



Transliterationslexikon

Susanne Burger

Ludwig Maximilians Universität München



Technisches Dokument 36
Oktober 1995

Oktober 1995

Susanne Burger

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

Tel.: (089) 2180 - 2806

e-mail: burger@sun1.phonetik.uni-muenchen.de

Gehört zum Antragsabschnitt: 14 VERBMOBIL/PHONDAT

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 102 L/4 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Achtung!

Dieses Lexikon besteht aus mehreren Teilen, weil es teils mit HTML, teils mit LATEX erstellt wurde.

Mit den meisten PostScript Applikationen kann man nur den ersten Teil sehen. Beim Ausdrucken sollten dann aber alle Teile erscheinen.

Eine weitere und eigentlich auch bessere Möglichkeit ist, das Lexikon mit <http://www.phonetik.uni-muenchen/VMTraLex.html> (deutsch) bzw.

<http://www.phonetik.uni-muenchen/VMTraLexeng.html> (englisch) als Web-Page aufzurufen.

Inhaltsverzeichnis

1	Definition	3
2	Einleitung	3
3	Benutzung	4

1 Definition

Dieses Dokument beschreibt im Lexikon-Format die Konventionen und Symbole zur Transliteration von Spontansprache im VERBMOBIL-Projekt.

Das Lexikon beinhaltet in kurzer, schnell auffindbarer Form das gesamte Zeicheninventar, das zur orthographischen Verschriftung von Spontansprache in VERBMOBIL genutzt wird, beschreibt den Aufbau eines Dialog-Textfiles und gibt außerdem Auskunft über die in der Spontansprache etwas anders handzuhabende Interpunktion.

Die englische Version wurde in Zusammenarbeit mit dem IPDS in Kiel erstellt. Bei der englischen Übersetzung halfen Adrian Simpson (IPDS Kiel), Christoph Draxler, Phil Hoole und Florian Schiel (IPSK München) mit.

Eine umfassende Beschreibung des Zeicheninventars und eine ausführliche Anleitung zur Transliteration von Spontansprache findet sich im 'Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL – 3.0' (Kohler, Lex, Pätzold, Scheffers, Simpson, Thon Kiel, September 1994).

Das Lexikon soll nicht das Handbuch ersetzen, sondern zusätzlich Hilfe bei der Transliteration sowie später bei der Entschlüsselung der Dialog-Texte bieten.

2 Einleitung

Im VERBMOBIL-Projekt werden mit großem Zeit- und Personalaufwand riesige Datenmengen an Spontandialogen verschriftet.

Für eine adäquate Niederschrift von Spontansprache mußten neben der Orthographie Sonderzeichen eingeführt werden, die spontansprachliche Phänomene wie z.B. Atmen, Häsitationen und Abbrüche symbolisieren. Im Laufe der Arbeit an der Transliteration der Verbmobil-Dialoge erwies es sich als sinnvoll, neben dem Handbuch noch ein überschaubareres Nachschlagewerk zu den Transliterationskonventionen zur Verfügung zu haben, das einen schnellen Zugriff auf Symbol und Bedeutung zuläßt.

Insbesondere soll das Lexikon

- anzulernenden Transliterierern einen schnellen Überblick geben und bei Unsicherheit in der Beschreibung der spontansprachlichen Phänomene das zeitaufwendige Suchen im Handbuch ersparen

- geübten Transliterierern bei seltener verwendeten Symbolen oder bei komplizierten Symbolkombinationen ein schnelles Nachschlagen erlauben
- und schließlich bei der Weiterarbeit mit den Dialogtexten zur Verständlichkeit und Entschlüsselung beitragen.

3 Benutzung

Das Lexikon wurde mit Hypertext als WWW-Page erstellt. Auf diese Art steht es allen Anwendern sofort zur Verfügung und bietet weiter die dem WWW eigene bequeme und schnell 'anklickbare' Benutzeroberfläche, sowie schnelle Verbindung zu anderen mit dem Lexikon zusammenhängenden WWW-Pages.

Adressen:

<http://www.phonetik.uni-muenchen.de/VMTraLex.html> (deutsch)

<http://www.phonetik.uni-muenchen.de/VMTraLexeng.html> (englisch)

Wenn die WWW-Page geladen ist, finden sich folgende anklickbare Begriffe:

1. Allgemeines zur Transliteration
 - 1.1 Interpunktion
 - 1.2 Filestruktur
2. Symbole
3. Literatur

Interpunktion gibt Hinweise zur in der Spontansprache etwas schwierigen Interpunktionszeichensetzung, *Filestruktur* beschreibt den optischen Aufbau eines Dialog-Textfiles.

