the other hand, other parameters will inverse this performance, giving good results in calm contexts. Using several lattices for the coupling will have the same effect as if the recognition had been made using all the parameters at each step, so the final list of recognized sentences will be that of combining all the different parameters for each of the recognition steps (mixture of experts).

4.2 - Conclusions on the results of the coupling

The main conclusion of the work described in this report is that the speech recognition can be substantially improved by the integration of syntactic constraints. In addition, this improvement can be achieved without significantly increasing the overall processing time. Several future alternatives still however need to emphasize this conclusion for scenarios different from those presented in the modest context considered here.

The dependency of the coupling on the speech recognizer's outputs carries the responsibility of the efficient coupling to the speech recognizer's side. The performed experiments demonstrate that good outputs from the speech recognizer benefit more from the coupling than not sub-optimal results, therefore determining the secondary role of the syntactic parser in the recognition process. Once these outputs are optimized to fit the requirements of the syntactic parser, the coupling will reveal the real advantages of implementing the high linguistic constraints on the speech recognition.

Although the recognition is not yet being significantly improved, there are several characteristics of the final output that suggest that the syntactical information is fundamental to obtain high quality results. For instance, in the hypotheses of the speech recognizer where the recognition was slightly wrong the syntactic parser has shown to improve the solution. It is also important to point out that the total list of interpretations of the syntactical module is always going to be better than the recognizer's set of hypotheses, because the syntactic parser limits the answers of the recognizer to those that are syntactically correct.

APPENDIX A – DETAILS ABOUT EXPERIMENTS

This appendix describes in detail and the experiments that were used in our work. All experiments where realized over a Ultra Spark 1 Sun Workstation running O.S. Solaris 2.5.

The speech recognizer used for the experiments was the “Speech training and recognition unified tool” (STRUT)[Strut96]. The sub-module that produces the lattices is called Noway [Now94].

Sample sentences

A sample of 10 sentences (Table A.1) has been randomly chosen from a list of 3 200 different sentences said by different speakers [Polyphone].

Table A.1 – Description of the 10 sample sentences used in our experiments

Sentence	Text transcription
1	La Suisse recule dans tous les domaines
2	Le président du parti est aussi président du comité national
3	Un paradis terrestre possède un fleuve
4	Le plastique a renvoyé l’acier dans les limbes de l’âge de fer
5	Je n’ai jamais supporté le monde
6	Ils s’installeront en France
7	Les problèmes restent nombreux
8	Les causes de cet engouement sont variées
9	Je me sens bien avec eux
10	La roue tourne

Parameter files

The speech recognizer needs two main resources: a dictionary containing the words and their phonetic transcription and a language model (mono-grams and bi-grams). The dictionary used for the experiments had 11 064 words and the language model had 1 186 664 mono-grams and 28 861 312 bi-grams.

Description of the experiments

The definition of the most important characteristics used to produce the lattices for each experiment must be done in order to clearly establish the parameters used for each of them. It is

important not only to realize experiments with different examples, but also with different characteristics of recognition [Now94].

The experiments used in the context of this report used a constant beam of 12 and different acoustic scales (called α): 0.15, 0.20, 0.25, 0.30, 0.35 and 0.70. Each experiment is referred in the report by the value of its acoustic scale. The acoustic scale is the factor by which to scale the acoustic log likelihood. This is often known as the acoustic model/language model match factor.

APPENDIX B – SPEECH RECOGNIZER'S RESULTS

This appendix contains all the results obtained from the speech recognizer running the experiments described in appendix A. It is subdivided it in two parts, the 1-best hypotheses (appendix B1) and the recognition time for each sentence (appendix B2).

