Proc. of the Workshop on Multi–lingual Spontaneous Speech Recognition in Real Environments, Nancy, June 1996.

Prosodic Parsing of Spontaneous Speech

E. Nöth, R. Kompe, A. Kießling, H. Niemann
Universität Erlangen-Nürnberg, Lehrstuhl für Mustererkennung (Informatik 5),
Martensstr. 3, 91058 Erlangen, Germany
A. Batliner
L.–M. Universität, Institut für Deutsche Philologie,
Schellingstr. 3, 80799 München, Germany
S. Schachtl, T. Ruland, U. Block
Siemens AG, Otto-Hahn-Ring 6, 81730 München, Germany

November 22, 1996

Abstract

In automatic speech understanding, the division of continuously running speech into syntactic chunks is a great problem. Syntactic boundaries are often marked by prosodic means. We use syntactic boundaries for disambiguation during parsing. For the training of statistic models for prosodic boundaries large databases are necessary. For the German VERBMOBIL project (automatic speech–to–speech translation), we developed a labeling scheme for syntactic–prosodic boundaries. Two main types of boundaries (major syntactic boundaries and syntactically ambiguous boundaries) and some other special boundaries are labeled for a large VERBMOBIL spontaneous speech corpus. We compare the results of classifiers (multi–layer perceptrons and language models) trained on these syntactic–prosodic boundary labels with classifiers trained on perceptual–prosodic and pure syntactic labels. Recognition rates of up to 96% were achieved. We show that the boundary scores computed by these classifiers can successfully be integrated into the syntactic parsing of word graphs and currently improve the parse time by 87% and reduce the number of parse trees by 94%. This is achieved by introducing a special **P**rosodic **C**lause **B**oundary **S**ymbol PSCB into our grammar and guiding the search for the best word chain by the prosodic boundary scores. The turns that we need to parse are up to 90 seconds long and frequently contain sequences of partial sentence equivalents due to restarts, ellipsis, etc.

1 INTRODUCTION

Prosody is used to mark boundaries while speaking. This structures the utterance and helps the listener to understand and disambiguate the meaning. To our knowledge so far nobody has really integrated information about prosodic phrase boundaries into a complete automatic speech understanding system. This paper presents the syntactic analysis of word hypotheses graphs using prosodic

[1]This work was partly funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the VERBMOBIL Project under Grant 01 IV 101 AO and funded under Grants 01 IV 102 F/4 and 01 IV 102 H/0. The responsibility for the contents lies with the authors. We would like to thank Thomas Kemp (University of Karlsruhe) for providing us with the word graphs used in the experiments described in this paper.

clause boundary information. The research is carried out in the speech–to–speech translation project VERBMOBIL [16]. The influence of prosody can already be evaluated in an end–to–end system evaluation. Here we will restrict ourselves to show the influence of prosody on parsing. The domain of VERBMOBIL is appointment scheduling, i.e. two persons try to fix a meeting date, time, and place. We currently look at German utterances to be translated into English.

A corpus analysis of VERBMOBIL data, which were collected in simulated human–human dialogs, showed that about 70 % of the utterances contain more than a single sentence [15]. About 25 % of the utterances are longer than 10 seconds. The use of prosody in parsing is crucial for two reasons:

1. To ensure that most of the words that were spoken are recognized, a large word hypotheses graph has to be generated. Currently, word hypotheses graphs of about 12 hypotheses per spoken word are generated. Finding the correct (or approximately correct) path through a word hypotheses graph is thus an enormous search problem that needs to use all knowledge sources. We will show that prosody can help significantly for this problem.

2. Spontaneous speech contains many elliptical constructions. So even if the spoken word sequence has been recovered by word recognition correctly, there still might be many different parses possible, especially with longer utterances. Consider the following two of the at least 36 different syntactic readings for a word sequence taken from the VERBMOBIL corpus
   "*Ja zur Not. Geht's auch am Samstag?*" vs.
   "*Ja zur Not geht's auch am Samstag.*"
   The appropriate English translations are
   "*O.K., if necessary. Is Saturday possible as well?*" vs.
   "*Well, if necessary, Saturday is possible as well.*"
   In this example only the sentence boundary disambiguates between the two different semantic meanings and pragmatic interpretations. We will show that by using prosody the number of possible parses for the best interpretation is reduced.

