

Filtering Errors and Repairing Linguistic Anomalies for Spoken Dialogue Systems

David Rousset* and Ariane Halber†

Thomson-CSF

Laboratoire Central de Recherches, F-91404 Orsay Cedex, France

email: {rousset,ariane}@thomson-lcr.fr

Abstract

Our work addresses the integration of speech recognition and language processing for whole spoken dialogue systems.

To filter ill-recognized words, we design an on-line computing of word confidence scores based on the recognizer output hypothesis. To infer as much information as possible from the retained sequence of words, we propose a bottom-up syntactico-semantic robust parsing relying on a lexicalized tree grammar and on integrated repairing strategies.

1 Introduction

Spoken dialogue systems enable people to interact with computers using speech. However, a key challenge for such interfaces is to couple successfully automatic speech recognition (ASR) and natural language processing modules (NLP) given their limits.

Several collaboration modalities between ASR and NLP have been investigated. On the one hand, the speech recognition task can benefit from linguistic decision to uncover the correct utterance, see (Rayner et al., 1994) among others. On the other hand, NLP components can be robust with respect to recognition errors. The straightforward approach is to be robust by focusing only on informative words (Lamel et al., 1995; Meteer and Rohlicek, 1994). By nature, it misses some existing information in the sentence and it can be misled in case of errors on informative words. A more controlled robustness is expected with a complete linguistic analysis (Young, 1994; Hanrieder and Görz, 1995; Dowding et al., 1994). In a practical application, a dialogue module

can then handle interactive recovery, as illustrated by (Suhm, Myers, and Waibel, 1996).

The current work attempts to repair misrecognitions by mobilising available acoustic cues and by using linguistic abstraction and syntactico-semantic predictions. We present a filtering method and a repairing parsing strategy which fit in a complete system architecture.

An advantage of our approach is the use of a core module that is independent from any application.

Another advantage, for real applications, is to be aware of the expected performances of the ASR systems. Indeed, there are obstacles that prevent ASR systems to be fully reliable. In particular, the decoding algorithms enforce models which do not exploit all linguistic knowledge, mainly due to computational complexity. This hinders somehow the decoding so that the right solution is sometimes just not available.

2 System architecture

The system architecture consists in a speech recognizer, a word confidence scoring module, a robust parsing module and higher modules –around a dialogue module (Normand, Pernel, and Bacconnet, 1997).

The modules of the system articulate in a complementary way. The scoring module goal is to provide word acoustic confidence scores to help the robust parser in its task. The parsing module takes the best recognition hypothesis. It attempts to repair recognition errors and transmits a semantic representation of the sentence to the dialogue module. It relies on a lexicalized tree grammar and on integrated repairing rules. They make use of the knowledge embedded in the lexical grammar and of candidates present in the N-best hypothesis. We have studied its capacities to detect and predict missing elements and to select syntactically and semantically well-formed sentences. The robust parser needs con-

*with Lab. CLIPS IMAG, Grenoble

†with Dept. Signal, ENST Paris

confidence scoring module to point out inserted and substituted elements.

The words identified as inserted or as substituted are marked but the decision is laid upon the robust parsing or subsequent linguistic processes. Moreover, falsely rejected words can give rise to deletion repairing procedures. The robust parsing strategy applies syntactic and semantic well-formedness constraints. It derives the meaning of the sentence out of available elements and furthermore predicts the missing elements required to meet the constraints. Whatever the case, initially well-formed sentence or not, the parsing produces a usable analysis for the higher layers to perform the final interpretation or to trigger a repairing dialogue.

3 Word Errors Filtering

Inserted and substituted elements are a major problem as they are a source of misunderstanding. If not treated early on in a spoken dialogue system, they weaken the dialogue interaction, caught between running the risk of confusing the user with irrelevant interactions or annoying the user with repetitive confirmation checks.