Appendix B1 - 1-best hypotheses

The name of each experiment (*1bestInNowayXX*) is followed by the ten 1-best hypotheses.. Each of the 1-best hypotheses is preceded by its corresponding number (appendix A) and are followed by the log of the probability of the hypothesis (between parenthesis). The *1bestInNowayXX* stands for the experiment where $\alpha = 0.XX$

```
::::::::::::::
1bestInNoway10
::::::::::::::
# 1 la suite recul dans tous deux  (-51.972401)
# 2 le président du parti est aussi président du comité national  (-84.573761)
# 3 un paradis terrestre aussi de huit  (-56.681129)
# 4 le parti quand la lame de l' âge de francs  (-105.872650)
# 5 nous allons sportive  (-43.593151)
# 6 il s' installe en francs  (-45.400517)
# 7 il est pas plus nombreuses  (-56.724342)
# 8 le camp de s' étend son arrivée  (-58.754246)
# 9 je ne sont bien vécu  (-31.039227)
# 10 la tour  (-25.659603)

::::::::::::::
1bestInNoway15
::::::::::::::
# 1 la suisse recul entouré de main  (-39.793842)
# 2 le président du parti tout aussi président du comité national  (-60.789272)
# 3 un paradis terrestre aussi de fruits  (-45.444828)
# 4 le parti quand elle a finalement l' âge de l' heure où  (-94.370361)
# 5 six à supporter l' homme  (-35.175446)
# 6 il s' installe en francs  (-35.470978)
# 7 le problème reste nombreuses  (-47.891846)
# 8 le camp de ces temps de façon paris  (-48.022949)
# 9 je ne sont bien vécu  (-20.851021)
# 10 la roue tourne  (-22.284855)

::::::::::::::
1bestInNoway20
::::::::::::::
# 1 la suisse recul entouré de main  (-26.034601)
# 2 le président du parti tout aussi président du comité national  (-36.709133)
# 3 un paradis terrestre aussi de fruits  (-32.942543)
# 4 le parti cas en vol à six mille vingt l' âge de l' heure où  (-73.955254)
# 5 si nous allons supporter l' homme  (-24.830992)
# 6 il s' installe en france  (-24.067322)
# 7 i problèmes restent nombreux six  (-35.954865)
# 8 le camp de ces temps de maçons paris  (-34.909580)
# 9 je ne sont bien avec eux  (-10.614474)
# 10 la route ou deux  (-15.062828)
```

```
::::::::::::::
1bestInNoway25
::::::::::::::
# 1 la suisse recul entouré de main  (-12.224090)
# 2 le président du parti tout aussi président du comité national  (-12.616423)
# 3 un paradis terrestre aussi de fruits  (-20.411814)
# 4 le parti cas en vol à six mille vingt l' âge de l' heure sous  (-53.080185)
# 5 si nous allons supporter le monde  (-13.925136)
# 6 il s' installeront en france  (-11.049776)
# 7 i problèmes restent nombreux se  (-23.659485)
# 8 le camp de ces temps de maçons paris  (-21.382879)
# 9 je ne sont bien avec eux  (0.096069)
# 10 la route ou deux  (-6.825910)

::::::::::::::
1bestInNoway30
::::::::::::::
# 1 la suisse recul de entouré de même  (1.851558)
# 2 le président du parti tout aussi président du comité national  (11.462983)
# 3 un paradis terrestre aussi de un fruit neuf  (-5.747544)
# 4 le parti terrain en vol à six mille vingt mètres l' âge de l' heure sous (-30.743086)
# 5 si nous allons supporter le monde  (-2.776604)
# 6 il s' installeront en france  (2.187006)
# 7 i problèmes restent nombreux se  (-11.249849)
# 8 le camp de ces temps de maçons paris  (-7.606792)
# 9 je ne sont bien avec eux  (10.864474)
# 10 la route ou deux  (1.408810)

::::::::::::::
1bestInNoway35
::::::::::::::
# 1 la suisse recul de entouré de même  (17.165728)
# 2 le président du parti tout si président du comité national  (35.644684)
# 3 un paradis terrestre aussi de un fruit neuf  (10.094335)
# 4 le plastique pain en voyait là si ni l' âme te l' âge de l' heure où sous (-6.847883)
# 5 si nous allons supporter le monde  (8.434304)
# 6 il s' installeront en france  (15.441489)
# 7 i propres le reste nombreuses se  (1.277707)
# 8 le camp de ces temps de maçons paris  (6.202743)
# 9 je ne sont bien avec eux  (21.661566)
# 10 la route ou deux  (9.654516)

::::::::::::::
1bestInNoway70
::::::::::::::
# 1 la suisse recul de entouré de main ni  (124.999557)
# 2 le président n' en du parti tout si président du comité national  (207.466629)
# 3 un paradis terrestre faut aussi de gain fuit ne sait f  (121.712822)
# 4 le pas ce ticket pain en vol yeux la six nuits liens me te l'âge de l'an vivre vous
ce nouveau (180.528961)
# 5 si nous allons six porte il le mon  (88.425674)
# 6 il feint à leur ont pas en france  (112.781326)
# 7 i propres les me reste non entre eux se  (96.783928)
# 8 il est quand de ces temps de manche ont paris  (105.902802)
# 9 je ne sont bien avec eux  (97.400993)
# 10 la roue que tous deux  (67.734322)
```

Appendix B2 - Recognizing timings

Following the experiment name (*RecNowayXX*) the timings to produce the lattices are given, as well as the result of the command `time`. *RecNowayXX* corresponds to experiment with $\alpha = 0.XX$.

