BMBF

Verbmobil
Verbundvorhaben

On the use of Prosody for Semantic Disambiguation in VERBMOBIL

Johan Bos
Anton Batliner
Ralf Kompe

Universität des Saarlandes
Universität München
Universität Erlangen–Nürnberg

Vm Memo 82
July 1995

July 1995

Johan Bos
Anton Batliner
Ralf Kompe

Computerlinguistik, Universität des Saarlandes, 66041 Saarbrücken
Phonetik, Universität München
Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen –Nürnberg

Tel.: (0681) 302 - 4344
Fax: (0681) 302 - 4351
e-mail: bos@coli.uni-sb.de
Anton.Batliner@phonetik.uni-muenchen.d400.de
kompe@informatik.uni-erlangen.de@

Gehört zum Antragsabschnitt: 3.12, 6.4, 8.1, 8.3, 8.7

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 R gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.

1 Introduction

The Verbmobil project aims to build a machine translation system for dialogues, with spontaneous speech as input (German and Japanese) as well as output (only English). This paper deals with the use of prosodic information to specify the semantic representation, in order to perform better and smoother translations[1]. To demonstrate this need and the way interfaces between semantics and prosody are established, we choose the phenomenon of *discourse particles*.

Although most discourse particles have a minimal contribution to the meaning of a sentence, their use is quite restricted with respect to the ongoing dialogue. Moreover, they are too much language specific to be translated directly: even for closely related languages like German and English, a huge number of differences exist. In a machine translation system like Verbmobil, one could undertake the "minimal" approach and just leave them out of the translation and save a lot of effort. The translation however will be very poor, and incoherent with respect to the ongoing dialogue. The approach we are in favour of is to analyse the precise function of discourse particles. Apart from syntactic information, prosodic information disambiguates the meaning of particles as well. This paper describes how the use of focus sensitive particles can be understood using prosodic information, from a semantic perspective that is.

As prosody effects the intended use of a particle, the translation of German into English, as it happens to be the case in Verbmobil, is straightforwardly influenced due to this fact. Consider for example:

(1) Lassen wir uns dann noch einen Termin ausmachen

Of interest here is the particle *noch*. If the speaker does not stress *noch*, this German utterance could be translated as: "then let us make an appointment". But the situation is different when the speaker stresses *noch*. A good translation then would be: "then let us make *another* appointment". The use of stress on *noch* has no effect on the meaning of the utterance, but it changes its *presuppositions*. A stressed *noch* in (1) presupposes that the speaker and hearer already made an appointment (just) before (1) was uttered. In the English translation, this presupposition is syntactically realised with the article *another*.

This is just one example where prosody effects the interpretation of an utterance. In this memo we discuss a set of particles with such properties, each of them with its own pecularities. These include *auch*, *noch*, *nicht*, *schon*, *gleich*, *nur*, *erst*, *immer*, *genau*, *bereits*, *sogar*, *lediglich*, and *zumindest*. Then we propose how to use prosodic information when building the semantic representation of an utterance.

[1] Prosodic information includes information on intonational patterns (pitch) as well as on duration, energy, and pauses. For the prosodic features that are used for the time being in automatic classification by ourselves, cf. e.g. (Kompe et al., 1995).

2 Data from the Verbmobil Corpus

In January 1995, 21 Verbmobil dialogues (N001K, ..., N0019K, N021K, N022K) including the so called "Blaubeurer Dialoge" have been labelled prosodically by M. Reyelt (TU Braunschweig) along the lines of (Reyelt & Batliner, 1994) (perceptual labeling combined with visual inspection of the pitch contour). They include 598 turns (69 minutes of speech) and constitute the data base of this paper. In the following, the tone labels will be disregarded; we will concentrate on the "functional" labels. We go through a number of selected particles and their occurrences in these dialogues. Accent is marked as follows in the label files:[2]

- Emphatic or contrastive accent (EK)
- Primary phrase accent (PA): the most prominent syllable in a prosodic phrase (if there are two equally prominent syllables in a phrase they are both labelled with PA)
- Secondary phrase accent (NA): the second most prominent syllable in a prosodic phrase (if prominent at all)

In the following examples, words with EK are written in capital letters (STRESS), words with PA in small capitals (STRESS), and words with NA in slanted letters (*stress*). Pauses are indicated by a comma. What we will refer to as Semantic Focus is marked with an f subscript and determined from the context, *not* from prosodic information. In case the context allows more possible foci, more f subscripts will appear.

