Multi-Tier Annotations in the Verbmobil Corpus

Karl Weilhammer, Uwe Reichel, Florian Schiel

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3, 80799 München, Germany
{weilkar, reichelu, schiel}@phonetik.uni-muenchen.de

Abstract

In very large and diverse scientific projects where as different groups as linguists and engineers with different intentions work on the same signal data or its orthographic transcript and annotate new valuable information, it will not be easy to build a homogeneous corpus. We will describe how this can be achieved, considering the fact that some of these annotations have not been updated properly, or are based on erroneous or deliberately changed versions of the basis transcription. We used an algorithm similar to dynamic programming to detect differences between the transcription on which the annotation depends and the reference transcription for the whole corpus. These differences are automatically mapped on a set of repair operations for the transcriptions such as splitting compound words and merging neighbouring words. On the basis of these operations the correction process in the annotation is carried out. It always depends on the type of the annotation as well as on the position and the nature of the difference, whether a correction can be carried out automatically or has to be fixed manually. Finally we present a investigation in which we exploit the multi-tier annotations of the Verbmobil corpus to find out how breathing is correlated with prosodic-syntactic boundaries and dialog acts.

1. Introduction

A typical characteristic of Language Resources (LR) in Spoken Language Processing (SLP) is the fact that they combine measurable, in most cases digitised signals with discrete symbolic data which denote some kind of semantics associated with the signals: The classic example is a corpus of recorded speech signals together with some kind of annotation. During the last decade a lot of technical approaches dealing with these representations have been developed and used by engineers and scientists. Examples are Birds annotation graphs (Bird, 2001) which is the most general approach, the BAS Partitur Format (Schiel et al., 1998), the representation within the Emu system (Cassidy and Harrington, 1996), annotation standards like TIMIT, SAM, Switchboard, UTF and plenty of others (see (Bird, 2001) for a good overview). In some of these also the problem of the 'multi-tier' representation of symbolic data associated to a signal was tackled and – more or less elegantly – solved.

In this contribution we will discuss how to successfully integrate several sources of symbolic information that are all based on the same LR, but were produced in a disorganised, organic fashion as it happens in many science projects – especially in those that do not have producing a re-usable LR as a top goal. How to deal with inconsistent input caused by manual alterations of baseline data, error updates that were not documented or propagated to all sources, with new unexpected semantic information that needs to be integrated etc.

One way, of course, to avoid the problem would be a clear definition of standards and semantics at the beginning of a project in which re-usable LRs are produced. The reality of SLP projects teaches us that this is not possible in most cases: formats and semantics of symbolic data are amongst the topics of the scientific process and cannot be forseen from the beginning. Therefore, we must expect all kinds of changing of the specs in the course of such a project. Exceptions are pure LR projects like SAM or the SpeechDat series where the specs were quite simple and stayed fixed forever. But even in the SpeechDat corpora we might face similar problems with several levels of error update, added layers of information etc. in the future.

Our main experience with this problem stems from the German Verbmobil project (VM). The Bavarian Archive for Speech Signals (BAS) located at the University of Munich agreed to take care of the long-term maintenance and distribution of the LRs produced during Verbmobil. This LR evolved over time into one of the most complex LR that exist in the German language.

This paper is organised as follows. First we will give a short overview about the Verbmobil project with regards to its LRs. Section 3. will briefly explain the basic principles of the BAS Partitur Format (BPF) that was used as the structural paradigm in the Verbmobil LRs. Section 4. describes our methods to deal with misaligned symbolic data in the integration process. Finally, to stress the point that all this effort is worth it, we present some interesting analysis results that could only be derived from the fully integrated Verbmobil LR.