As parsing is not always able to reject ill-recognized sentences, especially when they remain well-formed, cross-checking is required between acoustic and linguistic information. Our method is to isolate errors according to a scoring criterion and then transmit to the parsing suspected elements with the alternative acoustic candidates. They can be reactivated by the parsing if necessary, to achieve a complete analysis.

3.1 Scoring Method

A way to get a scoring criterion is to attribute a recognition confidence score to each word in the best sentence hypothesis.

A confidence score relates to the word being rightly recognized and not only to the word being acoustically close to an acoustic reference. It normally depends on the recognizer behaviour, the language to be recognized, and the application. For example (Rivlin, 1995) sees it as a normalisation of the phonemes acoustic scores and derives an exact estimation from a recognition corpus. We propose here a simple on-line computing of the word confidence score. It is not an exact measure but it has minimal knowledge requirements. The scoring relies on the observation of concurrent hypothesis of the recognizer and their associated acoustic scores. We have tested it with the N-best sentence hypothesis but lattice and word graph could be investigated further.

An initial score for each word in the best sentence is taken either from the word acoustic score or from the sentence score, distributed uniformly on the words. The score we have used here is the global sentence acoustic score. This initial word score is re-evaluated on the basis of concordances between the different recognition hypothesis. The major parameter for score estimation is the alignment between the word in the best hypothesis and the words in the other hypothesis. In our case this alignment is achieved by a dynamic programming method ¹.

For each N-best, an alignment value is defined from the words alignment. It disfavors especially the recidivist occurrences of a word candidate. Let w_i be the i^{th} word in the best hypothesis, the alignment value at rank n is:

$$Al_n(w_i) = \begin{cases} 1 & \text{when } w_i \text{ is aligned with itself} \\ -1 & \text{when } w_i \text{ is not aligned} \\ -r & \text{when } w_i \text{ is aligned for the } r^{th} \\ & \text{time with a given word} \end{cases} \quad (1)$$

The re-evaluation of a word score will derive from this word alignment value.

Each N-best gives rise to a re-evaluation of the current word score. This re-evaluation decomposes into two factors, a re-scoring potential V and a re-scoring amplitude ΔS . Let $S_n(w_i)$ be the score of the word w_i having observed N-best hypothesis up to rank n :

$$S_n(w_i) = S_{n-1}(w_i) + V_n(w_i)\Delta S_n \quad (2)$$

Where $V_n(w_i)$ is the potential for rescoring the word w_i according to hypothesis H_n -the sentence hypothesis at rank n and ΔS_n is the rescoring amplitude at rank n .

The first factor of the re-evaluation is the potential, defined in equation 3. It is based on the alignments and indicates the type of increase or decrease that a word deserves. A context effect is introduced in the potential in the form of penalties and bonus which are proportional to the direct neighbours alignment values (see equation 4), so that:

$$V_n(w_i) = Al_n(w_i) + \sum_{j \neq i} \delta Al_n(w_j, w_i) \quad (3)$$

$$\delta Al_n(w_j, w_i) = \begin{cases} \alpha^+ Al_n(w_j) & \text{if } Al_n(w_j) > 0 \\ \alpha^- Al_n(w_j) & \text{if } Al_n(w_j) < 0 \end{cases} \quad (4)$$

¹As no additional phonetic or temporal information is used to do the alignment, there might be seldom cases of bad alignment. The problem should not arise with lattice or word graph as they keep temporal information.

Where $Al_n(w_i)$ is the alignment value of word w_i between the first-best hypothesis H_1 and the N-best hypothesis H_n . $\delta Al_n(w_j, w_i)$ is the context effect of word w_j on word w_i (equation 4). Practically this is either a positive contribution if w_j is well aligned or a negative contribution if w_j is badly aligned. We consider context effect only from the immediate neighbours.

The second factor of the re-evaluation is the amplitude (cf. equation 2). The amplitude is the same for every word at a given rank. It is based on the n^{th} hypothesis score and the rank so that the amplitude decreases with the rank and with the relative score difference between H_1 and H_n . It expresses the rescaling power of hypothesis H_n and is calculated iteratively as:

$$\Delta S_n = \Delta S_{n-1} \left(1 - \frac{S(H_1) - S(H_n)}{|S(H_1)|} - \mu \right) \quad (5)$$

Where μ is a linear slope that ensures a minimal decrease. $S(H_n)$ is the global acoustic score of the hypothesis H_n .