Klickt man *Symbole* an, erscheint zuerst eine Liste, in der man alle Transliterationsymbole und ihre Bedeutung findet. Die Liste ist nicht alphabetisch, sondern nach funktionalen Gruppen geordnet, da die Begrifflichkeit der Symbolbedeutungen oft nicht einheitlich verwendet wird. Klickt man weiter ein entsprechendes Symbol oder die Bedeutung an, so erscheinen Verwendungsbeispiele, eine kurzer Benutzer-Hinweis und eine Liste der Zeichen, die unmittelbar vor dem entsprechenden Symbol stehen dürfen.

In Anhang A und Anhan B finden sich PostScript-Fassungen der WWW-Pages:

Anhang A (deutsche Version).....Seite 6 - 24

Anhang B (englischeVersion).....Seite 25 - 44

Anhang A – Transliterationslexikon (deutsch)

Konventionslexikon zur Transliteration von Spontansprache

Susanne Burger

Institut für Phonetik und sprachliche Kommunikation
München

(Lexikon zum "Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL -3.0"
Kohler, Lex, Pätzold, Scheffers, Simpson, Thon
Kiel, September 1994)

Eine Postscript-Version des Handbuchs ist am IPDS in Kiel erhältlich.

Inhalt

1. Allgemeines zur Transliteration

1.1 Interpunktion

1.2 Filestruktur

2. Symbole

3. Literatur

1. Allgemeines zur Transliteration:

Alle in **Verbmobil** aufgenommenen Dialoge werden transliteriert

- in einem einheitlichen Format
- nach Duden (mit Ausnahmen, z.B. reduzierte Wortformen)
- mit Hilfe von definierten Konventionen, die das Transliterieren von Spontansprache

unterstützen

Umlaute und Sz werden nach LaTeX-Format verschriftet.

1.1 Interpunktion

Es gibt nur **Punkt**, **Komma** und **Fragezeichen**.

Nach Punkt und Fragezeichen wird klein weiter geschrieben, sofern es sich nicht um ein Substantiv handelt. Vor und nach dem Interpunktionszeichen steht jeweils ein Leerzeichen. Generell ist die Zeichensetzung bei Spontansprache mit Schwierigkeiten verbunden, da grammatikalisch "richtige" Sätze häufig fehlen.

Punkt

Symbol: " . "

steht am Satzende zwischen Leerstellen

Beispiel:

```
... so , guten Tag , mein Name ist <!1 is'> J"ansch .  
<"ah> wir hatten bereits telefoniert<Z> . ...
```

Anmerkung:

In Zweifelsfällen entscheidet, ob ein " . " gesetzt wird oder nicht:

- Grammatik
- Intonation
- Pause, Atmen
- Beginn eines neuen Gedankens

Fragezeichen

Symbol: " ? "

steht am Ende einer Frage zwischen Leerstellen

Beispiel:

```
... <A> wie schaut 's denn aus , den darauffolgenden  
Sonntag , den <:<#Mikrobe> neunundzwanzigsten:>  
bei Ihnen ? geht 's da ? ...
```

Anmerkung:

In Zweifelsfällen entscheidet, ob ein " ? " gesetzt wird oder nicht:

- Fragewörter
- Satzbau
- Intonation
- Kontext

Komma

Symbol: " , "

steht zwischen Satzteilen oder Nebensätzen zwischen Leerstellen

Beispiel:

```
... <"ahm> morgen , Freitag , <h"as> wie ich seh' ,  
<"ah> mu"s ich feststellen , da"s ich <"ah>  
"überhaupt keine Zeit hab' . ...
```

Anmerkung:

In Zweifelsfällen entscheidet, ob ein " , " gesetzt wird oder nicht:

- Grammatik
- Nebensatz einleitende Partikel
- Intonation
- zusammengehörende Gedanken

z.B.:

- a) ich k"onnte schon , mittwochs hab' ich immer
Zeit .
- b) ich k"onnte schon . <P> <A> mittwochs hab' ich
immer Zeit .

vor Leerstelle und Interpunktionszeichen darf stehen:

Großbuchstabe

Kleinbuchstabe

:>

<A> (nur bei Ausatmen!)