2. The Verbmobil Corpus [1]

The Verbmobil project (1993 – 2000) aimed at the development of an automatic speech to speech translation system for the languages German, American English and Japanese (Wahlster, 2000). Within Verbmobil an empirical data collection was carried out by seven academic institutions in Tokyo, Pittsburgh, Kiel, Bonn, Hamburg, Karlsruhe and Munich. The main task of this data collection was to record a large corpus of spontaneous speech dialogs and provide annotations to train the acoustical models, the

[1] After the end of the Verbmobil project the corpus has been maintained by the BAS. The speech signals can be ordered on CD or DVD (`bas@bas.uni-muenchen.de`). The symbolic data can be downloaded for free via FTP `ftp://ftp.bas.uni-muenchen.de/pub/BAS/VM`.

language models, to build up the translators dictionary (together with most likely pronunciation variants), to train and test the syntactic/semantic analysis and the transfer. Aside from the main corpus some minor data collections were done for special tasks like command word spotting, module evaluation, concatenative speech synthesis, emotion detection and end-to-end evaluation. In this paper we will only deal with the main corpus, that is dialog recordings in three languages (mono- and multilingual).

The very first distributed VM volumes contained only the speech signals (cut in dialog turns) together with a complex 'transliteration' that included not only the orthographic text but also markers for many effects that occur in non-prompted spontaneous speech (Burger, 1997) [2]. Other partners started to work on these data. Many of them developed their own annotations. To integrate all these different kinds of symbolic data into one common structure the BAS Partitur Format (see next section) was defined in 1996. At the end of the first part of the Verbmobil project (1997) there existed already 9 different tiers in the VM LR: transliteration (TRL), lexical (ORT), pronunciation (KAN), two flavors of manual phoneme segmentation (SAP, PHO), automatic phoneme segmentation by MAUS (Kipp et al., 1997) (MAU), dialog act labeling (DAS), word segmentation (WOR) and a prosodic labeling in GTobi (PRB). At that time we faced the first problems caused by error updates in the transliteration that needed to be propagated through most of the tiers and we manually corrected the dependent tiers.

In the second part of Verbmobil the data collection was re-organised and more emphasis was given to English and Japanese as well as the multilingual recordings. Again new symbolic data were 'invented' by the partners. Some of the already existing annotations were modified, which means that old data had to be adjusted: syntactic based prosodic boundary labeling (PRO), signal based prosodic boundary labeling (LBP, LBG), syntax trees (LEX, SYN,FUN), syntactic word classes (POS), noise marker (NOI), VM2 transliteration (TR2), overlapped speech (SUP) and lemma tagging (LMA).

Tier	Turns
TR2	90025
PRO	29564
DAS	23560
LEX, SYN,FUN	22681
POS, LMA	61406
WOR	920
SAP	372
MAU	29115
PRB	917

Table 1: Selected tier of the Verbmobil corpus and the number of dialog turns (utterances) for which these annotations are available. The dialogues are in German, English or Japanese.

In 2000, after the official end of Verbmobil, all partners delivered their symbolic data and BAS started the integration of all these inputs into common BPF files and again we had to deal with the above discussed problems.

3. The BAS Partitur Format

A detailed and up-to-date description of the BPF can be found in the Internet [3]. Here we will just give the basic principle. The BPF links and aligns signal and symbolic data of a speech recording in a simple but effective way. There are basically two ways to link different tiers of symbolic information:

1. The physical absolute time measured from the beginning of the recording.
2. The discrete word number starting with zero.

Number 2 requires a definition of the concept of word, which is straight forward for English and German, but not trivial at all for the Japanese language. After all these two kinds of links are intuitive and convenient. In a speech signal we can label segments (time intervals) and singular events (points of time). Starting from this paradigm we find five different possible types of annotation:

1. Events attached to a word, a group of words or the gap between two words.
2. Events that denote a segment of time without a relation to the word structure.
3. Events that denote a singular time point without a relation to the word structure.
4. Events that denote a segment of time associated with a word, a group of words or the time slot between two words.
5. Events that denote a singular time point associated with a word, a group of words or the time slot between two words.

Within these five basic structures free syntax and semantics may be defined for an open number of annotation tiers based on the same signal. By adopting the label file structure of SAM it is possible to integrate all kinds of symbolic information linked to a physical signal.