The scoring stops in the case of the amplitude reaching zero. Fig 1 and 2 show evolution of the word score across N-best re-evaluation.

3.2 Filtering application

Once the word confidence scores are available, the filtering still needs a threshold to point out would-be errors. It is set on-line as the maximum score that different typical cases of words to be eliminated could reach. It is computed in the same time as word confidence scores. We consider the worst case score of several empirical cases independent from the two recognizer we tested. One of those cases is a word that would be not-aligned 80% of the time and always surrounded by aligned neighbours.

When the suspect words have been spotted, it remains to be decided whether they are substitutions or insertions. We distinguish them thanks to segmental cues and to local word variations between competitive hypothesis. Practically, the alignments previously calculated are scanned ; if the two bordering neighbours of a word w are once adjacent and well aligned in an hypothesis, w is marked as an insertion.

3.3 Evaluation

We have tested the word scoring module, with the incorporated filtering, on errors produced by two existing ASR systems from SRI and Cambridge University. The former, Nuance Communication recognizer system is constrained by a Context Free Grammar.

uttered:	DO YOU HAVE SOME RED ARMCHAIRS
H1:	DO YOU HAVE TWO RED COMPUTERS
H2:	DO YOU HAVE TWO RED ARMCHAIRS
H3:	DO YOU HAVE THOSE RED COMPUTERS
H4:	DO YOU HAVE THE RED COMPUTERS
H5:	DO YOU HAVE THOSE RED ARMCHAIRS
H6:	DO YOU HAVE THE RED ARMCHAIRS
H7:	DO YOU HAVE SOME RED COMPUTERS

Table 1: N-best hypothesis for the sentence “do you have some red armchairs”

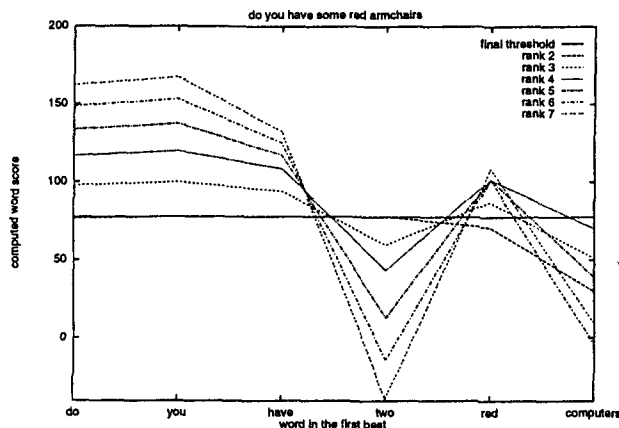


Figure 1: word scores across N-best ranks for the best hypothesis “do you have two red computers”

The latter, Abbot, uses an n-gram model (backed off trigram model) ².

The application domain is taken from the COVEN ³ project (Normand and Tromp, 1996), described on <http://chinon.thomson-csf.fr/coven/>. COVEN (Collaborative Virtual ENvironments) addresses the technical and design-level requirements of Virtual-based multi-participant collaborative activities in professional and citizen-oriented domains. Among the grounding testbed applications, an interior design application is being developed, which provides the background of the work described in this article. A typical interior design scenario deals with composition of pieces of furniture, equipment and decoration in an office room by several partici-

²The training corpus for the trigram was generated artificially by the context free grammar of the first recognizer mentioned. 15% of the testset is out of the Nuance Context Free Grammar. The sampling rate of acoustic models are 8 kHz for Nuance and 16 kHz for Abbot. The Nuance communication recognizer system exploits phonemes in context. Abbot uses a neural network to model standard phonemes.