The interpretation of the `time` command is the following:

```
 user seconds  CPU seconds  Real elapsed time  CPU % used
```

```
::::::::::::::
RecNoway10
::::::::::::::
1 403.71u 2.80s 13:39.13 49.6%
2 1241.46u 3.25s 41:54.53 49.5%
3 1196.81u 3.36s 40:07.77 49.8%
4 2052.40u 4.14s 1:08:43.49 49.8%
5 758.50u 3.17s 23:17.34 54.5%
6 472.96u 2.64s 15:55.58 49.7%
7 374.76u 2.69s 12:38.03 49.7%
8 885.07u 2.80s 29:40.87 49.8%
9 414.63u 2.58s 13:56.55 49.8%
10 355.79u 2.62s 11:58.54 49.8%

::::::::::::::
RecNoway20
::::::::::::::
1 128.65u 4.81s 2:28.84 89.6%
2 213.18u 4.35s 3:46.47 96.0%
3 193.94u 4.36s 3:23.37 97.5%
4 552.28u 4.89s 9:29.00 97.9%
5 179.42u 4.32s 3:11.56 95.9%
6 177.04u 3.79s 3:04.72 97.8%
7 117.41u 4.15s 2:04.41 97.7%
8 195.23u 4.37s 3:23.77 97.9%
9 109.83u 4.30s 1:56.83 97.6%
10 184.35u 4.36s 3:12.43 98.0%

::::::::::::::
RecNoway25
::::::::::::::
1 160.68u 2.30s 1:09.56 90.5%
2 77.64u 2.40s 1:25.49 93.6%
3 75.47u 1.89s 1:26.88 89.0%
4 192.86u 2.46s 3:23.20 96.1%
5 68.97u 1.76s 1:17.40 91.3%
6 71.75u 2.09s 1:15.75 97.4%
7 54.35u 2.02s 0:57.97 97.2%
8 74.56u 2.32s 1:17.67 98.9%
9 50.71u 2.25s 0:54.61 96.9%
10 74.22u 2.16s 1:18.21 97.6%

::::::::::::::
RecNoway30
::::::::::::::
1 56.11u 2.34s 2:00.75 48.4%
2 61.03u 2.19s 2:06.89 49.8%
3 65.83u 2.04s 2:16.11 49.8%
4 127.55u 2.25s 4:19.91 49.9%
5 57.76u 1.94s 1:59.97 49.7%
6 59.34u 2.03s 2:02.92 49.9%
7 49.24u 2.13s 1:43.19 49.7%
8 60.24u 2.20s 2:04.85 50.0%
9 47.40u 1.98s 1:38.87 49.9%
10 60.36u 1.93s 2:04.95 49.8%
```

```
::::::::::::::
RecNoway35
::::::::::::::
1 51.41u 2.03s 1:46.85 50.0%
2 53.98u 2.24s 1:52.47 49.9%
3 57.27u 2.16s 1:58.58 50.1%
4 95.09u 2.26s 3:14.96 49.9%
5 51.10u 2.05s 1:46.11 50.0%
6 52.20u 2.28s 1:48.64 50.1%
7 46.54u 2.00s 1:37.17 49.9%
8 53.28u 2.17s 1:50.90 50.0%
9 44.67u 2.18s 1:33.58 50.0%
10 52.84u 2.22s 1:49.98 50.0%

::::::::::::::
RecNoway70
::::::::::::::
1 42.99u 2.16s 1:30.80 49.7%
2 44.06u 2.14s 1:32.06 50.1%
3 43.00u 2.10s 1:30.57 49.7%
4 48.10u 2.07s 1:40.01 50.1%
5 42.69u 1.95s 1:29.17 50.0%
6 42.89u 2.14s 1:29.90 50.0%
7 41.61u 1.92s 1:26.83 50.1%
8 42.89u 1.99s 1:29.64 50.0%
9 41.28u 2.00s 1:26.85 49.8%
10 42.55u 2.06s 1:29.42 49.8%
```

APPENDIX C - WORD OCCURRENCE IN LATTICES

Following each of the sentences, the listing of each of the words in the sentence in alphabetical order is provided. For each of these words, the number of occurrences in the corresponding lattice is shown for each experiment.

::::::::::::::
Sentence 1
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
dans	7	3	1	1	1	1	1
domaines	**0**	**0**	**0**	**0**	**0**	**0**	**0**
la	4	5	3	2	3	3	2
les	12	10	9	7	8	7	0
recule	4	7	11	12	13	13	8
Suisse	17	12	8	2	1	1	1
tous	5	6	7	4	4	4	2

::::::::::::::
Sentence 2
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
aussi	3	3	3	4	4	4	2
comité	12	9	9	9	2	2	1
du	31	21	16	11	8	9	3
est	10	11	13	10	9	9	4
le	61	66	66	52	38	34	14
national	32	29	29	25	19	19	8
parti	26	20	20	15	16	14	7
président	39	45	39	33	25	25	12

::::::::::::::
Sentence 3
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
fleuve	**0**	**0**	**0**	**0**	**0**	**0**	**0**
paradis	7	7	9	13	11	11	2
possède	6	7	9	10	10	12	7
terrestre	15	13	12	10	10	12	9
un	30	31	33	31	32	34	15

::::::::::::::
Sentence 4
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
a	49	53	43	35	39	38	28
acier	**0**	**0**	**0**	**0**	**0**	**0**	**0**
dans	**0**	**0**	**0**	**0**	2	3	1
de	82	107	79	74	63	55	41
fer	**0**	**0**	**0**	**0**	**0**	**0**	**0**
l’	77	87	54	45	46	55	43
le	73	84	55	63	52	55	38
les	12	8	5	4	5	3	1
limbes	**0**	**0**	**0**	**0**	**0**	**0**	**0**
plastique	3	4	6	11	13	7	4
renvoyé	**0**	1	1	1	1	1	**0**
âge	33	18	19	20	19	20	12

::::::::::::::
Sentence 5
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
ai	**0**	**0**	**0**	**0**	**0**	**0**	**0**
jamais	**0**	**0**	**0**	**0**	**0**	**0**	**0**
je	1	3	**0**	1	**0**	**0**	**0**
le	43	35	33	35	25	21	21
monde	**0**	1	1	2	2	2	7
n’	5	4	5	9	10	8	10
supporté	2	1	1	2	1	2	1

::::::::::::::
Sentence 6
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
en	9	16	26	21	19	15	12
France	2	2	3	5	7	8	5
ils	10	16	11	4	5	5	12
installeront	7	11	19	13	10	6	**0**
s’	1	4	3	2	6	9	4

::::::::::::::
Sentence 7
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
les	1	**0**	2	2	4	5	3
nombreuses	10	10	9	9	12	3	1
problèmes	**0**	3	3	3	4	**0**	**0**
restent	**0**	7	9	6	8	7	8

::::::::::::::
Sentence 8
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
causes	3	3	4	3	1	1	**0**
cet	3	2	2	1	1	1	1
de	44	40	39	32	29	28	16
engouement	0	1	2	2	3	2	2
les	6	7	5	6	3	3	2
sont	2	4	5	4	4	5	2
variées	**0**	**0**	**0**	**0**	**0**	**0**	**0**

::::::::::::::
Sentence 9
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
avec	16	16	14	15	15	14	12
bien	16	16	18	13	16	13	3
eux	2	3	3	2	2	2	3
je	19	16	16	17	17	18	10
me	4	5	5	4	4	4	1
sens	4	5	7	6	5	4	0

::::::::::::::
Sentence 10
::::::::::::::

	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.30$	$\alpha = 0.35$	$\alpha = 0.70$
la	9	13	8	7	7	7	2
roue	**0**	7	14	10	6	6	8
tourne	**0**	2	2	**0**	**0**	**0**	**0**

APPENDIX D – THE RESULTS OF THE COUPLING

This appendix contains the listings of the results of the coupling for all experiments. The results are divided in two parts: the 1-best sentence found with the coupling without using the acoustic information of the lattice (appendix D1) and using it (appendix D2) and the sizes and timings of the operations concerning the coupling and the parser (appendix D3).