<Z>

<%>

<!n ..>

<i ..>

1.2 Filestruktur:

Ein Transliterations-File besteht aus einem **Header**

Beispiel:

```
; Dialog N057K
; zuletzt bearbeitet am 23.5.94
; Tonqualit"at: (allgemeine Kommentare zu
; Sprechern oder zur Aufnahmequalit"at des
; Dialogs)
```

und einem **Textteil**,
der durch eine Leerzeile vom Header getrennt ist.

Im Textteil werden die einzelnen Turns des Dialogs transliteriert. Jeder Turn beginnt mit dem dreibuchstabigen Sprecherkürzel und der Turnnummer, beginnend bei 000. Nach der Turnnummer steht ein Doppelpunkt, gefolgt von einem Leerzeichen. Dann folgt der eigentliche Text. Wörter werden nicht getrennt, an jedem neuen Zeilenanfang innerhalb des Turns stehen acht Leerzeichen. In der Zeile nach einem Turn kann noch ein mit ";" eingeleiteter Kommentar folgen, der sich auf den ganzen vorangegangenen Turn bezieht. Die einzelnen Turns sind durch jeweils eine Leerzeile voneinander getrennt.

Beispiel:

```
AAP000: so , guten Tag , mein Name ist <!1 is'>
      J"ansch . <"ah> wir hatten bereits
      telefoniert<Z> , mein Name J<Z>"ansch , $J
      $"A $N $$ $C $H , wegen <:<#Mikrobe> eines
      Arbeitstreffens:> .
```

```
BBP001: gr"u"s Gott , mein Name ist G<Z>"urtner , <A>
      <"ahm> $G $"U $R $T $N $E $R <A> . <A> <"ahm>
;Brummen "uber gesamtem Turn
```

```
AAP002: ja , <:<#Mikrobe> ich kuck' jetzt:> mal nach
      bei mir , wann ich <A> einen <!1 ein'> Termin
      frei h"atte . <A> das <:<#Mikrowind> erste:>
      w"are <Schmatzen> in der Woche oder die Tage
      vom vierzehnten Juli bis zum achtzehnten
      <:<#Mikrobe> Juli:> . <A> ginge das bei <:<#>
      Ihnen:> ?
```

```
BBP003: ...
```

2. Symbole:

Zu den Lexikoneinträgen

Symbol.....Bedeutung

\$.....	Buchstabieren
*.....	Nichtwort/Versprecher
<"ah>.....	
<"ahm>.....	
<hm>.....	
<h"as>.....	Häsitation/gefüllte Pause
<Schmatzen>.....	
<Schlucken>.....	
<R"auspern>.....	
<Husten>.....	
<Lachen>.....	
<Ger"ausch>.....	nonverbale artikulatorische Geräusche
<Z>.....	Zögern
<A>.....	Atmen
<P>.....	Pause
li_blank.....	Wortunterbrechung: linker Wortteil (Subline, Leerstelle)
blank_re.....	Wortunterbrechung: rechter Wortteil (Leerstelle, Subline)
=.....	Wortabbruch
/-.....	Phrasenabbruch
+/......	Selbstkorrektur: linke Grenze
/+.....	Selbstkorrektur: rechte Grenze
%.....	Schwerverständliches
<%>.....	Unverständliches
<i..>.....	Kommentar
< ;T>.....	technischer Abbruch
<! ..>.....	Aussprachekommentar
<#Klicken>.....	
<#Klingeln>.....	
<#Klopfen>.....	
<#Mikrobe>.....	
<#Mikrowind>.....	
<#Rascheln>.....	
<#Quietschen>.....	
<#>.....	nichtartikulatorische Geräusche
<:.....	Überlagerungsanfang
:>.....	Überlagerungsende
(.....	Überlappungsanfang (und (@
).....	Überlappungsende) und @)

Zu den Lexikoneinträgen:

Im Text der TRL-Files finden sich viele Sonderzeichen, die zur besseren Niederschrift von spontansprachlichen Phänomenen definiert wurden.

Im folgenden sollen alle verwendeten Symbole anhand von Beispielen kurz beschrieben werden.

Ein Eintrag enthält jeweils

- die Bezeichnung des spontansprachlichen Phänomens,
- ein Beispiel,
- das verwendete Symbol sowie Regeln zur korrekten Verwendung,
- eine Liste von Zeichen, die unmittelbar vor dem Symbol stehen dürfen

Buchstabieren:

Beispiel:

```
... mein Name ist J<Z>"ansch , $J $"A $N $$ $C $H ...
```

Symbol: \$

Jeder buchstabierte Buchstabe wird groß geschrieben,
\$ wird vorangestellt.

vor \$ darf stehen:

```
(  
+/  
-  
Leerstelle
```

Nichtwörter:

Beispiel:

```
... *haarknapp <:<#Rascheln> um einen <!1 ein'> Tag  
verfehlt ...  
... was *exkursieren Sie denn ? ...
```

Symbol: *

Als Nichtwort gelten:

- Neuschöpfungen,
- Versprecher,
- fremdsprachliche Wörter, die im Duden nicht aufgenommen sind,
- ausländische Namen.