Auch

The German *auch* associates with semantic focus. Whether *auch* is itself stressed or not, determines the location of the focus. This is illustrated by the following examples:

(2) das geht auch bei MIR$_f$

(3) das geht auch bei MIR$_f$

(4) das$_f$ geht AUCH bei mir

(5) ? das geht AUCH bei MIR

(6) ? das geht AUCH bei MIR

(7) ? DAS/DAS geht auch bei mir

The general rule says: if *auch* has no stress, the semantic focus is at the right and coincides with a PA; if *auch* is stressed, the semantic focus is at the left. An

[2] Note that the relationship between these three accent classes is most probably continuous and not strictly ternary. It is an empirical matter to be solved whether all three of them can be distinguished automatically or whether we, e.g., should combine PA and EK into one single label.

utterance with *auch* introduces a presupposition, which is best described as the scope of *auch* with an alternative substituted for the semantic focus. For example, the presupposition of (2) is "das geht bei X", where X is likely to be the hearer. For (3), the presupposition is "X geht bei mir", with X being an alternative for what "das" refers to.

In our data base, there are in total 53 occurences of *auch* which are prosodically labelled. 25 of them are non-stressed, and for these 21 confirm the rule above, i.e., the semantic focus is on its right and marked with PA or EK. So, 28 of the *auch*s are stressed (PA/NA), and in 26 cases the semantic focus is situated left of *auch*, but does not correspond with a certain stress pattern. We may tentatively conclude that the general rule sketched above holds for our data set.

Noch

As we saw in example (1), the presupposition that *noch* introduces depends on its intonation. A stressed *noch*, in a case where it modifies a NP (8), has a different presupposition than a *noch* without stress (9). The english translation would be "another". The presupposition also depends on what kind of sentences it modifies: modal sentences with *noch* (9) have different presuppositions than non-modal ones (11). Another interpretation of *noch* is where it is used in a comparitive construction: then it acts as an intensifier (10).

(8) Dann lassen wir uns NOCH einen Termin ausmachen

(9) Dann lassen wir uns noch einen Termin ausmachen

(10) Das ist noch besser

(11) Peter ist noch krank

An instance of *noch* without stress within a modal sentence (9) triggers a presupposition that there was some event in the past that logically and temporarily preceeds the event that *noch* is modifying. There seems also a presupposition that this event has not taken place yet. In summary: the weak presupposition of *noch* in a modal sentence is that there is a certain event expected which has not taken place yet.

Yet another interpretation of *noch* is where it modifies a temporal construction, like *im März noch*, which comes closes to "still in March". This interpretation comes close to the one of (11), which presupposes that Peter is sick for a longer time than you might have expected. An utterance like

(12) dann sollten wir im März noch ein Treffen ausmachen

is three way ambiguous, depending on whether *noch* is stressed or not, and whether it modifies the NP *ein Treffen* or the temporal expression on its left. Note that prosodic information might help us to disambiguate the latter: a prosodic boundary just before the particle (presumably in combination with a PA on *Treffen)* rules out the second reading (modification of the temporal expression), a prosodic boundary

just after it (presumably in combination with a PA on *März*) rules out the first reading (modification of the NP). Prosodic boundaries occur often in spontaneous speech and prosodic information that represents these boundaries could clearly be helpful.

There are exactly 100 occurrences of *noch* in the corpus. Unluckily, 98 of them are non-stressed. There are two occurrences of stressed *noch*; in these cases it modifies a comparitive adjective (*noch besser*) and one in an *auch noch* construction, both cases where prosody does not really matter.

Nur

The German "only" is, like the English one, characteristic because it *does* contribute to the meaning of an utterance. Take the utterance:

(13) Ich hab' Dienstag nur vormittags eine Vorlesung

The meaning of this utterance is ambiguous. It could mean something along "Ich habe Dienstag vormittags eine Vorlesung", and there is no time alternative to "vormittags", on "Dienstag", where this is true, i.e., "ich habe Dienstag nicht nachmittags eine Vorlesung". Since the first implication holds also under negation, it is the presupposition of the utterance. This first possible interpretation is most likely when *vormittags* is stressed, and should be translated into English with "on Tuesday, I have got a lecture only in the morning".