We use prosody only to guide the search for the best syntactic parse through the word graph; no hard decisions are made. Partial parses are ranked in an agenda according to a score which takes into account the prosodic probability for a clause boundary. At each step of the search the best partial parse is extended. So the main use of prosodic information will be to speed up the search for the best complete parse. However in a system with limited resources (i.e. the syntax has to produce a parse after $n\times$utterance length or it will receive a time out signal), this speed up will also increase the recognition rate of the syntax module.

In this paper we first present the prosodic boundary markers used for the training of the classifiers (Section 2). Then the speech data (Section 3) and the prosodic classifiers (Section 4) are described. Section 5 shows how the grammar and the search algorithm used during parsing were extended in order to utilize the prosodic clause boundary information. Experimental results are given in Section 6.

2 PROSODIC SYNTACTIC BOUNDARY MARKERS — THE M–LABEL SYSTEM

It is well known that there is a high correlation but no one–to–one correspondence between syntactic boundaries and prosodic phrase boundaries. We developed a prosodic syntactic labeling scheme for German that

- provides a coarse labeling of syntactic boundaries

class	label	context, example
M3	M3S	main/subordinate clause: vielleicht stelle ich mich kurz vorher noch vor M3S <Atmung> mein Name ist Lerch *perhaps I should first introduce myself* M3S *<breathing> my name is Lerch*
	M3P	non–sentential free element/phrase, elliptic sentence: <Atmung> guten Tag M3P Herr Meier *<breathing> hello* M3P *Mr. Meier*
	M3E	extraposition: da hab' ich ein Seminar M3E den ganzen Tag M3S <Atmung> *there I have a seminar* M3E *the entire day* M3S *<breathing>*
	M3I	embedded sentence/phrase: eventuell M3I wenn Sie noch mehr Zeit haben M3I <Atmung> 'n bißchen länger *possibly* M3I *if you've got even more time* M3I *a bit longer*
	M3T	pre–/post–sentential particle with pause: gut M3T <Pause> okay *fine <pause>* M3T *okay*
MU	M3D	pre–/post–sentential particle without pause: <Atmung> also M3D dienstags paßt es Ihnen M3D ja M3S <Atmung> *then* M3D *Tuesday will suit you* M3D *isn't it / after all* M3S
	M3A	syntactically ambiguous: würde ich vorschlagen M3A vielleicht M3A im Dezember M3A noch mal M3A dann *I would propose* M3A *possibly* M3A *in December* M3A *again* M3A *then*
M0	M2I	constituent, prosodically marked: wie sähe es denn M2I bei Ihnen M2I Anfang November aus *will it be possible* M2I *for you* M2I *early in November*
	M1I	constituent, prosodically not marked: M3S hätten Sie da M1I 'ne Idee M3S M3S *have you got* M1I *any idea* M3S
	M0I	every other word (default)

Table 1: The M–labels, with a typical example taken from the VERBMOBIL corpus.

- can be done fast and fairly reliable (it is based solely on the transliteration of the utterance without taking the speech signal or the context utterances into account)
- takes prosodic knowledge into account, i.e. syntactic boundaries are marked differently depending on whether they are likely to be marked prosodically
- takes typical spontaneous speech phenomena into account

The labeling scheme is described in [2, 3]. Currently we distinduish 10 labels, which are grouped into three major classes. Table 1 shows the labels and their corresponding class:

- M3: prosodic–syntactic clause boundary
- M0: no clause boundary
- MU: undefined, i.e. M3 or M0 cannot be assigned to this word boundary without context knowledge and/or perceptual analysis.