Dem Nichtwort wird * ohne Leerstelle vorangestellt.

vor * darf stehen:

(
+/
< ; T >
Leerstelle

Häsitationen:

Beispiel:

... <"ah> wir hatten bereits telefoniert<Z> ...
... gu<Z>t , <"ahm> wie w"ar' es bei Ihnen am<Z>
neunzehnten Juli ...
... ich denke/- <hm> also bei mir ginge es sehr
gut ...
... w"urd' ich sagen , <A> <h"as> wenn wir ...

Symbol: <"ah> <"ahm> <hm> <h"as>

Häsitationen stehen zwischen <>

- <"ah>: rein vokalische Artikulation, unabhängig von der Vokalqualität
- <"ahm>: vokalische Artikulation + nasale Artikulation
- <hm>: rein nasale Artikulation
- <h"as>: seltenere Artikulationen, die nicht in eine der ersten drei Klassen einzuordnen sind.
Hierunter fallen z.B. /brrt/ /pf /puh/ etc.

vor <h"as> .. darf stehen:

(
< ; T >
Leerstelle

Artikulatorische Geräusche:

Beispiel:

... Mittagessen <P> <Schmatzen> und<Z> , na ja ...
... ordentlich planen mu"s und <"ah> <Schlucken> wann
w"urde Ihnen ...
... einverstanden . <P> <R"auspern> auf
Wiedersehen ...
... +/je/+ <Husten> je schneller wir das machen ...
... bei <:<Lachen> mir:> terminlich sehr
ung"unstig ...
... dann/- <"ah> <A> <Ger"ausch> das sind sechs
Termine ...

Symbol: <Schmatzen> <Schlucken> <R"auspern> <Husten> <Lachen> <Ger"ausch>

<Ger"ausch> steht für nicht-zuordenbare Produktionen.
Artikulatorische Geräusche werden zum Zeitpunkt ihrer Produktion
zwischen Leerstellen transliteriert.
Sie können auch andere verbale und nonverbale Produktionen überlagern.

vor <Ger"ausch> .. darf stehen:

(
<:
Leerstelle

Zögerung/Dehnung:

Beispiel:

... ich h"atte Zeit +/am<Z>/+ <A> <Schmatzen> ab
Dienstag ...
... ich dachte ger<Z>ade ...

Symbol: <Z>

<Z> wird ohne Leerstelle an den gedehnten Laut angehängt, unabhängig davon, an welcher
Stelle des Wortes die Dehnung zu hören ist.

vor <Z> darf stehen:

Kleinbuchstabe
Großbuchstabe

Atmen:

Beispiel:

```
... ist G<Z>"urtner , <A> <"ahm> $G $"U $R $T $N  
$E $R <A> . <A> ...
```

Symbol: <A>

<A> steht hauptsächlich für Einatmen und deshalb hinter dem Interpunktionszeichen. Bei deutlichem Ausatmen am Ende eines Satzteils kann Atmen auch vor einem Interpunktionszeichen stehen.

vor <A> darf stehen:

(
Leerstelle

Pausen:

Beispiel:

```
... ich habe <P> M"oglichkeiten dazu ...
```

Symbol: <P>

An Satz-/Satzteilgrenzen steht <P> immer nach dem Interpunktionszeichen.

vor <P> darf stehen:

(
Leerstelle

Wortunterbrechung:

Beispiel:

```
... Ver_ <A> _pflichtungen ...  
... statt_ +/f=/+ <h"as> _findet ...
```

Symbol: _

Wird ein Wort unterbrochen und nach der Unterbrechung fortgesetzt, so steht nach der Abbruchstelle _ ohne Leerstelle und vor der Fortsetzungsstelle _ ohne nachfolgende Leerstelle. Alle während der Unterbrechung stattfindenden Ereignisse werden getrennt durch Leerstellen zwischen den Wortteilen notiert.

vor li_blank (linkes subline) darf stehen:

-

<Z>

Großbuchstabe

Kleinbuchstabe

vor blank_re (rechtes subline) darf stehen:

Leerstelle

Wortabbruch:

Beispiel:

... +/sieb=/+ siebzehnter ...