Appendix D 1

The 1-Best interpretations of the syntactic parser

bestXX refers to the experiment with $\alpha = 0.XX$

```
::::::::::::::
best10
::::::::::::::
Sentence 1: l’ Suisse recule entre le mai
Sentence 2: l’ président du parti est ainsi résident commissaire national
Sentence 3: l' qualité possède l' une
Sentence 4: une partie en elle lame le âge gros
Sentence 5: il exprime
Sentence 6: ils installent enfant
Sentence 7: il a plus nombreux
Sentence 8: les causes du fait sont parti
Sentence 9: je semble avec eux
Sentence 10: elle a tout

::::::::::::::
best15
::::::::::::::
Sentence 1: l' Suisse recule de dans de le homme
Sentence 2: les présidents du parti tous président l' commissaire nationale
Sentence 3: un huit est que possible une
Sentence 4: une plastique à envoyé a finalement large de l' rue
Sentence 5: un dix a sportive
Sentence 6: ils installent dans France
Sentence 7: il a plus récession de plus
Sentence 8: les causes des ces temps sont parti
Sentence 9: je est cent bien avec eux
Sentence 10: elle a tout

::::::::::::::
best20
::::::::::::::
Sentence 1: l' suisse recul entoure le demain
Sentence 2: l' président du parti est aussi résident du produit national
Sentence 3: un parti terrestre possède un fruit
Sentence 4: une plastique à envoyé a finalement large du heureux de nous
Sentence 5: ils avons supporté l' âme
Sentence 6: ils installent en France
Sentence 7: un dix populaire est nombreux de ce
Sentence 8: les causes de cette tendance ont parti
Sentence 9: je est cent bien avec eux
Sentence 10: la roue tourne
```

::::::::::::::
best25
::::::::::::::
Sentence 1: l’ Suisse recule entre le rude même
Sentence 2: l' président du parti est ainsi président du commissaire national
Sentence 3: un paradis terrestre possède un fruit
Sentence 4: une plastique a renvoyé le fini vingt large de rue de nous
Sentence 5: ils avons supporté la même
Sentence 6: ils se installeront dans France
Sentence 7: un dix populaire est nombreux de ce
Sentence 8: les causes de cet engouement sont paris
Sentence 9: je est sens bien avec eux
Sentence 10: la roue est ouvert

::::::::::::::
best30
::::::::::::::
Sentence 1: le Suisse recule entre le rude même
Sentence 2: l' président du parti est ainsi président du commissaire national
Sentence 3: un paradis terrestre possède un fruit
Sentence 4: l' plastique un violent a fini l' vingt large de revue de nous
Sentence 5: un dix a supporté le mot
Sentence 6: ils se installeront dans France
Sentence 7: un dix populaire est nombreux de ce
Sentence 8: les causes de cet engouement sont paris
Sentence 9: je est sens de bien avec eux
Sentence 10: la roue est ouvert

::::::::::::::
best35
::::::::::::::
Sentence 1: elle a suisse recul devant tour rude même
Sentence 2: l' président du parti est tout site résident du commissaire national
Sentence 3: un paradis terrestre possède un fruit neuf
Sentence 4: l' plastique un en violent à fini la maître las à je remise nous
Sentence 5: un six à la est supporté le mot
Sentence 6: ils se installeront dans France
Sentence 7: un dix populaire est nombreux si
Sentence 8: les causes dès cet engouement sont paris
Sentence 9: je sent de liens avec eux
Sentence 10: la roue est ouvert

Appendix D2

The 1-Best interpretations of the syntactic parser using the acoustic information