In [2] we compared these labels with purely prosodic labels (B–labels)[1] [11], and precise syntactic labels (S–labels) [5]. This comparison showed that there is a high agreement between these labels and, hence, justifies our rather coarse labeling scheme. The advantage of the M–labels is that a high number of labeled data can be produced within a short time, because they are not very precise and they do not rely on perceptual evaluations. Meanwhile, there are 7,286 turns (about 150,000 words) labeled with the Ms, which took only a few months. In about the same amount of time only 648 turns were labeled with the precise S–labels, and providing B–labels for so far 861 turns took even more time.

3 SPEECH DATABASE

For the VERBMOBIL project a large database is currently being collected. It contains German–German, English–English, Japanese–Japanese, and German–English appointment scheduling dialogs. Here we only report on the German–German data. The data are transcribed according to [8]. Currently about 637 German–German dialogs are available.

For the classification experiments in Section 4 we used 3 dialogs for testing (64 turns of 3 male and 3 female speakers, 1513 words, 12 minutes in total). For the training of the multi–layer perceptron (MLP) all the available data labeled with the B–labels were used (797 turns) except for the test set; for the language model (LM), trained using the M labels, 6297 turns were used.

For the parsing experiments in Section 5 we chose 16 dialogs with 315 turns. These dialogs had been selected by the groups working on semantic processing within VERBMOBIL, since they contain semantic phenomena that should be covered by the VERBMOBIL demonstrator. Part of these turns were contained in training data of the MLP. 41 turns were not used for the experiments, because they consisted only of short elliptic utterances like time–of–day expressions. Parsing of these turns is trivial and does therefore not give much insight in the usefulness of prosody with respect to parsing. For these turns, word graphs were provided by University of Karlsruhe[2]. The word graphs contained 12.2 hypotheses per spoken word. The word accuracy, i.e. the lowest accuracy of any of the paths contained in the graph, was 91.3%. 128 word graphs were correct, i.e. they contained the spoken word chain.

4 AUTOMATIC BOUNDARY CLASSIFICATION

We will now compare classification results obtained with a multi–layer perceptron (MLP), a stochastic language model (LM), and a combination of both classifiers. The MLP serves as an acoustic–prosodic classifier getting acoustic and few lexical features as its input. The LM estimates probabilities for boundaries given a few words in the context of the word. With these classifiers for each of the words in a word chain or in a word graph a probability for a clause boundary being after the word is computed.

The computation of the acoustic–prosodic features is based on an automatic time alignment of the phoneme sequence corresponding to the spoken or recognized words. For the boundary classification experiments we only use the aligned spoken words thus simulating 100% word recognition. For each word a vector of prosodic features is computed automatically from the speech signal. For the word–final syllable, the entire word and currently ±2 syllables and ±2 words in the context the following features are considered (a total of 276):

[1] In the following we use B3 for a word boundary, which is perceived as a major prosodic boundary.

[2] We would like to thank Thomas Kemp, who provided us with these word graphs using the word recognizer described in [17].

cases	B3 vs. ¬B3 165 vs. 1284	M3 vs. M0 190 vs. 1259
MLP	87/87	87/83
LM_M	92/85	95/86
$MLP+LM_M$	94/89	96/89

Table 2: Percentage of correct classified word boundaries for different combinations of classifiers: total vs. class–wise average

- duration (+/– normalized),
- normalized F0 minimum, maximum, onset, and offset values, and their resp. relative positions on the time axis;
- energy, minimum and maximum values, and their resp. relative positions on the time axis;
- linear regression coefficients for F0 and energy contours;
- length of the pause before and after the word;
- flags indicating whether the syllable carries a lexical word accent or whether it is in a word final position.

The feature set is described in more detail in [7]. One MLP was trained to recognize the B–labels based on the features and data as described above. In order to balance for the a priori probabilities of the different classes, during training the MLP was presented with an equal number of feature vectors from each class. For the experiments, MLPs with 40/20 nodes in the first/second hidden layer showed best results.