Another possibility appears when we stress *Vorlesung*. Then the meaning of *nur* changes significantly: it should be paraphrased as something along: "the only problem is that I have got a lecture on Tuesday morning" when it is a rejection (*Ablehnung*), or "I have only got a lecture", nothing more that costs time.

Like *auch*, *nur* is sensitive to focus. In the first case *nur* has *vormittags* as focus, in the second the whole sentence is the semantic focus (scope equals focus) and in the third case it is the *Vorlesung* which is in focus.

There are only 18 occurences of *nur* of which we have prosodic information. Eleven times there is a PA/EK on the right of an unstressed *nur*, in all cases of which the PA/EK is the semantic focus. Three cases exist where *nur* itself is stressed with PA/EK, with the focus on its right with NA, and three cases where *nur* has NA and its focus PA. There was one case with *nur* unstressed, occuring at the end of a sentence, immediately following its focus with PA: "aber UNGERN nur".

3 Prosody in Verbmobil

Prosodic information for speech data in Verbmobil is written down in a special section in the "infostring" that is part of the ASCII interface between acoustics/phonetics on the one hand and linguistics on the other hand; as for details, cf. (Noeth & Plannerer, 1994) and below. That means that for each word in the word graph, scores

can be found for the prosodic marking of three different phenomena; these scores are most likely normalized onto the region between 0 and 1 and represent the probability that a certain event takes place:

- **boundaries**, e.g. strong (phrase) boundary B3, weak (intermediate) phrase boundary B2, no boundary B0
- **accents** (phrase accent)
- **sentence mood**, e.g. question vs. non-question

The following description of the prosodic part in the infostring is taken from (Noeth & Plannerer, 1994), pp. 4,5 (translation by the authors):

> **Prosody**
> `(PR (G ...) (M ...) (A ...))`
>
> - `PR` denotes that prosodic information follows;
> - `G` that information with respect to the marking of boundaries follows;
> - `M` denotes that information with respect to the prosodic marking of sentence mood follows;
> - `A` denotes that information with respect to the prosodic marking of phrase and sentence accents follows.
>
> The symbols `G`, `M`, or `A` are followed by a series of values for possible classes. For example, `(G .1 .8 .1)` says that there is a possibility[3] of .1 that the phrase has no boundary, a possibility of .8 that the phrase has a weak boundary, and a possibility of .1 that there is a strong boundary. The names of the classes are excluded to reduce the size of the infostring. The number, order, and value ranges of the classes are described in an off–line document.

`(A .1 .9)`, e.g., says that there is a possibility of .1 that the word in question is not accented, and a possibility of .9 that the word in question carries the phrase accent. Note that the number of classes is not fixed. This is due to the fact that it is not clear yet which prosodic phenomena can be classified with a sufficient reliability. We already know e.g. that the intra- and intersubjective reliability for "weak" boundaries is distinctly smaller than for "strong" boundaries (Reyelt, 1993). "Weak" boundaries could, however, be used for the proper classification and recognition of PP-attachment etc.[4]

[3] In the interface definition it is not defined what these possibilities are; however in the VERBMOBIL "Forschungsprototyp" they actually will be probabilities.

[4] Classification results for the ERBA corpus, cf. (Kompe et al., 1995), and preliminary results for the Verbmobil corpus show that it might be possible in the near future to classify weak boundaries with sufficient reliability.

In linguistics, continuous information is not usual; the scores can easily be mapped onto a binary or ternary distinction as, e.g. +/− stressed, +/− boundary, +/− question or strong/weak/no stress, strong/weak/no boundary, and terminal/progredient/interrogative. E.g. in the example above, `(G .1 .8 .1)` can be mapped onto "the word in question is followed by a weak boundary" (second column is the greatest of all three columns). Note that there are different mappings possible. E.g. the following rules could be used to decide if a word is regarded as accented: a word carries a phrase accent (PA),