bestacousticXX refers to the experiment with $\alpha = 0.XX$

```
::::::::::::::
bestacoustic10
::::::::::::::
Sentence 1 :  l’ Suisse recule entre le homme
Sentence 2 :  l’ président du parti est ainsi résident commissaire national
Sentence 3 :  l' qualité possède ce huit
Sentence 4 :  l' plastique en elle lame le âge gros
Sentence 5 :  il exprime
Sentence 6 :  il installe dans France
Sentence 7 :  ils ont plus nombreuses
Sentence 8 :  les causes du sept sont parti
Sentence 9 :  je semble avec eux
Sentence 10:  elle a tout

::::::::::::::
bestacoustic15
::::::::::::::
Sentence 1 :  l' Suisse recule entre le est de main
Sentence 2 :  l' président du parti est ainsi président du commissaire national
Sentence 3 :  un parti terrestre possède un fruit
Sentence 4 :  l' parti à envoyé a finalement large de rue
Sentence 5 :  un dix a sportive
Sentence 6 :  ils se installeront à France
Sentence 7 :  un dix populaire est nombreux de ce
Sentence 8 :  les causes de cet engouement sont prix
Sentence 9 :  je est cent bien avec eux
Sentence 10:  elle a tout

::::::::::::::
bestacoustic20
::::::::::::::
Sentence 1 :  l' suisse recul entoure le demain
Sentence 2 :  l' président du parti est ainsi président du commissaire national
Sentence 3 :  un parti terrestre possède un fruit
Sentence 4 :  l' plastique à envoyé a finalement large de l' oeuvre pour je
Sentence 5 :  ils avons supporté l' mot
Sentence 6 :  ils se installeront dans France
Sentence 7 :  un dix populaire est nombreux de ce
Sentence 8 :  les causes de cet engouement sont paris
Sentence 9 :  je est cent bien avec eux
Sentence 10:  la route tourne

::::::::::::::
bestacoustic25
::::::::::::::
Sentence 1 :  l' suisse recul entoure vu de mains
Sentence 2 :  l' président du parti est ainsi président du commissaire national
Sentence 3 :  un paradis terrestre possède un fruit
Sentence 4 :  l' plastique a envoyé le fini vingt large de aire pour je
Sentence 5 :  ils avons supporté le même
Sentence 6 :  ils se installeront en France
Sentence 7 :  un dix populaire est nombreux de ce
Sentence 8 :  les causes de cet engouement sont paris
Sentence 9 :  je est sens bien avec eux
Sentence 10:  la roue est ouvert
```

::::::::::::::
bestacoustic30
::::::::::::::
Sentence 1 : la suite recule entre le rude même
Sentence 2 : l' président du parti est ainsi président du commissaire national
Sentence 3 : un paradis terrestre possède un fruit
Sentence 4 : l' plastique un violent a fini l' vingt large de revue de nous
Sentence 5 : un six a supporté le mot
Sentence 6 : ils se installeront en France
Sentence 7 : un dix populaire est nombreux de ce
Sentence 8 : les causes de cet engouement sont paris
Sentence 9 : je est sens de bien avec eux
Sentence 10: la roue est ouvert

::::::::::::::
bestacoustic35
::::::::::::::
Sentence 1 : elle a suisse recul devant tour rude même
Sentence 2 : l' président du parti est tout six résident du commissaire national
Sentence 3 : un paradis terrestre possède un fruit neuf
Sentence 4 : l' plastique vingt en violent à fini la maître las à je remise nous
Sentence 5 : un six à la est supporté le mot
Sentence 6 : ils se installeront en France
Sentence 7 : il est populaire est nombreux si
Sentence 8 : les causes dès cet engouement sont paris
Sentence 9 : je sent de liens avec eux
Sentence 10: la roue est ouvert