The example displayed in figure 1 is a very short utterance from a German Verbmobil dialog recording (only selected tiers are shown to keep it brief). The speaker said: "Am Georgengarten. Ja, das habe ich mir notiert". ("At Georgengarten. Ok, I jotted that down."): The BPF in figure 1 contains a phonemic segmentation of type 4 (MAU) and several tiers of type 1. The successful integration of different layers of type 1, 4 or 5 is only possible, if the correct word structure of all tiers is in synchrony. However, if the data stem from different sources, you may never be sure about that. For instance the group creating the lemma tagging might have split all compound names into single items (for whatever reason; keep in mind that these groups do not work together to produce one single corpus, but rather to

[2] The English version of the Verbmobil transcription conventions: http://www.is.cs.cmu.edu/trl_conventions/

[3] Up-to-date description of the BPF: http://www.bas.uni-muenchen.de/Bas/BasFormatseng.html

```
TR2:    0       ~Am-Georgengarten .
TR2:    1       ja ,
TR2:    2       das7@ <!1 des>
TR2:    3       habe7@ <!1 haw>
TR2:    4       ich7@
TR2:    5       mir7@
TR2:    6       notiert7@ . <#Klicken>
SUP:    2,3,4,5,6       g015acn1_034_ABE.par
        @7also , +/@7das <!1 des> @7is<Z>t/+
        <!1 is'> <#Klicken> <P>
ORT:    0       Am-Georgengarten
ORT:    1       ja
ORT:    2       das
ORT:    3       habe
ORT:    4       ich
ORT:    5       mir
ORT:    6       notiert
KAN:    0       Q'am#geQ'O6g@n#g"a:6t@n
KAN:    1       j'a:
KAN:    2       das+
KAN:    3       ha:b@+
KAN:    4       QIC+
KAN:    5       mi:6+
KAN:    6       no:t'i:6t
NOI:    6;7     <#Klicken>
DAS: 0,1,2,3,4,5,6      @(INFORM AB)
SYN:    0       1       NX
SYN:    1       1       DM
SYN:    2       1       NX
...
FUN:    0       0       HD
FUN:    0       1       --
FUN:    1       0       -
...
LEX:    0       0       NE
LEX:    1       0       PTKANT
LEX:    2       0       PDS
...
POS:    0       NE
POS:    1       ITJ
POS:    2       PDS
POS:    3       VAFIN
POS:    4       PPER
POS:    5       PPER
POS:    6       VVPP
LMA:    0       Am-Georgengarten
LMA:    1       ja
LMA:    2       d
LMA:    3       haben
LMA:    4       pper
LMA:    5       pper
LMA:    6       notieren
MAU: 0 479 -1 <p:>
MAU: 480 479 0 Q
MAU: 960 639 0 a
MAU: 1600 2239 0 m
MAU: 3840 479 0 g
...
PRO: 0;1      LS2
PRO: 1;2      DS1
PRO: 6  SM3
```

Figure 1: BAS Partiture File of the Verbmobil Corpus

solve their specific task in the project). Then the lemma tagging and the baseline transliteration wouldn't be in synchrony any more.

4. Alignment

In this section we will describe the process of integrating different kinds of annotations into one coherent data structure. We will consider only annotations or sets of annotations that are independent of each other, but linked to only one reference. In the Verbmobil data the word numbers are the main references and the transliteration (TR2) is the basis annotation (anchor tier) on which all others depend. The main task is therefore to synchronise the links and dependencies between the different sources and the reference tier.

In the case of machine generated annotations, which can be easily reproduced, the problem of synchronisation is trivial, because an adjusted annotation can be created by applying the automatic algorithm and its knowledge base on the new anchor tier. Examples are automatic phoneme segmentation (MAU), part of speech tagging (POS) or the orthographic forms (ORT) extracted from the transliteration.

4.1. The Structure of Links

We will further focus on synchronising multi-tier annotations that are prepared by humans and therefore not easily reproducible. The relevant annotations correspond to BPF tiers of type 1, 4 and 5. For the task of synchronisation it is useful to distinguish between the following types of dependencies:

1. The dependent tier refers to the gap between two successive items (words) of the anchor tier (e.g. syntactic or prosodic boundaries).