³COVEN is a European project of the ACTS Programme (Advanced Communications Technologies and Services).

pants, within the limits of a common budget. Elements of the design are taken from a set of possible furniture, equipment and decoration objects, with variable attributes in value domains. The user may ask information to the system which provides guidance for the user decision.

The evaluation results of the speech recognizers are given with others results in table 5. Here are two examples of scoring and filtering. Figure 1 shows the evolution across seven N-best of an ill-recognized sentence score profile. At the end, the two ill-recognized words (*some* and *armchairs*) are identified as errors, they are classified as substitutions according to their type of alignment in the different N-best. The recognition hypothesis are displayed in table 1 (the recognizer is Nuance).

In the second example table 2 (from Abbot), the word *is* is inserted, but not in all N-best hypothesis. The confidence scores succeed in pointing *is* as ill-recognized, the alignment considerations will then classify it as an insertion.

uttered:	CAN YOU GIVE ME THE BUDGET
H1:	CAN YOU GIVE ME IS A BUDGET
H2:	CAN YOU GIVE ME IS THE BUDGET
H3:	CAN YOU GIVE ME A BUDGET
H4:	CAN YOU GIVE ME IT BUDGET
H5:	CAN YOU GIVE IT THE BUDGET
H6:	CAN YOU GIVE ME THE BUDGET
H7:	CAN YOU GIVE ME THESE BUDGET

Table 2: N-best hypothesis for the sentence "can you give me the budget"

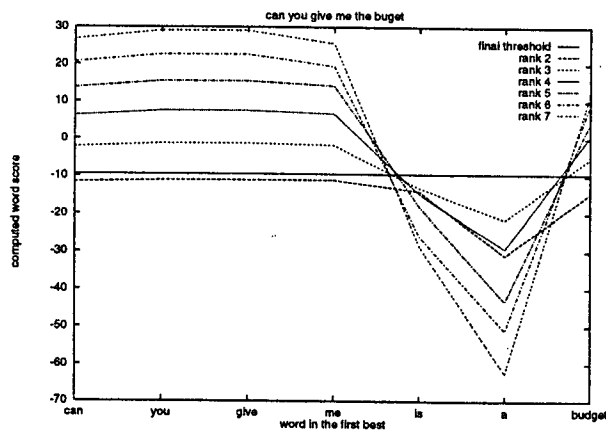


Figure 2: word scores across N-best ranks for the best hypothesis "can you give me is a budget"

First evaluation of the filtering hints that it may be a good guidance but not a sufficient criterion: some parameter settings, such as the threshold, remain problematic. Table 5 displays rather lim-

ited performances for the filtering taken alone and we suspect that even with future improvements, it will remain limited. A better filtering can only be achieved if it is informed by other knowledge sources. Performances of filtering, when coupled with the robust parsing, are indeed much more satisfactory.

4 Repairing Parsing Strategy

The aim of the robust parser presented here is to build a semantic representation needed by higher layers of the system while faced with possible ill-formed sentences. The parsing itself is led by a Lexicalized Tree Grammar (Schabes, Abeillé, and Joshi, 1988). It relies on a set of elementary trees (defined in the lexicon) which have at least one terminal symbol on its frontier, called the anchor. Trees can be combined through two simple operations : substitution⁴ and furcation (de Smedt and Kempen, 1990). Those operations are theoretically equivalent to Tree Adjoining Grammar operations. However an original property of our Lexicalized Tree Grammar is to integrate a set of semantic operations which lay down additional constraints. The parser handles semantic features, attached to the trees, and propagates them according to specific rules (Roussel, 1996). The result is a semantic representation built synchronously to the syntactical tree.