Trigram language models (LM) were additionally used for the classification of boundaries. They model word chains where the M3–boundaries have been inserted. This method as well as the combination of LM and MLP scores is described in more detail in [9].

In Table 2, we compare the results of different classifiers for the two main classes boundary vs. not–boundary deterimned using two different types of reference boundaries: B and M. In the latter, the 'undefined' boundaries MU are not taken into account. The first number shows the overall recognition rate, the second is the average of the class–wise recognition rates. The recognition rates take all word boundaries except the end of turns into account; the latter can be classified in a trivial way.

It can be noticed that, roughly, the results get better from top left to bottom right. Best results can be achieved with a combination of the MLP with the LM no matter whether the perceptual B or the syntactic–prosodic M labels serve as reference. The LM alone is already very good; we have, however, to consider that it cannot be applied to the 'undefined' classes MU, which are of course very important for a correct syntactic/semantic processing and which account for about 4% of all word boundaries and for 23% of all non–M0 boundaries. Especially for these cases, we need a classifier trained with perceptual–prosodic labels. Note however, that even on the M3/M0–task the combination of the two classifiers, MLP+LM, shows slightly better results than the LM alone.

Due to the different a priori probabilities, the boundaries are recognized worse than the non–boundaries with the LMs (e.g., 80.8% for M3 vs. 97.7% for M0 for the MLP+LM classifier); this causes the lower average of the class–wise recognition rates compared to the overall recognition rates. It is of course possible to adapt the classification to various demands, e.g., in order to get better recognition rates for the boundaries if more false alarms can be tolerated.

(rule1)	`input`	$\rightarrow$	`phrase`	`input .`
(rule2)	`phrase`	$\rightarrow$	`s`	`PSCB .`
(rule3)	`phrase`	$\rightarrow$	`s_ell`	`PSCB .`
(rule4)	`phrase`	$\rightarrow$	`np`	`PSCB .`
(rule5)	`phrase`	$\rightarrow$	`excl`	`PSCB .`
(rule6)	`phrase`	$\rightarrow$	`excl .`	

Table 3: Grammar 1 for multiple phrase utterances

In the following section word graphs are prosodically scored using these classifiers. In this case, for each of the word hypotheses contained in the graph the probability for a clause boundary following this word is computed. The computation of the acoustic features as well as of the LM score is based on ± 2 context words. In the case of the word graphs, the best scored word hypotheses being in the context of a word hypothesis are used. This approach is sub–optimal, but we could show in [9], that recognition rate does not decrease very much when classifying word graphs instead of the spoken word chain.

5 GRAMMAR AND PARSER

In VERBMOBIL two alternative syntax–modules exist. Here we describe the interaction of prosody with the syntax–module developed by Siemens (Munich). For the interaction with the module developed by IBM (Heidelberg) cf. [1].

In the module described here, we use a **T**race and **U**nification **G**rammar (TUG) [4] and a modification of the parsing algorithm of Tomita [14]. The basis of a TUG is a context free grammar augmented with PATR-II-style feature equations. The Tomita parser uses a graph-structured stack as central data structure [**?**, 13]. After processing word w_i the top nodes of this stack keep track of all partial derivations for $w_1 ... w_i$. In [12], a parsing-scheme for word graphs is presented using this parser. It combines different knowledge sources when searching the word graph for the spoken utterance: a TUG, a statistical trigram or bigram model and the score of the acoustic component. In the work described here we added another knowledge source: the score for clause boundaries computed as indicated in Section 4.

In order to make use of the prosodic information, the grammar had to be slightly modified. The best results were achieved by a grammar that neatly designed the occurrence of `PSCB`s between the multiple phrases of the utterance:

A CF-grammar for spontaneous speech has to allow for a variety of possible input phrases following each other in a single utterance, cf. (rule1) in Table 3. Among those count normal sentences, (rule2), sentences with topic ellipsis, (rule3), elliptical phrases like PPs or NPs, (rule4) or presentential particle phrases, (rule5) and (rule6). Those phrases were classified as to whether they require an *obligatory* or *optional* `PSCB` behind them. The grammar fragment in Table 3 says that the phrases `s`, `s-ell` and `np` require an obligatory `PSCB` behind them, whereas `excl`(amative) may also attach immediately to the succeding phrase (rule 6).