2. The dependent tier refers to a single item of the anchor tier (e.g. POS).

3. The dependent tier refers to a number of successive items of the anchor tier (e.g. dialogue acts).

4. The dependent tier refers to both single items and groups of items within a set of annotations representing a (hierarchical) framework (e.g. syntax trees).

It is useful to specify the dependency types as exactly as possible, since using knowledge about the nature of the dependency increases the amount of corrections that can be treated automatically. If the anchor tier is modified (e.g. after an error correction process) all the dependent annotations have to be adjusted accordingly.

4.2. Detection of Differences

Given an old, uncorrected anchor tier with its dependent annotations and a new, corrected anchor tier we will outline an algorithm similar to dynamic programming to detect in a first step the differences in the two anchor tiers and generate in a second step a corrected version of the dependent tiers.

We specify a hierarchical set of operations that will enable us to transform the old anchor tier into the new anchor tier. The advantage of such a set of operations for difference detection is that for each operation, we can define a

old dependent tier	old anchor tier	new anchor tier	new dependent tier
	$\vdots$	$\vdots$	
$a_1 \longrightarrow$	w_1	w_1	$\longleftarrow a_1$
	w_2	w_2	
	w_3	$w_3 w_4$	
$a_2 \longrightarrow$	w_4		$\overset{?}{\longleftarrow} a_2$
	w_5	w_5	
	$\vdots$	$\vdots$	

Figure 2: Alignment and Correction.

correction process in the dependent annotation. We used the following hierarchy:

1. Transform a particular word or word chain into another word or word sequence
2. Split compound words and join neighbouring words (e.g. *pianobar* → *piano bar*)
3. Insert or delete a word or a group of words
4. Replace an unspecified word sequence by another unspecified word sequence

The highest level of the hierarchy and therefore the most determined case are specific word transformations. An example would be

$$\textit{can't} \quad \longleftrightarrow \quad \textit{can not.}$$

More general operations are splitting of compound words or joining of neighbouring words.

$$w_1 \quad w_2 \quad \longleftrightarrow \quad w_1 w_2$$

In this case the words that are to be split are not specified, i. e. *Piano Bar* would be transformed into *Pianobar* as well as *non smoker* into *non-smoker*. Insert and delete are applied if there is a word or a word sequence missing in either the new or the old anchor tier. Replace is used, if the old and the new anchor tier differ in a word or a sequence of words. The hierarchy is necessary, because if the level 4 replace was executed first, non of the other transformations would ever have a chance to be applied.

The process of difference detection between an old and a new anchor tier is organised as follows: We start with the first items of each anchor tier and compare them. If they are equal we continue with the next pair of items until the two tiers differ. At this point we test if one of the operations specified above can be applied to the *old anchor tier* to derive a sequence identical to the *new anchor tier*. Beginning with the most determined operation 1 and finishing with the most basic operation 4. If an operation leads to a satisfactory repair, the process of difference detection is stopped and the repair in the dependent tier is carried out.

If necessary the levels of the above hierarchy can be split into sublevels or a distance measure like the Levenshtein Distance can be used for instance to further distinguish a replace that is just due to a typo from a replace that changes the word sense.

In principal it is possible to apply a sequence of different operations. With a certain number of insert and delete operations each sequence of items can be transformed into any other sequence of items. The same holds for replace operations. For an automatic error correction it is important to find the set of operations that represents the logic structure in terms of the annotation best. In the actual work sequences of operations do not play an important role, because it is often to difficult to correct complex differences automatically in the dependent tier.

4.3. Correction

The process of error correction depends strongly on the nature and the complexity of the annotation. Therefore structural information as well as the actual content of the dependent annotation can be used for the corrections. In many cases they can be fixed automatically, in some a human expert is needed. We will discuss examples for some of the basic annotation types that are listed in section 4.1.