Appendix D3
Times of the total recognition process

interXX corresponds to experiment where $\alpha = 0.XX$. Afterwards each of the sentences appears with its number and its results for that experiment.

```
                  :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
                                          inter15
                  :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
::::::::::::::
Sentence 1
::::::::::::::
       Size of the table : 627.40 Ko
       Highest cell in the table: (65, 1) and cell (65, 1) contains P

GVs  a 2 272 793 interprétations
GNP  a 1 425 911 interprétations
GNmp a   102 342 interprétations
GNms a 1 857 835 interprétations
P    a 1 962 262 interprétations
GNfp a    47 914 interprétations
N+mp a    81 710 interprétations
GNfs a   454 528 interprétations
N+ms a   540 375 interprétations
GNp  a   150 256 interprétations
GNs  a 2 312 363 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.06 s
       Time to analyse the CYK table: 0.38 s
       Total time(translate + analyse+read grammar+printings,...): 0.50 s

::::::::::::::
Sentence 2
::::::::::::::
       Size of the table : 1965.05 Ko
       Highest cell in the table: (153, 1) and cell (153, 1) contains P

GVs  a 3 963 332 416 interprétations
GNP  a 3 963 332 416 interprétations
N+ms a   476 615 672 interprétations
P    a 2 200 556 844 interprétations
GNms a 3 389 359 544 interprétations
GNmp a   573 972 872 interprétations
GVp  a    10 333 808 interprétations
GNs  a 3 389 359 544 interprétations
GNp  a   573 972 872 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.14 s
       Time to analyse the CYK table: 1.51 s
       Total time(translate + analyse+read grammar+printings,...): 1.78 s

::::::::::::::
Sentence 3
::::::::::::::
       Size of the table : 709.95 Ko
       Highest cell in the table: (69, 1) and cell (69, 1) contains P

GVs  a 1 100 397 interprétations
GNP  a   804 511 interprétations
GNms a 1 016 792 interprétations
N+ms a 2 456 593 interprétations
GNmp a    95 006 interprétations
P    a   654 443 interprétations
N+mp a    62 410 interprétations
GNfs a       353 interprétations
GNs  a 1 017 145 interprétations
GNp  a    95 006 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.44 s
       Total time(translate + analyse+read grammar+printings,...): 0.60 s
```

```
::::::::::::::
Sentence 4
::::::::::::::
       Size of the table : 2403.43 Ko
       Highest cell in the table: (166, 1) and cell (166, 1) contains P

GVs  a  1 115 442 interprétations
GNP  a    933 630 interprétations
GNms a  1 026 816 interprétations
P    a 42 082 407 interprétations
N+ms a  1 590 360 interprétations
GNfs a    195 414 interprétations
GNfp a    102 348 interprétations
GNmp a    102 348 interprétations
N+fs a      1 968 interprétations
GNs  a  1 222 230 interprétations
GNp  a    204 696 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.16 s
       Time to analyse the CYK table: 2.07 s
       Total time(translate + analyse+read grammar+printings,...): 2.36 s

::::::::::::::
Sentence 5
::::::::::::::
       Size of the table : 187.38 Ko
       Highest cell in the table: (50, 1) and cell (50, 1) contains P

GVs  a   216 interprétations
GNP  a   144 interprétations
GNms a   216 interprétations
GNmp a    72 interprétations
N+ms a   256 interprétations
P    a 2 133 interprétations
N+mp a   144 interprétations
GNs  a   216 interprétations
GNp  a    72 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.06 s
       Time to analyse the CYK table: 0.08 s
       Total time(translate + analyse+read grammar+printings,...): 0.20 s

::::::::::::::
Sentence 6
::::::::::::::
       Size of the table : 355.34 Ko
       Highest cell in the table: (87, 1) and cell (87, 1) contains P

GVs  a 2 396 interprétations
GNP  a 2 264 interprétations
GNms a 1 668 interprétations
N+ms a 6 788 interprétations
GNmp a   264 interprétations
GNfs a   464 interprétations
P    a   230 interprétations
GNs  a 2 132 interprétations
GNp  a   264 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.08 s
       Time to analyse the CYK table: 0.18 s
       Total time(translate + analyse+read grammar+printings,...): 0.34 s
```

```
::::::::::::::
Sentence 7
::::::::::::::
       Size of the table : 559.55 Ko
       Highest cell in the table: (78, 1) and cell (78, 1) contains P

GVs  a 16 563 interprétations
GNP  a  8 159 interprétations
GNms a 13 440 interprétations
GNmp a  1 364 interprétations
P    a 28 986 interprétations
GNfs a  1 107 interprétations
N+ms a  3 194 interprétations
N+mp a  1 364 interprétations
GNs  a 14 547 interprétations
GNp  a  1 364 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.08 s
       Time to analyse the CYK table: 0.36 s
       Total time(translate + analyse+read grammar+printings,...): 0.49 s

::::::::::::::
Sentence 8
::::::::::::::
       Size of the table : 309.36 Ko
       Highest cell in the table: (60, 1) and cell (60, 1) contains P

P a 18 360 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.05 s
       Time to analyse the CYK table: 0.17 s
       Total time(translate + analyse+read grammar+printings,...): 0.30 s

::::::::::::::
Sentence 9
::::::::::::::
       Size of the table : 361.73 Ko
       Highest cell in the table: (74, 1) and cell (74, 1) contains P

GVs  a  1 984 interprétations
GNP  a    992 interprétations
GNfs a    992 interprétations
P    a 25 670 interprétations
N+fs a    992 interprétations
GNs  a    992 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.20 s
       Total time(translate + analyse+read grammar+printings,...): 0.34 s

::::::::::::::
Sentence 10
::::::::::::::
       Size of the table : 122.39 Ko
       Highest cell in the table: (38, 1) and cell (38, 1) contains P

GVs  a 75 interprétations
GNP  a 53 interprétations
GNmp a 18 interprétations
GNms a 69 interprétations
P    a 88 interprétations
GNfs a  3 interprétations
N+mp a 18 interprétations
N+ms a 35 interprétations
GNp  a 18 interprétations
GNs  a 72 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.03 s
       Time to analyse the CYK table: 0.05 s
       Total time(translate + analyse+read grammar+printings,...): 0.13 s
```

```
::::::::::::::
Sentence 1
::::::::::::::
       Size of the table : 585.38 Ko
       Highest cell in the table: (63, 1) and cell (63, 1) contains P

GVs  a 588 836 interprétations
GNP  a 374 907 interprétations
GNmp a  30 272 interprétations
GNms a 503 351 interprétations
P    a 521 234 interprétations
GNfp a   5 965 interprétations
N+mp a  23 847 interprétations
GNfs a  95 752 interprétations
N+ms a 124 741 interprétations
GNp  a  36 237 interprétations
GNs  a 599 103 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.06 s
       Time to analyse the CYK table: 0.36 s
       Total time(translate + analyse+read grammar+printings,...): 0.50 s

::::::::::::::
Sentence 2
::::::::::::::
       Size of the table : 1873.16 Ko
       Highest cell in the table: (149, 1) and cell (149, 1) contains P

GVs  a 2 012 715 456 interprétations
GNP  a 2 012 715 456 interprétations
N+ms a    93 262 372 interprétations
P    a   862 334 592 interprétations
GNms a 1 705 688 232 interprétations
GNmp a   307 027 224 interprétations
GVp  a     5 144 128 interprétations
GNs  a 1 705 688 232 interprétations
GNp  a   307 027 224 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.12 s
       Time to analyse the CYK table: 1.45 s
       Total time(translate + analyse+read grammar+printings,...): 1.69 s

::::::::::::::
Sentence 3
::::::::::::::
       Size of the table : 1002.66 Ko
       Highest cell in the table: (80, 1) and cell (80, 1) contains P

GVs  a 2 665 214 interprétations
GNP  a 2 030 481 interprétations
GNms a 2 559 517 interprétations
N+ms a 4 711 706 interprétations
GNmp a   337 270 interprétations
N+mp a   117 448 interprétations
P    a 4 594 041 interprétations
GNs  a 2 559 517 interprétations
GNp  a   337 270 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.09 s
       Time to analyse the CYK table: 0.65 s
       Total time(translate + analyse+read grammar+printings,...): 0.80 s
```

```
::::::::::::::
Sentence 4
::::::::::::::
       Size of the table : 2961.55 Ko
       Highest cell in the table: (142, 1) and cell (142, 1) contains P

GVs  a 1 853 736 641 interprétations
GNP  a 1 580 682 437 interprétations
GNms a 1 787 834 618 interprétations
P    a 3 326 421 643 interprétations
N+ms a 2 714 993 380 interprétations
GNfs a   309 280 647 interprétations
GNfp a   179 843 482 interprétations
GNmp a   194 867 722 interprétations
N+fs a     3 772 152 interprétations
GNs  a 2 097 115 265 interprétations
GNp  a   374 711 204 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.15 s
       Time to analyse the CYK table: 2.66 s
       Total time(translate + analyse+read grammar+printings,...): 2.94 s

::::::::::::::
Sentence 5
::::::::::::::
       Size of the table : 173.33 Ko
       Highest cell in the table: (49, 1) and cell (49, 1) contains P

GVs  a  42 interprétations
GNP  a  24 interprétations
GNms a  36 interprétations
GNmp a  12 interprétations
N+ms a  36 interprétations
P    a 984 interprétations
N+mp a  12 interprétations
GNs  a  36 interprétations
GNp  a  12 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.04 s
       Time to analyse the CYK table: 0.08 s
       Total time(translate + analyse+read grammar+printings,...): 0.20 s

::::::::::::::
Sentence 6
::::::::::::::
       Size of the table : 613.95 Ko
       Highest cell in the table: (92, 1) and cell (92, 1) contains P

GVs  a 1 034 414 interprétations
GNP  a   778 854 interprétations
GNms a   998 960 interprétations
N+ms a 1 025 856 interprétations
P    a   290 658 interprétations
GNfs a    35 454 interprétations
GNs  a 1 034 414 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.37 s
       Total time(translate + analyse+read grammar+printings,...): 0.53 s
```

```
::::::::::::::
Sentence 7
::::::::::::::
       Size of the table : 440.73 Ko
       Highest cell in the table: (72, 1) and cell (72, 1) contains P

GVs  a 1 943 interprétations
GNP  a   176 interprétations
GNms a   528 interprétations
P    a 5 873 interprétations
N+ms a   176 interprétations
GNs  a   528 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.27 s
       Total time(translate + analyse+read grammar+printings,...): 0.41 s

::::::::::::::
Sentence 8
::::::::::::::
       Size of the table : 695.36 Ko
       Highest cell in the table: (94, 1) and cell (94, 1) contains P

P a 92 322 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.10 s
       Time to analyse the CYK table: 0.49 s
       Total time(translate + analyse+read grammar+printings,...): 0.68 s

::::::::::::::
Sentence 9
::::::::::::::
       Size of the table : 467.66 Ko
       Highest cell in the table: (80, 1) and cell (80, 1) contains P

GVs  a 70 090 interprétations
GNP  a 41 942 interprétations
GNms a 68 970 interprétations
GNfs a    560 interprétations
P    a 26 740 interprétations
N+ms a 27 588 interprétations
N+fs a    560 interprétations
GNs  a 69 530 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.29 s
       Total time(translate + analyse+read grammar+printings,...): 0.44 s

::::::::::::::
Sentence 10
::::::::::::::
       Size of the table : 235.79 Ko
       Highest cell in the table: (61, 1) and cell (61, 1) contains P

GVs  a 286 interprétations
GNP  a 234 interprétations
GNmp a  40 interprétations
GNms a 252 interprétations
N+ms a 103 interprétations
P    a 273 interprétations
GNfs a  24 interprétations
N+mp a  40 interprétations
GNp  a  40 interprétations
GNs  a 276 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.06 s
       Time to analyse the CYK table: 0.11 s
       Total time(translate + analyse+read grammar+printings,...): 0.24 s
```