The segmentation of utterances according to a grammar like in Table 3 is of relevance to the text understanding components that follow the syntactic analysis, cf.:

(1) Result for lattice VM1/N011K/NHW3K002.A16:
`[ja,also,bei,mir,geht,prinzipiell,jeder,Montag,und,jeder, Donnerstag,PSCB]`
Well, as far as I'm concerned, in principle every Monday or Thursday is possible.

(rule 7)	`input`	→	`phrase , PSB , input .`
(rule 8)	`phrase`	→	`s .`
(rule 8)	`phrase`	→	`s_ell .`
(rule 9)	`phrase`	→	`np.`
(rule 10)	`phrase`	→	`excl.`

Table 4: Grammar 2 for multiple phrase utterances

(2) Result for lattice VM4/G275A/G275A002.B16:
`[ja,PSCB,das,pa"st,mir,Dienstag,PSCB,ist,der,f"unfzehnte,PSCB]`
Yes. This Tuesday, that suits me. That is the fifteenth.

Those two examples differ w.r.t. the attachment of the exclamative `ja`. In the first example it is followed immediately by a sentence (rule6), whereas in the second it is separated by a `PSCB` from the following sentence (rule5). Semantic analysis or dialog can make use of these different rules. The exclamative in example(1) p.ex. might be identified as introduction, in example(2) it might be interpreted as affirmation. The occurrence of the second PSCB in example(2) is not so fortunate. Here the `PSCB` divides the intended subject `Dienstag` from its matrix clause `ist der f"unfzehnte`. A hesitation in the input that did not get detected as false alarm might be responsible for this. However (2) is a syntactically correct segmentation since a grammar for spoken language has to allow for topic ellipsis and the phrase `ist der f"unfzehnte` constitutes a correct sentence according to (rule 3). The grammar therefore retrieves the somewhat clumsy interpretation for this lattice as indicated by the English translation. Hence, we also tested with a grammar that *obligatorily* required a `PSCB` behind every input phrase, see Table 4.

When searching the word graph, partial sentence hypotheses are organized as a tree. A graph-structured stack of the Tomita parser is associated with each node. In the search an agenda of score–ranked orders to extend a partial sentence hypothesis ($\text{hypo}_i = \text{hypo}(\text{w}_1,...,\text{w}_i)$) by a word w_{i+1} or by the `PSCB` symbol, respectively, is processed: The best entry is taken; if the associated graph–structured stack of the parser can be extended by w_{i+1} or by `PSCB`, respectively, new orders are inserted in the agenda for combining the extended hypothesis hypo_{i+1} with the words, which then follow in the graph, and, furthermore, the hypothesis hypo_{i+1} is extended by the `PSCB` symbol. Otherwise, no entries will be inserted. Thus, the parser makes hard decisions and rejects hypotheses which are ungrammatical.

The acoustic, prosodic and trigram knowledge sources deliver scores which are combined to give the score for an entry of the agenda. In the case the hypothesis $hypo_i$ is extended by a word w_{i+1} the score of the resulting hypothesis is

$$\begin{aligned} score(hypo_{i+1}) \;=\; & score(hypo_i) \\ & + acoustic\_score(w_{i+1}) + \alpha \cdot trigram\_score(w_{i-1}, w_i, w_{i+1}) \\ & + \beta \cdot prosodic\_score(\neg \mathtt{PSCB}) \\ & + \text{`}score\ of\ optimal\ continuation' \ . \end{aligned}$$

If $hypo_i$ is extended by the `PSCB` symbol, the score of $hypo_{i+1}$ is given by

$$score(hypo_{i+1}) \;=\; score(hypo_i)$$

	analysis with PSCBs:	analysis without PSCBs:
# successful analyses	178	165
average # of syntactic readings	8.2	128.2
average parse time (secs)	4.9	38.4