4.3.1. Sparse Distributed Annotations

In type 1 annotations the dependent tier refers to the gap between two successive items in the anchor tier. An example would be prosodic or syntactic boundaries (`PRO`). As it can be seen in figure 1 the labels of the PRO tier are typically sparse distributed. The label `LS2` refers to the gap between the first and the second word, `DS1` to the gap between the second and the third word and finally a `SM3` boundary closes the sentence after the sixth word. There are no entries for the gaps 2;3, 3;5 and 5;6. We can exploit this fact in the correction process. Differences originating from level 1 word-transformations or level 2 compound word operations will in general not affect the PRO tier unless they are carried out across a boundary, which is extremely rare and can be checked easily. Level 3 and level 4 differences, that are far away from a boundary, are not very likely to change the syntactical structure of the entire sentence, therefore no new boundary will have to be inserted or deleted and this case can be treated automatically. If an insertion is detected next to a boundary then it is a priori not clear if it goes before or after the boundary. In the case of syntactical boundaries we exploited punctuation – if available in the new transliteration tier – to decide whether a word was inserted before or after the boundary.

With deletions it is not clear if adjoining boundaries have to be canceled or not. Just imagine word 1 *ja* would have been deleted in figure 1. It stands between a `LS2` and a `DS1` boundary. Which of them is to be deleted? Decisions like that must be made by a human expert.

The dialogue act labeling, which is of type 3 represents another example of a sparse annotation. In this case word sequences are labeled as dialogue acts, not as in the example before the boundaries between them. Analogous to what was explained above differences occurring at the beginning or at the end of a dialogue act have to be examined more carefully than differences inside a dialogue act.

4.3.2. Complex Annotations

The most complex annotation we had to deal with was the hierarchical structure of a syntax tree represented in three tiers. The terminal symbols, syntactic word classes, are listed in LEX. LEX is a type 2 dependence. SYN is of type 3 and denotes syntactical phrases and their position in the hierarchy of the syntax tree. FUN is also type 3 and denotes the functions of the phrases and their positions in the tree. Each detected difference causes corrections in all three tiers.

The concatenation of a number of words entails the following procedure: For the correction of the LEX tier it is necessary to find out which of the words had the function of a *head* in the old annotation. The compound word inherits the word class of the last head. If the words were previously grouped in phrases, these phrases have to be deleted in SYN and the functions of the terminal symbols in FUN as well, involving a re-construction of the the syntax tree.

Splitting a compound word is not possible without additional linguistic knowledge, which can either be included in the correction algorithm or must be supplied by a human expert. For splitting German verbs into two words we chose the following procedure: The first word gets the LEX label *verbal particle* and the second word receives the verb-class label of the old composition and the function *head*. Most of the other corrections were processed manually.

An overview of the specification of the syntax trees, which were originally annotated in NEGRA format can be found in (Hinrichs et al., 2000) [4].

4.3.3. Practical Problems with Correctness

From the examples discussed above it is clear that concerning the repairs there is a trade-off between automatisation vs. correctness. There are repairs that can be implemented automatically without loss of correctness. Others can only be implemented with a high probability of correctness and finally there are those that must be done manually because a satisfactory heuristic solution would be to difficult to implement. For instance the assumption in section 4.3.1. that a difference occurring far away from a boundary would not change the annotation is highly probable, but there remain rare cases in which it might be incorrect. Since we are dealing with a finite corpus these cases can – if identified – be treated as exemptions. This is where the biggest amount of manual work has to be invested. And this is the point, where a project manager can define the degree of automatisation and correctness for the alignment.

5. Breathing in Spontaneous Speech

A data base of several aligned annotations stored in a well established format such as BAS-Partitur is much more valuable than each annotation alone. It provides the basis for the application of powerful data models. In the last part of this paper we want to demonstrate an analysis involving the positions of breathing, dialogue-act boundaries and syntactic-prosodic boundaries in the Verbmobil dialogues, exploiting information that comes from various aligned tiers.

5.1. The Breathing Cycle

Using only the TR2 tier we can obtain a histogram of the duration of the respiratory cycle during speech. The upper plot of figure 3 shows the breathing interval in words.