:::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
inter25
:::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::

```
::::::::::::::
Sentence 1
::::::::::::::
       Size of the table : 422.47 Ko
       Highest cell in the table: (59, 1) and cell (59, 1) contains P

GVs  a 287 674 interprétations
GNP  a 168 194 interprétations
GNms a 257 718 interprétations
P    a 149 501 interprétations
GNfs a  52 008 interprétations
N+ms a  79 974 interprétations
GNs  a 309 726 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.05 s
       Time to analyse the CYK table: 0.25 s
       Total time(translate + analyse+read grammar+printings,...): 0.36 s

::::::::::::::
Sentence 2
::::::::::::::
       Size of the table : 1683.59 Ko
       Highest cell in the table: (131, 1) and cell (131, 1) contains P

GVs  a 3 719 536 028 interprétations
GNP  a 4 130 371 784 interprétations
N+ms a   811 294 684 interprétations
P    a   452 405 888 interprétations
GNms a 2 441 588 436 interprétations
GNmp a   867 111 836 interprétations
GNfs a   821 671 512 interprétations
GVp  a    20 855 281 interprétations
GNs  a 3 263 259 948 interprétations
GNp  a   867 111 836 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.12 s
       Time to analyse the CYK table: 1.23 s
       Total time(translate + analyse+read grammar+printings,...): 1.46 s

::::::::::::::
Sentence 3
::::::::::::::
       Size of the table : 838.58 Ko
       Highest cell in the table: (84, 1) and cell (84, 1) contains P

GVs  a 38 213 394 interprétations
GNP  a 30 368 279 interprétations
N+ms a 74 810 974 interprétations
GNms a 28 562 890 interprétations
GNmp a  4 204 632 interprétations
N+mp a  4 204 632 interprétations
P    a 24 613 775 interprétations
GNs  a 28 562 890 interprétations
GNp  a  4 204 632 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.09 s
       Time to analyse the CYK table: 0.56 s
       Total time(translate + analyse+read grammar+printings,...): 0.71 s
```

```
::::::::::::::
Sentence 4
::::::::::::::
       Size of the table : 2844.63 Ko
       Highest cell in the table: (150, 1) and cell (150, 1) contains P

GVs  a 1 694 233 256 interprétations
GNP  a 3 817 112 212 interprétations
GNms a 4 230 218 648 interprétations
P    a 3 052 598 603 interprétations
N+ms a 1 590 047 602 interprétations
GNfp a 1 092 423 672 interprétations
GNmp a 1 200 011 064 interprétations
GNfs a 1 638 635 508 interprétations
GNs  a 1 573 886 860 interprétations
GNp  a 2 292 434 736 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.16 s
       Time to analyse the CYK table: 2.63 s
       Total time(translate + analyse+read grammar+printings,...): 2.90 s

::::::::::::::
Sentence 5
::::::::::::::
       Size of the table : 209.84 Ko
       Highest cell in the table: (53, 1) and cell (53, 1) contains P

GVs  a  34 interprétations
GNP  a  16 interprétations
GNms a  36 interprétations
GNmp a   4 interprétations
N+ms a  48 interprétations
P    a 505 interprétations
N+mp a   4 interprétations
GNs  a  36 interprétations
GNp  a   4 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.05 s
       Time to analyse the CYK table: 0.10 s
       Total time(translate + analyse+read grammar+printings,...): 0.22 s

::::::::::::::
Sentence 6
::::::::::::::
       Size of the table : 646.15 Ko
       Highest cell in the table: (81, 1) and cell (81, 1) contains P

GVs  a 4 121 159 interprétations
GNP  a 3 104 393 interprétations
GNms a 4 037 714 interprétations
N+ms a 4 058 616 interprétations
P    a 1 677 661 interprétations
GNfs a   118 214 interprétations
GNs  a 4 155 928 interprétations

               TIMING RESULTS
              ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.40 s
       Total time(translate + analyse+read grammar+printings,...): 0.54 s
```

```
::::::::::::::
Sentence 7
::::::::::::::
       Size of the table : 384.00 Ko
       Highest cell in the table: (67, 1) and cell (67, 1) contains P

GVs  a   795 interprétations
GNP  a    21 interprétations
GNms a    63 interprétations
P    a 3 201 interprétations
N+ms a    21 interprétations
GNs  a    63 interprétations

                TIMING RESULTS
               ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.23 s
       Total time(translate + analyse+read grammar+printings,...): 0.37 s

::::::::::::::
Sentence 8
::::::::::::::
       Size of the table : 802.97 Ko
       Highest cell in the table: (90, 1) and cell (90, 1) contains P

P a 167 152 interprétations

                TIMING RESULTS
               ----------------
       Time to translate the lattice into the CYK table: 0.09 s
       Time to analyse the CYK table: 0.56 s
       Total time(translate + analyse+read grammar+printings,...): 0.72 s

::::::::::::::
Sentence 9
::::::::::::::
       Size of the table : 426.11 Ko
       Highest cell in the table: (75, 1) and cell (75, 1) contains P

GVs  a 47 780 interprétations
GNP  a 23 890 interprétations
GNms a 71 280 interprétations
GNfs a    130 interprétations
N+ms a 23 760 interprétations
P    a 31 416 interprétations
N+fs a    130 interprétations
GNs  a 71 410 interprétations

                TIMING RESULTS
               ----------------
       Time to translate the lattice into the CYK table: 0.07 s
       Time to analyse the CYK table: 0.25 s
       Total time(translate + analyse+read grammar+printings,...): 0.39 s

::::::::::::::
Sentence 10
::::::::::::::
       Size of the table : 205.87 Ko
       Highest cell in the table: (57, 1) and cell (57, 1) contains P

GNP  a 134 interprétations
GNmp a  26 interprétations
GVs  a 111 interprétations
GNms a 160 interprétations
N+ms a  73 interprétations
P    a 212 interprétations
N+mp a  26 interprétations
GNp  a  26 interprétations
GNs  a 160 interprétations

                TIMING RESULTS
               ----------------
       Time to translate the lattice into the CYK table: 0.06 s
       Time to analyse the CYK table: 0.10 s
       Total time(translate + analyse+read grammar+printings,...): 0.24 s
```