... im Ja=/- also ich sag' Ihnen jetzt ...

Symbol: =

An ein abgebrochenes Wort wird = ohne Leerstelle angehängt.

vor = darf stehen:

<Z>

Großbuchstabe

Kleinbuchstabe

Phrasenabbruch:

Beispiel:

... ab dem dritten August <A> bis zum/- <P> Moment ,
ich ...

... ja , ich hab' da eigentlich/- also ich bin vom
neunzehnten bis ...

Symbol: /-

Ein Abbruch liegt dann vor, wenn der Sprechakt angehalten und nach dem Abbruch ein neuer Gedankengang begonnen wird. An der Abbruchstelle wird /- ohne Leerstelle angefügt. Hinter /- stehen keine Interpunktionszeichen.

vor /- darf stehen:

=
<A> (nur bei Ausatmen)
<Z>
< ;T>
Großbuchstabe
Kleinbuchstabe

Wiederaufnahme:

Beispiel:

.... +/am/+ <:#Mikrobe> am Donnerstag:> kann ich
erst ...
... +/im Sep=/+ im September ...
... die Woche +/von/+ <"ah> mit Freitag ...
... also +/das/+ das/+ das zweite ...

Symbol: +/../+

Bei einer Wiederaufnahme wird nach der Unterbrechungsstelle das vorher Abgebrochene entweder wiederholt oder korrigiert. An der Unterbrechungsstelle wird /+ ohne Leerstelle angefügt. Der Äußerung, die nach der Unterbrechungsstelle dann wiederholt/korrigiert wird, wird +/ ohne nachfolgende Leerstelle vorangestellt, sodaß +/.. /+ das Reparandum klammert.

vor +/ darf stehen:

(
Leerstelle

vor /+ darf stehen:

=
<A> (nur bei Ausatmen)
<Z>
< ;T>
Großbuchstabe
Kleinbuchstabe

Schwerverständliches:

Beispiel:

```
... %eins , %zwei , %drei , %vier ...  
... wann h"atten Sie %da bitte Zeit ...
```

Symbol: %

Vor schwerverständlichen Wörtern steht % ohne Leerstelle.

vor % darf stehen:

```
(  
+/  
<;T>  
Leerstelle
```

Unverständliches:

Beispiel:

```
... aber <A> <"ah> <%> <"ahm> wie w"ar's denn ...  
... <%> fr"uher geht 's leider nicht ...
```

Symbol: <%>

Für Unverständliches steht <%> zwischen Leerstellen.

vor <%> darf stehen:

```
(  
Leerstelle
```

Kommentare:

Beispiel:

```
; Tonqualit"at: viele Nebenger"ausche ...  
;gesamter Turn verrauscht ...  
... zum Beispiel <;"ubersteuert> ...  
... einen Termin vereinbaren <;heiser> ...
```

Symbol: ;

Ein den ganzen Dialog betreffender Kommentar steht vor dem gesamten Dialog im Header, am Zeilenanfang steht ein Semikolon, nach dem Kommentar folgt eine Leerzeile. Ein Kommentar zu einem Beitrag eines Sprechers wird am Zeilenanfang mit Semikolon eingeleitet und ohne Leerzeile an den Sprecherbeitrag angefügt.

Symbol: <;..>

Lokale Kommentare werden im Text hinter der relevanten Stelle und einer Leerstelle eingefügt, durch ein Semikolon eingeleitet und in spitze Klammern eingeschlossen.

vor <;..> darf stehen:

Leerstelle

Technischer Abbruch:

Beispiel:

```
... ab ein U<;T> ...
... <;T>neunzehnten is<Z>t etwas schwierig ...
... erste Woche ginge <A> <;T> <#Klicken> <A>
  <"ah> Montag ...
... unsere Termine alle untergebra<;T> <;T>ke ...
```

Symbol: <;T>

Durch den Knopfdruck kann die Sprachaufnahme am Anfang oder am Ende abgeschnitten werden, aber auch innerhalb eines Dialogbeitrags unterbrochen sein. Fand der Abbruch während einer lexikalischen Einheit statt, wird <;T> ohne Leerstelle davor oder danach angehängt. Zwischen zwei aufeinanderfolgenden <;T> steht eine Leerstelle. Wurde keine lexikalische Einheit abgeschnitten, steht <;T> zwischen Leerstellen.