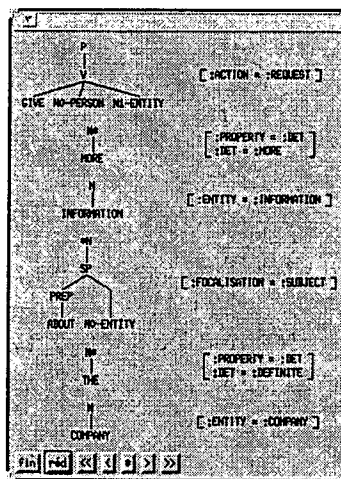


Figure 3: elementary trees and attached semantic features for the sentence "give me more information about the company"

In figures 3 and 4, the heads of the trees are standard syntactic categories, the star symbol on the

⁴It should be borne in mind that the term *substitution* when speaking of Tree Grammars has nothing to do with the term *substitution* that refers to a recognition error

right or left of the head indicates an auxiliary tree that will combine with a compatible tree ; a X* head symbol indicates a tree which combines with a node of category X on its right, a *X node combines with a node X on its left. Nodes X0, X1, or more generally Xn, are substitution sites, they are awaiting a tree whose head symbol is X. Substitution sites bear syntactic and semantic constraints on their possible substituors. Here, the semantic constraints are made visible in the node symbol (e.g. NO-PERSON means the substitutor of this node must be of category N -noun- and must possess a semantic feature :PERSON).

The parsing reveals, through linguistic anomalies, errors that wouldn't be spotted efficiently by acoustic criteria. The linguistic context allows to enrich and complete the analysis in case of an error, either detected during the parsing as a linguistic anomaly or signalled previously from confidence scores.

Actually, the robust parsing strategy articulates around a single parser, which is used iteratively according to the anomalies encountered. Three passes can each provide analysis when anomalies are detected -for correct sentences, the first pass is sufficient. Each pass will in turn modify the result of the previous pass and hand it back to the parser.

In the first pass, lexical items are first matched with their corresponding elementary tree in the lexicon. Concurrent trees for one item give rise to parallel concurrent branches of parsing, but they are taken into account in a local chart parsing.

For example the verb *want* is associated in the COVEN⁵ lexicon with two entries, one for the infinitive construction and one for the transitive construction. As preposition *to* exists in the lexicon, a sentence in which the words *want* and *to* appear calls two lexicon matching, thus two parsing branches. Figure 4 displays the trees involved. The parser will select the right matching along the syntactico-semantic operations thanks to expectations of substitution sites.

The first pass includes a first feature of robustness since unreliable words signalled by the filtering as probable substitutions are represented by an automatically generated "joker" tree. A joker tree is an overspecified tree that cumulates semantic features from different candidates whose elementary tree share the same structure⁶. Several alternative joker trees are generated when word candidates belong to different categories. Initially all semantic features in an overspecified joker tree are marked

⁵cf. section 3.2

⁶joker trees are similar to elementary tree. They can also be defined manually to fit identified cases

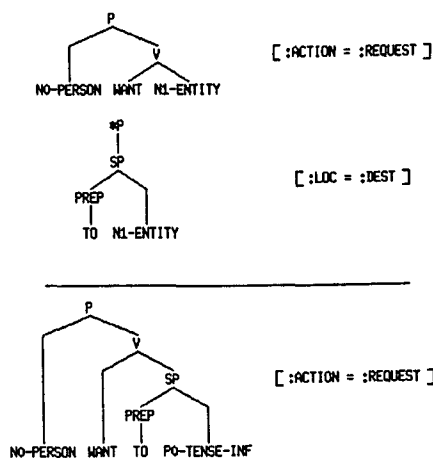


Figure 4: concurrent elementary trees and attached semantic features for the words *want to*

as uncertain, not to confuse the higher levels, then, during the parsing the semantic features mobilised for the tree operations are relieved from their uncertain status. To avoid a heavy combinatorial search, directly operations to combine two adjacent jokers are not attempted.