Table 5: Parsing statistics for 274 word graphs

$$+\beta \cdot prosodic\_score(\texttt{PSCB})$$
$$+\text{`}score\ of\ optimal\ continuation' \ .$$

The weights α and β are determined heuristically. Prior to parsing a Viterbi–like backward pass approximates the scores of optimal continuations of partial sentence hypotheses (A^*–search). After a certain time has elapsed, the search is abandoned. With these scoring functions, hard decisions about the positions of clause boundaries are only made by the grammar but not by the prosody module. If the grammar rules are ambiguous given a specific hypothesis $hypo_i$, the prosodic score guides the search by ranking the agenda.

6 EXPERIMENTAL RESULTS

In experiments using a preliminary version of the sub–grammars for the individual types of phrases, we compared the two different grammars explained in Section ??. With the grammar shown in Table 3 149 word graphs could successfully be analyzed; with the one given in Table 4 only 79 word graphs were analyzed. This indicates that often the prosody module computes a high score for ¬PSCB after exclamatives so that parsing fails if a PSCB is obligatorily required as in the grammar of Table 4.

With an improved version of the grammar for the individual phrases, we repeated the experiments using the grammar of Table 3 and compared them with the parsing results using a grammar *without* PSCBs. For the latter, we took the category PSCB out of the grammar and allowed all input phrases to adjoin recursively to each other. The graphs were parsed without taking notice of the prosodic PSCB information contained in the lattice. An immense increase of readings and decrease in efficiency manifested itself, cf. Table 5.

The statistics show that on the average the number of readings decreases by 93% when prosodic information is used, and the parse time drops by 87%. If the lattice parser does not pay attention to the information on possible PSCBs, the grammar has to determine by itself where the phrase boundaries in the utterance might be. It may rely only on the coherence and completeness restrictions of the verbs that occur somewhere in the utterance. These restrictions are furthermore softened by topic ellipsis, etc. Any simple utterance like *Er kommt morgen.* results therefore in a lot of possible segmentations, see Table 6.

[er,kommt,morgen]	*He comes tomorrow.*
[er],[kommt,morgen]	*He? Comes tomorrow!*
[er kommt],[morgen]	*He comes. Tomorrow!*
[er],[kommt],[morgen]	*He? Comes! Tomorrow.*

Table 6: Syntactically possible segmentations

An even more serious problem showed itself w.r.t. the treatment of empty categories. The grammar uses empty categories in order to deal with verb movement and topicalisation in German. The binding of those empty categories has to be checked inside a single input phrase, i.e. the main sentence. No movement across phrase boundaries is allowed. Now, whenever a `PSCB` signals the occurrence of a boundary, the parser checks whether all binding conditions are satisfied and accepts or rejects the path that was found so far. This mechanism works efficiently in the case prosodic information was used. For the grammar without `PSCB`s, no signal, where to check the binding restrictions, is available. Therefore, the uncertainty about segmentation of multiple phrase utterances led to indefinite parsing time for some of the lattices in the corpus. Those lattices were analyzed correctly with `PSCB`s.

7 CONCLUSION

In this paper we showed that prosodic clause boundary information can reduce the parse time of word graphs computed for spontaneous speech by 87%. The number of parse trees of the resulting analyses decreases by 94%. This is especially due to the high number of elliptic and interrupted phrases contained in spontaneous speech, which cause that the position of clause boundaries is highly ambiguous. Apart from differences in the particular technical solutions to sub–problems the main difference of our approach with respect to the prosodic parse–rescoring described in [10] lies in the fact that we first compute prosodic scores based on the word hypotheses generated by the word recognizer. This allows the use of prosodic information within the parsing process and thus not only reduces the number of readings but also the parse time. In a speech understanding system, the speed–up is as important as the reduction of the number of parse trees.