- if the accent score is greater than 0.5
- if the accent score is greater than 0.8
- if the accent score is greater than the non-accent score
- if the difference between accent score and non-accent score is greater than .3
- if the accent score is greater than 0.5 and less than 0.8 (if it is greater than 0.8 the word carries an emphatic accent)
- if the accent score is the greatest one in the phrase no matter what its absolute value is

All of these rules might be used by the different linguistic modules or even by the same module, and which one is used depends on the actual problem to be solved during the analysis of an utterance. Thus the mapping will not be performed by the prosody module, but the prosody module just scores the word graph and the actual mapping is performed by the linguistic module itself, when it uses the prosodic information. (E.g. the syntax module from Siemens currently used in the VERBMOBIL demonstrator uses the prosodic boundary scores directly within the search for the best path/parse of the word graph (Bakenecker et al., 1994).) Note, that the labels PA and NA are relative quantities by definition: PA is the most prominent and NA the second most prominent syllable within a phrase regardless the absolute prominence. Thus an automatic distinction between PA and NA can only be done using the accentuation score computed by a classifier. We analyzed the accent probabilities computed by a classifier for the 21 dialogues with respect to this. We found that the average probabilities are 7.5 for PA, 6.7 for NA and 7.7 for EK. Since the variances are quite high, the absolute score does seem to be appropriate to distinguish these classes. However, in 72% of the cases PA got a higher probability than NA in the same phrase, and about 80% of the EK got higher probabilities than PA or NA. In 76% of the cases a word labeled with PA, NA, or EK has a higher accent probability than any other non–labeled word in the same phrase (whereas 86% of the PA/NA/EK–words got an accent probability of at least 0.5).

The representation of prosodic marking in general is not fixed. If necessary, this information can be supplemented with other prosodic information along the lines of (Reyelt & Batliner, 1994). In this memo, an inventory of prosodic labels for the speech data in Verbmobil is described. Accents, e.g. are specified as being

phrase accents (PA), secondary phrase accents (NA, i.e. "Nebenakzent"), and emphatic or contrastive accents (EK). In addition to B1, B2, and B3 explained above, "agrammatical" boundaries indicating hesitations or repairs are labelled as B9. The intonation (pitch contour) of accents is labelled with a modified tone sequence approach.

In the following, we will take the prosodic hand labels that were given to the utterances "at face value" - as if we really could work with these labels. In reality, these labels constitute the references for training and testing. The outcome of such an automatic classification are the above mentioned scores in the infostring that can be converted into "hard decisions". Note that this sort of information has two drawbacks:

First, it might not be correct; the so far best classifier for accented vs. non–accented syllables yields a recognition rate of 83.3% (i.e., 82% of the words are correctly classified). It is a neural network which gets as input durational, intonational, pausal and energy features derived from the time alignment of the word hypotheses to be classified and of a few hypotheses from the context; for details cf. (Kompe et al., 1995; Kiessling et al., 1994).

Second, this information is so to speak blind as for the phonetic/linguistic context of the word in question: in a word graph, it cannot be decided for purely phonetic/prosodic reasons which one of the possible adjacent words are the "right" ones. Context information as it, e.g. might be necessary in order to compute durational variations is therefore confined to a sub–optimal solution, where one of all adjacent edges is used to compute this information. Furthermore, the features extracted from the pitch contour (e.g. regression coefficients) are computed over a certain window. We received better results when not taking a fixed window, but a window which is defined by certain syllables in the context. Also in this case, only approximative methods can be used on the basis of the word graph. Furthermore, as in word recognition a stochastic language model can be helpful for prosodic classifying or scoring. This models the probability that a certain word is accented or succeded by a boundary given a certain word (sub–) chain surrounding the word. Also in this case a suboptimal approach has to be applied when using word graphs. (For more details concerning the problem of prosodic scoring of word graphs cf. (Kompe et al., 1995).)

From this point of view, a top down procedure can be imagined where semantics e.g. wants to disambiguate two possible interpretation of a phrase with the same words and word order and initiates an inquiry which of two possible phrase accent positions is most likely. For the time being, a – maybe suboptimal – equivalent to this information is simply passed through with the mapping of the scores in the infostring onto hard decisions. That makes it possible, however, that a disambiguation cannot be achieved because the two possible positions are both marked as accented. If the scores are not dichotomized but passed through as is, a disambiguation might almost always be possible because most certainly, the exact scores of two events differ while the hard decisions based on these scores may not.