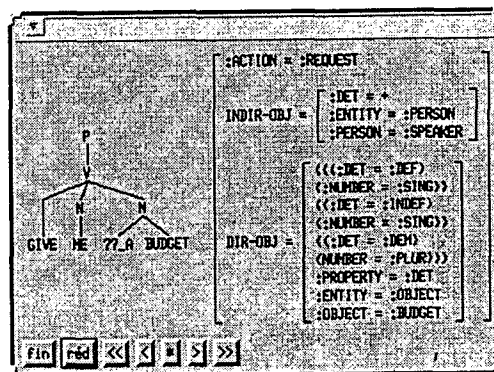


Figure 5: analysis of "give me is a budget", recovery from a substitution

Concerning insertions, the parser checks whether a local analysis is possible without a word suspected to be inserted, if so, the decision is made to eliminate the word, if not, the word is considered as substitution, and processed as described above. This is not an absolute criterion, in particular optional words falsely considered to be insertions by the filtering are not recovered.

The repairing capacities at this stage apply for instance to the case mentioned table 2. In sentence "can you give me is a budget", the word *a* is marked as a substitution (cf. 3.2). It triggers the genera-

tion of joker trees. the candidates *a, the, this, these* are represented by a single joker tree while *it*, in the 4th best hypothesis, involves a different joker tree –it is in fact its own tree, but with semantic features marked as uncertain. The branch of parsing containing this joker is eliminated on syntactic grounds, whereas the first branch of parsing turns into a complete analysis (figure 5). The word *is* which is marked as a possible insertion is confirmed in its status and definitely eliminated.

The second pass aims at recovering from would-be deleted words by re-inserting expected co-occurring words. We use knowledge about co-occurrences implicitly described in some elementary trees: elementary trees defined for more than one anchor are now being selected even if all their anchors are not present in the recognized sentence. It is however checked whether the anchors appear in given competitive recognition hypothesis at compatible positions⁷. In the following example in table 3 the recognizer (here, Abbot) has recognized the sentence *whom is this chair are too light* instead of the actual utterance *whom is this chair chosen by*.

uttered:	WHOM IS THIS CHAIR CHOSEN BY
H1:	WHOM IS THIS CHAIR ARE TOO LIGHT
H2:	WHOM IS THIS CHAIR TO AN BY
H3:	WHOM IS THIS CHAIR TO AN WALL I
H4:	WHOM IS THIS CHAIR CHOSEN IT
H5:	WHOM IS THIS CHAIR TO AN WALL MIND
H6:	WHOM IS THIS CHAIR TO AN WALL MY
H7:	WHOM IS THIS CHAIR TO AN BY A

Table 3: N-best hypothesis for the sentence “whom is this chair chosen by”

The sequence *are too light* is spotted by the filtering as a probable substitution. At pass one, the parser doesn’t succeed in putting together the elementary trees which span the whole sentence.

Now, in pass two it is observed that in the sure part of the sentence *whom is this chair*, two words *whom* and *be* are the beginning of several multi-anchor elementary trees. The aligned candidates with the sequence *are too light* allow to select only one multi-anchor tree WHOM-BE-N1-CHOSEN-BY. This provides a complete analysis.

The second pass enables a lexical recovery. The knowledge exploited here about dependencies between words at arbitrary distance can operate particularly efficiently with an n-gram driven recognizer. Indeed, the co-occurrences captured by an n-gram model suffer from a limited scope and an adjacency condition.

⁷The position is figured out from the hypothesis alignment, see section 3.1

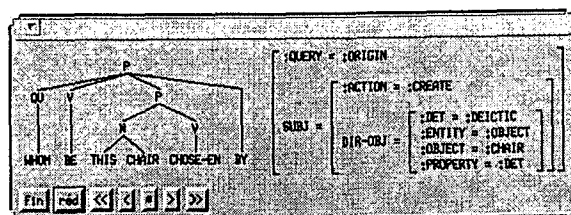


Figure 6: analysis of “whom is this chair chosen by”; the original sentence is recovered

The third pass differs from previous passes ; instead of initiating the recovery from the lexical elements at hand, it summons predictions from the grammatical expectations.

This pass is meant to detect the other errors and complete the analysis with underspecified elements.

Each anomaly revealed by the parsing has the trees around it examined to determine whether it is possible to restore a local well-formedness by inserting a tree.