[4]Format specifications of the BPF are available via Internet: http://www.bas.uni-muenchen.de/Bas/BasFormatseng.html#SYN

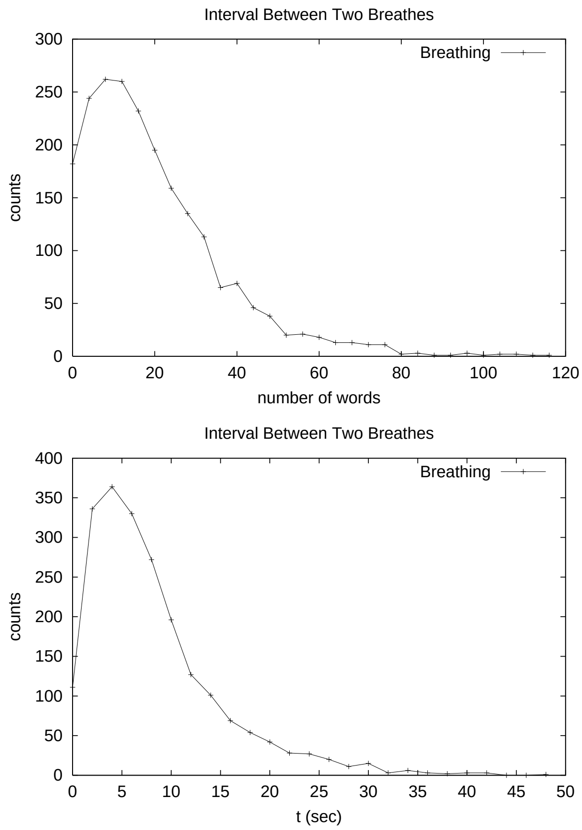

Figure 3: Duration of the Respiratory cycle in words (upper plot) and seconds (bottom plot).

The MAU tier establishes the relation between the transcription TR2 and the speech signal and thereby to time. Usually the automatic segmentation system assigns a pause symbol to a breath in the signal, or directly continues with the next word when the breath is very short. We obtained the positions of breaths by taking the value midways between the end of the word before and the start of the word after the breath (lower plot in figure 3). Both distributions have a similar shape. They rise quickly to a maximum at around 5 seconds or 12 words respectively, and after that decline in a wide tail.

5.2. Correlations with Prosodic-Syntactic Boundaries and Dialogue Act Boundaries

There are many publications in phonetic journals that deal with breathing during speech. (Winkworth et al., 1995) and (Henderson et al., 1965) report that inspirations are largely taken at sentence boundaries or other positions appropriate to the grammatical structure of spontaneous speech.

We used the syntactic and prosodic boundaries that are listed in the PRO tier (Batliner et al., 1998) to verify this statement for the Verbmobil corpus. Additionally we did the same tests with the more semantically oriented dialog act annotation of the DAS tier (Alexandersson et al., 1998). To avoid artefacts we did not consider breaths and boundaries that occurred at the begin and end of a turn.

The a priori probabilities for occurring between any two transcribed words have been calculated for breathing, prosodic-syntactic boundaries and dialog act boundaries.

$$P(B) = 0.07 \qquad P(\texttt{PRO}) = 0.30 \qquad P(\texttt{DAS}) = 0.08$$

Breathing occurs almost as often as dialogue act boundaries while prosodic-syntactic boundaries are about four times more frequent. The conditional probabilities for breathing on the position of a PRO or DAS boundary are

$$P(B|\texttt{PRO}) = 0.14 \qquad P(B|\texttt{DAS}) = 0.46.$$

Almost half of the dialogue act boundaries coincide with breaths, whereas only for 14 percent of the much more frequent prosodic-syntactic boundaries this is the case. To find out how good the positions of breath predict a PRO or DAS boundary we calculated the following conditional probabilities.

$$P(\texttt{PRO}|B) = 0.65 \qquad P(\texttt{DAS}|B) = 0.58.$$

About two third of all breaths occur on prosodic-syntactic boundaries and substantially more than half of them on dialogue act boundaries. Considering the fact that they are four times less frequent, the dialog acts come off well. To get a clearer picture we calculated the conditional probabilities for a DAS boundary given a PRO boundary and vice versa.

$$P(\texttt{PRO}|\texttt{DAS}) = 0.96 \qquad P(\texttt{DAS}|\texttt{PRO}) = 0.26$$

This reveals that the dialog act boundaries can approximately be understood as a subset of the PRO boundaries.