In the future, we intend to use prosodic boundary information for resolving other types of ambiguities such as the attachment of prepositional phrases, of appositions and of adverbials. Especially the attachment of prepositional phrases is rather ambiguous without information about phrase boundaries; e.g. *"I saw the man with a telescope"*, or *"I want to take the train to Munich"*.

References

[1] A. Batliner, A. Feldhaus, S. Geißler, T. Kiss, R. Kompe, and E. Nöth. Prosody, Empty Categories and Parsing — A Success Story. In ICSLP 96 [6], pages 1169–1172.

[2] A. Batliner, R. Kompe, A. Kießling, M. Mast, and E. Nöth. All about Ms and Is, not to forget As, and a comparison with Bs and Ss and Ds. Towards a syntactic–prosodic labeling system for large spontaneous speech data bases. Verbmobil Memo 102, Februar 1996.

[3] A. Batliner, R. Kompe, A. Kießling, H. Niemann, and E. Nöth. Syntactic–prosodic Labelling of Large Spontaneous Speech Data–bases. In ICSLP 96 [6], pages 1720–1723.

[4] H.U. Block and S. Schachtl. Trace & Unification Grammar. In *Proc. of the Int. Conf. on Computational Linguistics*, volume 1, pages 87–93, Nantes, July 1992.

[5] A. Feldhaus and T. Kiss. Kategoriale Etikettierung der Karlsruher Dialoge. Verbmobil Memo 94, Oktober 1995.

[6] *Int. Conf. on Spoken Language Processing*, Philadelphia, October 1996.

[7] A. Kießling. Extraktion und Klassifikation prosodischer Merkmale in der automatischen Sprachverarbeitung. Dissertation. Technische Fakultät der Universität Erlangen–Nürnberg, 1996.

[8] K. Kohler, G. Lex, M. Pätzold, M. Scheffers, A. Simpson, and W. Thon. Handbuch zur Datenaufnahme und Transliteration in TP14 von Verbmobil, V3.0. Verbmobil Technisches–Dokument 11, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel, Kiel, September 1994.

[9] R. Kompe, A. Kießling, H. Niemann, E. Nöth, E.G. Schukat-Talamazzini, A. Zottmann, and A. Batliner. Prosodic Scoring of Word Hypotheses Graphs. In *Proc. European Conf. on Speech Communication and Technology*, volume 2, pages 1333–1336, Madrid, September 1995.

[10] M. Ostendorf, C.W. Wightman, and N.M. Veilleux. Parse Scoring with Prosodic Information: an Analysis/Synthesis approach. *Computer Speech & Language*, 7(3):193–210, 1993.

[11] M. Reyelt and A. Batliner. Ein Inventar prosodischer Etiketten für Verbmobil. Verbmobil Memo 33, Juli 1994.

[12] L.A. Schmid. Parsing Word Graphs Using a Linguistic Grammar and a Statistical Language Model. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, volume 2, pages 41–44, Adelaide, 1994.

[13] N. Sikkel. *Parsing Schemata.* CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, 1993.

[14] M. Tomita. *Efficient Parsing for Natural Language: A Fast Algorithm for Practical Systems.* Kluwer Academic Publishers, 1986.

[15] H. Tropf. Spontansprachliche syntaktische Phänomene: Analyse eines Korpus aus der Domäne “Terminabsprache”. Technical report, Siemens AG, ZFE ST SN 54, München, April 1994.

[16] W. Wahlster. Verbmobil — Translation of Face–To–Face Dialogs. In *Proc. European Conf. on Speech Communication and Technology*, volume “Opening and Plenary Sessions”, pages 29–38, Berlin, September 1993.

[17] A. Waibel, M. Finke, D. Gates, M. Gavalda, T. Kemp, A. Lavie, L. Levin, M. Maier, L. Mayfield, A. McNair, K. Shima, T. Sloboda, M. Woszczyna, T. Zeppenfeld, and P. Zhan. JANUS–II — Translation of Spontaneous Conversational Speech. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, volume 1, pages 409–412, Atlanta, 1996.