Patterns of anomaly that fits in this case are defined in a compact way thanks to the general tree types used in the grammar. There are about twenty patterns, each of them is made to insert the required tree, in the form of an underspecified joker tree. This type of joker tree has a full syntactic structure but undefined semantic features: some semantic features can be added along the syntactico semantic operations.

The third pass can chose to ignore joker trees introduced in the first pass. This allows to correct irrelevant matching of joker in the first pass. This occurs when two words are substituted for a single word, or when an insertion is classified as a substitution.

uttered:	CAN YOU GIVE ME MORE INFORMATION ABOUT THE COMPANY
H1:	CAN YOU GIVE ME MORE INFORMATION THE COMPANY
H2:	CAN YOU GIVE ME MORE INFORMATION BY THE COMPANY
H3:	CAN YOU GIVE ME MORE INFORMATION THAT COMPANY
H4:	CAN YOU GIVE ME MORE INFORMATION ABOUT SECRETARY
H5:	CAN YOU GIVE ME MORE INFORMATION THE OVER COMPANY
H6:	CAN YOU GIVE ME MORE INFORMATION BOW TO THE COMPANY
H7:	CAN YOU GIVE ME MORE INFORMATION ONE THE COMPANY

Table 4: N-best hypothesis for the sentence “can you give me more information about the company”

Example table 4 stands for a typical omission recovery. The word *about* was deleted so that neither of the first passes can span the entire sentence. The third pass succeeds in inferring an analysis by inserting a generic prepositional tree that meets the syntactic and semantic expectations (see figure 7). Yet the recovery lets the information introduced by

		Nuance	Abbot	
Well Recognized Sentences	correct preserving	32 %	62 %	Correct Interpretation
	wrong filtering weakly recovered by the robust parsing	27,5 %	6,5 %	
Ill Recognized Sentences	correct filtering	11 %	8,5 %	Potentially Correct Interpretation
	partial filtering	8,5 %	5 %	
	wrong filtering. sentence rightly rejected by the parsing	0 %	5 %	Rejection
	wrong filtering. sentence falsely analysed as well formed	17 %	6,5 %	False Interpretation
	wrong filtering. sentence analysed through the robust parsing	4 %	6,5 %	
		100 %	100 %	

Table 5: results on filtering and subsequent repairing strategy

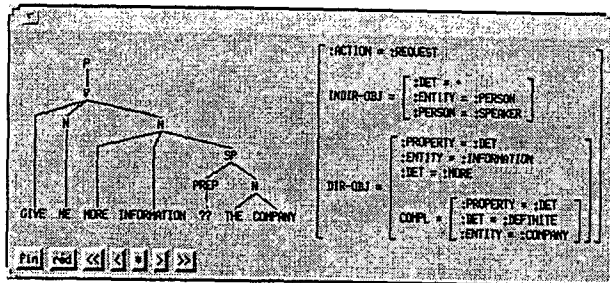


Figure 7: analysis of “give me more information the company”, recovery from an omission

the preposition undefined. However a look at compatible aligned words in the N-best hypothesis can instantiate the joker once an analysis is found.

5 Evaluation

The parser has been tested on a 200 words application⁸. The robust parsing runs in real time on an SGI Indigo-2 Impact (R4400 250 MHz). Table 6 shows the processing performances for each parsing pass.

Results on the repairing capacities according to the filtering behaviour are presented in table 5.

“Weakly recovered” means that all the information is present in the semantic representation, but part of it may be marked as uncertain with other parasite information (see figure 5 for an example). “Potentially correct interpretation” means that a valid semantic representation has been reached with

⁸The application task and the recognizer systems are described section 3.3.

1 st pass	1 st to 2 nd pass	1 st to 3 rd pass
32,5 ms.	44 ms.	113 ms.