A randomly generated subset of the PRO boundaries of the same size as the dialogue act boundaries, would have led to conditional probabilities of

$$P(\texttt{PRO}_{rand}|B) = 0.17 \qquad P(B|\texttt{PRO}_{rand}) = 0.14.$$

This shows that a lot of semantic information relevant to our problem was added in the process of selecting the dialogue acts. We used a section of the Verbmobil corpus which had the size 90k words for this investigation. All the results are more than highly significant.

5.3. Conclusion to Breathing in Spontaneous Speech

On the basis of our analysis we can confirm, that breaths are largely taken on prosodic-syntactic boundaries. Especially on those that coincide with the end of a dialog act. That is when a semantic unit is finished.

6. Acknowledgements

We would like to thank all the groups of the Verbmobil data collection who contributed their annotations and supported us in the process of building a homogeneous corpus. We would especially like to thank Heike Telljohann, Valia Kordoni and Yasu Kawata (SYN, FUN, LEX), Michael Kipp (DAS), Anton Batliner (PRO) , Martin Emele (POS, LMA), Harald Lüngen, Thorsten Trippel (lexicon), Marcus Bäumler (PRB), Volker Warnke and Kerstin Fischer (emotional data), Daniela Oppermann, Susanne Burger, Akira Kurematsu and Susanne Jekat (TR2). This work was funded by the German Federal Ministry of Education and Science, Research and Technology (BMBF) in the framework of the Verbmobil project and State of Bavaria via the Bavarian Archive for Speech Signals BAS.

7. References

Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Michael Kipp, Stefan Koch, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, and Melanie Siegel. 1998. Dialogue acts in VERBMOBIL-2 – second edition. Report 226, Verbmobil.

Anton Batliner, Ralf Kompe, Andreas Kießling, Marion Mast, Heinrich Niemann, and Elmar Nöth a. 1998. M = Syntax + Prosody: A syntactic-prosodic labelling scheme for large spontaneous speech databases. *Speech Communication*, 25:193–222.

Steven Bird. 2001. A formal framework for linguistic annotation. *Speech Communication*, 33(1,2):23–60.

Susanne Burger. 1997. Transliteration spontansprachlicher Daten, Lexikon der Transliterationskonventionen in Verbmobil II. Technical Document 56, Verbmobil.

Steve Cassidy and Jonathan Harrington. 1996. EMU: an enhanced hierarchical speech data management system. In *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*.

A. Henderson, F. Goldman-Eisler, and A. Skarbek. 1965. Temporal patterns of cognitive activity and breath control in speech. *Language and Speech*, 8:336–242.

Erhard W. Hinrichs, Julia Bartels, Yasuhiro Kawata, Valia Kordoni, and Heike Telljohann. 2000. The Tübingen treebanks for spoken German, English, and Japanese. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, Artificial Intelligence, pages 550–575. Springer-Verlag, Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokio.

Andreas Kipp, Barbara Wesenick, and Florian Schiel. 1997. Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proceedings of the EUROSPEECH, Rhodos, Greece*, pages 1023–1026.

Florian Schiel, Susanne Burger, Anja Geumann, and Karl Weilhammer. 1998. The partitur format at BAS. In *Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain*, volume 2, pages 1295–1301.

Wolfgang Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Artificial Intelligence. Springer-Verlag, Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokio.

Alison L. Winkworth, Pamela J. Davis, Roger D. Adams, and Elizabeth Ellis. 1995. Breathing patterns during spontaneous speech. *Journal of Speech and Hearing Research*, 38:124–144.