In principle, the incorporation of prosodic information into semantics can proceed along the following lines:

- First, an analysis can be based on the hand labels and on the spoken word chain; this is the "best case" analysis and we followed this approach in the first part of this paper.
- Then, we can use the outcome of an automatic classification and the spoken word chain. This is a sort of "medium" case analysis because the automatic classification is not always correct, cf. above.[5]
- Later on, syntax and semantics have to work with the outcome of the word lattice, cf. e.g.(Bakenecker et al., 1994). This is the "worst case" analysis but the "real life" task at the same time.

Note that the impact of prosody must not be the same in the "best case" analysis as in the "worst case" analysis. This is an empirical matter to be solved.

4 Semantic Representations for Focusing Particles

In this section we exemplify how focusing particles are represented in the semantics using prosodic information. We use LUD (a Description Language for Underspecified DRSs, (Bos, 1995)) as a description language for Discourse Representation Structures (DRSs) of DRT (Kamp, 1981). LUD enables us to underspecify certain semantic phenomena, like quantifier scoping and anaphora resolution. The mapping from a LUD representation to a DRS is hence a one-to-many mapping, the (model) interpretation of a LUD representation is the set of interpretation in the object language (DRSs in DRT).

A LUD representation is characterized by its descriptional power. Every single piece of information is uniquely labeled, and these labels make it very easy to talk about structures and express relationships between different pieces of structure. For scoping, LUD includes so-called *holes*, which are meta-variables over semantic representation, to underspecify e.g. quantifier scoping. Every LUD representation has at least one hole (the top-hole), and scope ambiguities arise when there is more than one hole. Below we will restrict ourself to simple example and avoid scoping ambiguities. A LUD-R is therefore a triple of holes, conditions, and constraints. Consider the LUD-R for (14):

(14) Der Montag geht bei mir

[5] For example, we had a look at the scores obtained with a preliminary automatic classification for each occurrence of *auch* in our database. We defined a word as accented if its score was greater than 0.5. In 25% of the cases, a mismatch could be observed between PA label and a score smaller than 0.5, and between no accent label and a score greater than 0.5. The percentage correctly classified was thus 75%.

holes:	conditions:	constraints:
h_0	l_1: e	$l_5 \leq h_0$
	l_2: gehen(e)	l_8 α_{def} l_5
	l_3: theme(e,x)	l_{11} α_{deic} l_5
	l_4: bei(e,y)	
	l_5: $\wedge\{$ l_1 , l_2 , l_3 l_4 $\}$	
	l_6: x	
	l_7: montag(x)	
	l_8: $\wedge\{$ l_6 , l_7 $\}$	
	l_9: y	
	l_{10}: speaker(y)	
	l_{11}: $\wedge\{$ l_9 , l_{10} $\}$	

The labels l_1, l_6, and l_9 introduce discourse markers. Grouping of conditions is done by l_5, l_8, and l_{11}. The alfa constraints express that "der Montag" and "mir" are anaphoric (since they are definite descriptions resp. deictic) with respect to "gehen". The DRS (to save space, DRSs are shown in linear format) that can be transformed from this LUD-R is for example:

(15) [e x y | montag(x) gehen(e) theme(e,x) bei(e,y)] <y, speaker>

In this DRS "der Montag" is accommodated (since there is no suitable antecedent available), and the discourse marker y is anchored to the current speaker. Scoping ambiguities are not relevant in this example (in fact, l_5 is plugged into hole h_0). Now we make our example slightly more interesting by adding a focusing particle to it (16):

(16) Der Montag geht auch bei mir

As we know, this example is without any prosodic clues ambiguous in its presuppositions. Depending on the focus of "auch", the presuppositions could be that another time than "der Montag" is possible for the speaker, or that "the Montag" is also possible for the hearer. And there might be even more possible interpretations. Anyway, the LUD-R for (16), introducing a new kind of condition for focusing adverbs, is:

holes:	conditions:	constraints:
h_0	l_1: e	$l_5 \leq h_0$
h_1	l_2: gehen(e)	l_8 α_{def} l_5
	l_3: theme(e,x)	l_{11} α_{deic} l_5
	l_4: bei(e,y)	$l_5 \leq h_1$
	l_5: $\wedge\{$ l_1 , l_2 , l_3 l_4 $\}$	$l_{12} \leq h_0$
	l_6: x	
	l_7: montag(x)	
	l_8: $\wedge\{$ l_6 , l_7 $\}$	
	l_9: y	
	l_{10}: speaker(y)	
	l_{11}: $\wedge\{$ l_9 , l_{10} $\}$	
	l_{12}: auch(h_0,h_1)	

