BMBF

Verbmobil
Verbundvorhaben

The conventions for phonetic transcription and segmentation of German used for the Munich Verbmobil corpus

Anja Geumann
Daniela Oppermann
Felix Schaeffler

Ludwig Maximilians Universität München

Vm **Memo 129**
Dezember 1997

Dezember 1997

Anja Geumann
Daniela Oppermann
Felix Schaeffler

Institut für Phonetik und Sprachliche Kommunikation
Ludwig Maximilians Universität München
Schellingstr. 3/II
80799 München

Tel.: (089) 2180 -
e-mail: geumann, daniela, felix@phonetik.uni-muenchen.de

Gehört zum Antragsabschnitt: VERBMOBIL

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 701 D1 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den AutorInnen.

Contents

1 Introduction

This work describes the conventions for phonetic transcription and segmentation used for manual phonetic labeling of parts of the Verbmobil corpus (parts of CD 2, and CD 7). Transcription was done by Marion Libossek, Sonja Neubauer, Daniela Oppermann, Regina Petermann, Felix Schaeffler.
The symbol inventory is in large parts the same as used in the Phondat corpus (see Pompino-Marschall 1992). Also, the mode of segmentation is much the same as for the Phondat corpus. A major difference to the Phondat 2 segmentation is, that only linguistic units are transcribed. Therefore, the segmentation may contain breaks.
The data are stored in the Partitur files, in the so called SAP tier. For a description see Burger et al. 1997.

2 Symbol inventory

2.1 Consonants

IPA-Number	SAMPA	Example(orth.)	Example(SAMPA)
102	b	bei	baI
104	d	du	du
128	f	verfahren	f6fa:r@n
110	g	Gast	gast
146	h	Hast	hast
153	j	ja	ja
109	k	Kahn	ka:n
155	l	Licht	lICt
114	m	Mann	man
116	n	neun	nOYn
101	p	Platz	plats
122	r	Rauch	raUx
132	s	las, Maš	la:s, ma:s
103	t	Torte	tO6t@
129	v	Vase, wann	va:se
133	z	lesen	le:z@n
134	S	Tasche	taS@
135	Z	Loge	lo:Z@
138	C	dich	dIC
140	x	Dach	dax
119	N	Junge	jUN@

113	Q	ich	QIC

2.2 Vowels

IPA-Number	SAMPA	Example (orth.)	Example(SAMPA)
304 503	a:	Kahn	ka:n
304	a	kann	kan
302 503	e:	Beet	be:t
302	e	Meteor	meteo:6
303 503	E:	Käse	kE:z@
303	E	Bett	bEt
301 503	i:	riet	ri:t
301	i	Politik	politi:k
319	I	ritt	rIt
307 503	o:	bog	bo:k
307	o	Politik	politi:k
306	O	Bock	bOk
308 503	u:	Mus	mu:s
308	u	Kulisse	kulIs@
321	U	muš	mu:s
309 503	y:	Hüte	hy:t@
309	y	kyrillisch	kyrIlIS
320	Y	Hütte	hYt@
310 503	2:	Höhle	h2:l@
310	2	Ökonom	2kono:m
311	9	Hölle	h9l@
322	@	lesen	le:z@n
324	6	Leser	le:s6

2.3 Diphthongs

IPA-Number	SAMPA	Example (orth.)	Example(SAMPA)
304 319	aI	zwei	tsvaI
304 321	aU	Bauch	baUx
306 320	OY	neun	nOYn

2.4 Diacritics, not used in the canonical form

IPA-Number	SAMPA	Example
406	q	(Laryngealization, creaky) aq
424	∼	(Nasalization) a∼
404	H	(Aspiration) tH

2.5 Other Symbols not used in the canonical form

IPA-Number	SAMPA	Meaning, Syntax
960	-	Modifications of the canonical form Elision @-

		Replacement O-@
		Insertion -t
953	%	Uncertain boundary of a segment p
		Uncertain beginning p%<
		Uncertain end p%>
		Uncertain beginning and end p%<%>

3 Rules of transcription

- TR1
  A transcription symbol is chosen mainly by the auditory judgement of the segment in its context. The context should have the size of at least one syllable.

- TR2
  A deviation from the canonical form is only quoted if another category is perceived, e.g. /I/ instead of /i:/ as in /fy:zIk/ or /fy:zi:k/. More subtle variants induced by coarticulatory effects are neglected.

- TR3
  Only the symbols given above may be used.

- TR4
  If two adjacent homorganic segments merge (e.g. in 'hat den') the final segment is deleted (see TR5). If the auditory impression of a geminate remains, both symbols are preserved; the uncertain boundary can be marked by % , e.g. /kOm%>m%<al/.

- TR5
  In case of ambiguities concerning the elision of adjacent segments or assignment of symbols in reduced forms, the following rules have to be considered.

  1. A postvocalic segment in an unaccented syllable can be elided.
  2. Next, non-initial segments of an accentable syllable can be elided.
  3. Finally, a word- or syllable-initial segment can be elided.

- TR6
  If the deviation from the canonic form affects entire words, and an erroneous canonic form is obvious, the canonic form is corrected.

3. Rules of segmentation

- SR1
  Segment boundaries are positioned at positive zero crossings in the oscillogram.

- SR2
  The placing should be checked in the spectrogram.

- SR3
  In transient areas between two succeeding elements, when the sound quality of both appears simultaneously, the border is set in the middle of the transient area, if no other criterion fits. Examples are /sf/, /sS/, possibly /l/ and /r/ in intervocalic position. The boundary then is marked as %, e.g. /s%>f%</.