Table 6: comparison of average cpu time required for different parsing options

some biased information. This bias might be ignored or detected by the higher level modules. The last two lines of the table distinguish between two kinds of wrongly filtered sentence: the first appear well-formed to the parser –there is no way to recover from those–, the second contain anomalies detected by the parser –there might be some way to repair or reject those ones. It can be observed that the approach is basically non-destructive toward well-recognized sentences. There is a theoretical case that would result in a loss of information: the false rejection of an optional word. But it didn’t show up. For ill-recognized sentences, at least 27% are fully recovered, for Nuance as well as for Abbot (this concerns line 3 of table 5). In both cases too, a little less than 50% appear difficult to recover, given the current filtering (last two lines of the table).

6 Conclusion

The results enlighten the repairing capacities of a couple filtering module/robust parsing module. In addition this couple presents some original desirable features that we intend to push further. First, although the parser belongs to the family of robust parsers –since it can process ill-formed sentence– it is still able to reject a subset of ill-formed sentences, which may be produced by a recognizer. Second, thanks to the lexical recovery from word candidates in the N-best hypothesis, the spoken input can be decoded further.

The scoring module can be seen as achieving not so much a filtering than a narrowing of the search space of recognition candidates. However, the approach has limitations: the parser cannot handle a large number of candidates so that the number of N-best must be limited and hence the correct candidates sometimes missed.

Moreover, spurious hypothesis generated along the passes are still hard to eliminate. This suggests the need for cross-checking with other knowledge sources, like statistical cues derived from text corpora or from recognition errors corpora.

To sum up, our work described an integration of speech recognition and language processing which is independent from a given recognition system. The basic idea was to make use of available acoustic information in order to point out a limited set of words to suspect –especially inserted words– and to exploit the potential of linguistic knowledge in order to repair the best sentence hypothesis. It can serve as a basis for many more developments.

References

- de Smedt, K and G. Kempen. 1990. Segment grammar : a formalism for incremental generation. In C. Paris et al., editor, *Natural language generation and computational linguistics*. Dordrecht, Kluwer.
- Dowding, J., R. Moore, F. Andry, and D. Moran. 1994. Interleaving syntax and semantics in an efficient bottom-up parser. In *ACL'94*.
- Hanrieder, G. and G. Görz. 1995. Robust parsing of spoken dialogue using contextual knowledge and recognition probabilities. In *ESCA Tutorial and Research Workshop on Spoken Dialogue Systems*, Denmark.
- Lamel, L., S.K. Bennacef, H. Bonneau-Maynard, S. Rosset, and J.L. Gauvain. 1995. Recent developments in spoken language systems for information retrieval. In *VIGSO'95*, Denmark.
- Meteer, M. and R. Rohlicek. 1994. Integrated techniques for phrase extraction from speech. In *Human Language Technology Workshop*, pages 228–233.
- Normand, V., D. Pernel, and B. Bacconnet. 1997. Speech-based multimodal interaction in virtual environments: Research at the Thomson-CSF corporate research laboratories. *PRESENCE: Teleoperators and Virtual Environments*. to appear as lab-review.
- Normand, V. and J. Tromp. 1996. Collaborative Virtual Environments : the COVEN project. In *FIVE'96*, Pisa, December.
- Rayner, M., D. Carter, V. Digalakis, and P. Price. 1994. Combining knowledge sources to reorder N-Best speech hypothesis lists. In *Human Language Technology Workshop*, pages 217–221.
- Rivlin, Z. 1995. Confidence measure for acoustic likelihood scores. In *Eurospeech'95*.
- Roussel, D. 1996. A lexicalized tree grammar with morphological component for spoken language processing : in french. In *Colloque Représentation et Outils pour les Bases Lexicales*, Grenoble, November.
- Schabes, Y., A. Abeillé, and A. Joshi. 1988. Parsing strategies with lexicalized grammars : Tree adjoining grammar. In *COLING'88, Budapest*, pages 578–583.
- Suhm, B., B. Myers, and A. Waibel. 1996. Interactive recovery from speech recognition errors in speech user interface. In *ICSLP'96*, pages 865–868.
- Young, S.R. 1994. Spoken dialog systems: Basic approach and overview. In *NCAI'94, Seattle*, pages 116–121.