[BIBLIOGRAPHY]

[Now94]
Steve Renals
Noway's manual page
University of Cambridge

[Gil97]
Daniel Gildea
Using a Stochastic Context-Free Grammar as a Language Model for Speech Recognition
Diploma project in University of Berkeley – EE225d
May 12, 1997

[Strut96]
Step by step guide to using the Speech Training and Recognition Unified Tool
http://tcts.fpms.ac.be

[ChRaj98a]
J. -C. Chappelier, M. Rajman
A Generalized CYK Algorithm for Parsing Stochastic CFG
In TAPD'98 Workshop, pp 133-137, Paris
April 1998

[ChRaj98b]
J. -C. Chappelier, M. Rajman
A practical bottom-up algorithm for on-line parsing with stochastic context-free grammars
Technical report no 98/284
July 9,1998
Département Informatique - EPFL

[Arag98]
Ramón Aragüés
Integration of syntactic constraints within a speech recognition system
Diploma project in Centro Politécnico Superior de Zaragoza
September 1998

[Polyphone]

G. Chollet, J.-L. Cochard, Contantinescu, C. Jaboulet, Ph Langlais
Swiss French Polyphone and Polyvar:
telephone speech databases to model inter--- and intra --- speaker variability
IDIAP, RR-96-01
April 1996