- SR4
  Voiced (periodic) segments start with the first clearly noticable oscillation. The initial boundary matches the first significant zero crossing. Otherwise, segments judged as voiced can be devoiced initially. In this case, the devoiced portion is assigned to the voiced element (e.g. vowel) and the auditory judgment is crucial.

- SR5
  In fricatives with low intensity (especially /h/) the segment starts when the signal to noise ratio becomes positive. The position of the end boundary (with utterance final fricatives) follows the same rule. Utterance final exhalation noise should be separated from friction or aspiration.

4 The treatment of single sound classes

4.1 Stops

1. The phases closure and burst, are both included in the stop segment. If the aspiration of /p,t,k/ or /b,d,g/ is clearly noticable and has a duration of more than 15 ms, H is inserted.

2. The beginning of utterance-initial stops or stops following a pause is, in most cases, uncertain, /p%</.

3. In a non-homorganic stop cluster in which the first stop is unreleased (but audible!), the first half of the closure is labelled as the first stop, the

remainder and the burst are labelled as the second stop. The boundary should be labelled as uncertain, e.g. /t%>p%</. The end of utterance-final unreleased stops is to be marked as uncertain.

4. The closure phase of a voiced stop starts after the last identifiable vowel oscillation.

5. Utterance final stops end after the burst. If necessary aspiration has to be inserted, but should be distinguished from exhalation noise.

6. The beginning of a voiced plosive with a preceding homorganic nasal is often difficult to determine. In these cases the low level amplitude phase of the nasal is considered as part of the stop. The burst is often only marked by a slight irregularity of the following oscillation.

7. Stops with an incomplete closure are treated as stops only if perceived as a stop. Otherwise it should be replaced by the corresponding fricative, or as appropriate.

8. The voiced - voiceless distinction keeps to the canonical proposal as far as possible. The canonical proposal is only rejected if a categorial change is audible and appears in the acoustic signal. E.g. /p,t,k/ is realised with a voice bar, /b,d,g/ is voiceless (and aspirated) in syllable initial or medial position.

9. The glottal stop is treated like other stops; therefore it consists of a closure and a burst. Usually the first oscillation is counted as burst phase. Utterance initial glottal stops begin with an arbitrary boundary. If no closure phase exists, the glottal stop is elided. In this case the adjacent vowel is often laryngealised.

4.2 Affricates

1. Affricates (e.g. ts, pf, tS) are treated as two segments.

2. The boundary between the two segments is placed after the burst. Only if a distinct aspiration is intelligible, is an H inserted, and the fricative follows the aspiration.

4.3 Fricatives

1. The glottal fricative may be elided following other fricatives.

2. Succeeding homorganic fricatives should be kept apart. If no separation is possible, the rules TR4, and TR5 are applied, as for stops.

4.4 Nasals

1. Syllabic nasals are separated from an adjacent nasal, as far as audible. Internal structure or longer duration may serve as a clue.
2. A nasal might be realized by nasalization of an adjacent vowel. In this case the nasal is deleted and the vowel is nasalized e.g. /a~/.
3. Devoiced nasals are not denoted any further.

4.5 Rhotics

The symbol /r/ summarizes the following articulations:

- uvular trill
- alveolar trill/flap
- uvular fricative voiced/voiceless
- velar fricative or glide

Postvocalic 'r' is canonically represented as /6/, as in /hambu6k/ 'Hamburg'. If postvocalic /6/ is realized as a trill or fricative, it has to be replaced by /r/. If only the preceding vowel is lengthened, the (short) vowel is replaced by a long vowel and the /6/ is deleted, e.g. /dE:/ instead of /de:r/, or /va:/ instead of /va:6/. The diphthong quality of vowel and following /6/ is marked by an arbitrary boundary, e.g. /va:%>6%</

4.6 Vowels

1. Long tense vowels are marked for length. Short tense vowels can be used, if necessary, but use should be limited.
2. Differing vowel quality should be noted as far as a categorial (perceptual) switch occurs.

3. If a vowel has diphthong quality, it may be replaced by /aI/, /OY/, or /aU/.
4. Devoicing is not marked.
5. Strong laryngealyzation/creaky voice is marked even if it does not replace a glottal stop.
6. Nasalization can be marked by a diacritic if it replaces an adjacent nasal, or in words like 'Restaurant' /rEsto:ra~/

4.7 Diphthongs

1. If diphthongs have monophthong quality the deviation from the canonic form has to be notated.
2. Additive diphthong qualities are avoided. Only in rare cases can new diphthtong qualities be built up by combining two monophthongs with an intermediate arbitrary boundary, e.g. /o%<U%>/.

5 Reduced forms

The following forms are derived from the rules in TR4 and TR5.

'haben wir' realized as /h a m 6/

h	→	h
a:	→	a:-a
b	→	b-m
@	→	@-
n	→	n-
v	→	v-
i:	→	i:-
6	→	6

'Abend' realized as /Q a: m t/

Q	→	Q
a:	→	a:
b	→	b-m
@	→	@-
n	→	n-

t	→	t

'hast du einen' realized as /h a s n/

h	→	h
a	→	a
s	→	s
t	→	t-
d	→	d-
u:	→	u:-
Q	→	Q-
aI	→	aI-
n	→	n
@	→	@-
n	→	n-

'einen neuen' realized as /aIq n OY n/

Q	→	Q-
aI	→	aI-aIq
n	→	n-
@	→	@-
n	→	n-
n	→	n
OY	→	OY
@	→	@-
n	→	n

6 References

Pompino-Marschall, B. (1992): PhonDat - Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank für gesprochenes Deutsch. *Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation München (FIPKM)* 30/1992, pp. 99-128.

Burger, S.; A. Geumann; F. Schiel; K. Weilhammer (1997): The Partitur Format at BAS. *Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation München (FIPKM)* 95/1997.