This focus condition expresses both focus and scope of *auch*, using holes as arugments. The first argument is its focus, the second its scope. This leaves us "free" to choose a focus in this example, since every label is subordinated by the top hole h_0. Possible DRSs are for example:

(17) assert:
[e x y | montag(x) gehen(e) theme(e,x) bei(e,y) x≠x'] <y, speaker>
presup:
[e' x' y' | gehen(e') theme(e,x') bei(e',y')] <y', speaker>

(18) assert:
[e x y | montag(x) gehen(e) theme(e,x) bei(e,y) y≠y'] <y, speaker>
presup:
[e' y' | gehen(e') theme(e,x) bei(e',y')]

The DRS in (17) is fine for (16) with stress on "auch", DRS (18) is fine for (16) with "auch" unstressed and "mir" stressed. Therefore, as soon as prosodic information is available, one of these interpretation can be ruled out, and we can constrain this in the LUD-R: auch(l_8,h_0) for the first interpretation and auch(l_{11},h_0) for the second one. These rules can be specified in the lexicon.

The interface between prosody and semantics is simply established via syntax. For each word in the sign, the logical nodes are specified, which correspond with the ones in the infostring. The infostring then tells us what the score for a particular word is. A score higher than .5 (a chance of 50 percent), for example, could be interpreted as "stressed". The exact figure has to be found empirically, though.

5 Conclusion

We showed why prosodic information could be used to disambiguate semantic interpretation of German utterances, resolving at semantic representations, which

are by themself underspecified, but could be made more specific by directly using information from the prosody component. With no doubt the resulting semantic representations help both the transfer and generation component to make better translations in Verbmobil.

In the very near future the architectural set-up described in this memo is going to be implemented as well, using the speech recognizers of Daimler Benz or Univ. of Karlsruhe, the prosody module of Univ. of Erlangen, the Siemens TrUG–parser, and the Semantic Formalism of Univ. of Saarbrücken.

References

(Bakenecker et al., 1994) G. Bakenecker, U. Block, A. Batliner, R. Kompe, E. Nöth, and P. Regel-Brietzmann. Improving Parsing by Incorporating 'Prosodic Clause Boundaries' into a Grammar. In *Int. Conf. on Spoken Language Processing*, volume 3, pages 1115–1118, Yokohama, September 1994.

(Bos, 1995) Johan Bos. LUD, a Desription Language for Underspecified Discourse Representation Structures. draft, Universität des Saarlandes, July 1995.

(Kamp, 1981) Hans Kamp. A Theory of Truth and Semantic Representation. *Formal Methods in the Study of Language*, 1, 1981.

(Kiessling et al., 1994) A. Kießling, R. Kompe, H. Niemann, E. Nöth, and A. Batliner. Detection of Phrase Boundaries and Accents. In Niemann, de Mori, and Hanrieder, editors, *Progress and Prospects of Speech Research and Technology: Proc. of the CRIM/FORWISS Workshop (München, Sept. 1994)*, pages 266–269, Sankt Augustin, 1994. infix.

(Kompe et al., 1995) 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*, page (to appear), Madrid, September 1995.

(Noeth & Plannerer, 1994) E. Nöth and B. Plannerer. Schnittstellendefinition für den Worthypothesengraphen, Verbmobil–Memo–2–94, Januar 1994.

(Reyelt, 1993) M. Reyelt. Experimental investigation on the perceptual consistency and the automatic recognition of prosodic units in spoken German . In *Proc. ESCA Workshop on prosody*, pages 238–241, Lund, September 1993.

(Reyelt & Batliner, 1994) M. Reyelt and A. Batliner. Ein Inventar prosodischer Etiketten für VERBMOBIL, Verbmobil–Memo–33–94, Juli 1994.