vor <;T> darf stehen:

```
< :
+ /
<h"as>..
<Z>
Großbuchstabe
Kleinbuchstabe
Leerstelle
```

Aussprachevarianten:

Beispiel:

```
... damit w"ar' das <!1 des> eigentlich klar ...
... dann kommen Sie <!2 komm' Se> doch ...
... wenn wir <!2 wemma> 's die Woche noch machen ...
```

Symbol: <!n ..>

Dialektaussprachen, andere Stilformen oder sonstige Abweichungen werden in korrekter Form nach Duden verschriftet. Die eigentlich geäußerte Version wird hinter der korrekten Form

nach einer Leerstelle

in spitzen Klammern ,

nach einem ! ,

einer Ziffer, die sich auf die betroffenen lexikalischen Einheiten bezieht,

und einer Leerstelle transliteriert.

Hinweis: Genauer definierte Regeln zur konsistenteren Verschriftung von

Aussprachekommentaren finden sich im Verbmobil-Memo 111: Aussprachevarianten in der VERBMOBIL-Transliteration - Regeln zur konsistenteren Verschriftung (Burger, Kachelrieß, München, August 1996).

(<http://www.phonetik.uni-muenchen.de/VMTechDoks.html>)

vor <!n ..> darf stehen:

Leerstelle

nichtartikulatorische Geräusche:

Beispiel:

```
.. de<Z>m <:<#Klicken> neunundzwanzigsten:>
  August ...
... <:<#Klingeln> am Mittwoch st"anden:> ...
... <:<#Klopfen> gern so machen:> ...
... <#Mikrobe> das tut mir leid ...
... ich k"onnte Ihnen <:<#Mikrowind> vorschlagen:> ...
... nach Berlin <P> ausmachen . <#Rascheln> <"ahm> ...
... der gesamte <:<#Quietschen> Rest des Mais:> ...
... k"ame mir gelegen , <A> <#> allerdings ...
```

Symbol: <#Klicken> <#Klingeln> <#Klopfen> <#Mikrobe> <#Mikrowind>
<#Rascheln> <#Quietschen> <#>

<#Klicken> für z.B. Klicken des Knopfdrucks
<#Klingeln> Telefonklingeln
<#Klopfen> z.B. Tischberührung
<#Mikrobe> Mikrofonberührung
<#Mikrowind> ins Mikrofon blasen oder ins Mikrofon atmen
<#Rascheln> z.B. Papierrascheln
<#Quietschen> z.B. Stuhlquietschen
<#> Geräusche, die in keine der anderen Kategorien passen

Nichtartikulatorische Geräusche werden durch vorangestelltes # in spitzen Klammern verschriftet.

Sie können auch andere verbale Produktionen oder nichtartikulatorische Geräusche überlagern.

vor <#>.. darf stehen:

(
< :
Leerstelle

Geräuschüberlagerung:

Beispiel:

```
... de<Z>m <:<#Klicken> neunundzwanzigsten:>
  August ...
... bei <:<Lachen> mir:> terminlich sehr
  ung"unstig ...
... <:<#Klopfen> <Lachen> k"amen mir die Monate:>
  April ...
... <:<Lachen> ich habe:> <:<#Klopfen> <Lachen>
  heute:> keine Zeit ...
```

Symbol: <:...:>

Nichtartikulatorische Geräusche oder nonverbale artikulatorische Produktionen, die Wörter oder andere Ereignisse überlagern, werden zusammen mit den Wörtern/Ereignissen mit <: :> geklammert.

Nach <: folgt das überlagernde Geräusch, eine Leerstelle und die überlagerte Produktion. :> schließt die Klammerung und wird nach dem Überlagerten ohne Leerstelle gesetzt.

Bei mehreren überlagernden Ereignissen gilt folgende Reihenfolge:

- nichtartikulatorische Geräusche,
- artikulatorische nonverbale Produktionen,
- verbale Produktionen.

vor <: darf stehen:

(
Leerstelle

vor :> darf stehen:

/+
/-
-
<.. >
<#.. >
<h"as>
<A>
<Z>
<%>
<;T>
Großbuchstabe
Kleinbuchstabe

Überlappendes Sprechen (nur bei Dialogen ohne Knopfdruck):

Beispiel:

PEG004: wollen wir mal die Termine ausmachen ,
(f"ur die n"achsten@) Monate ?
CLK005: (@ja gerne).. gerne.

CLK087: nee , nicht den elften , den<Z>
(f"unfzehnten <A>@)..
PEG088: (@f"unfzehnten) und den
(achtundzwanzigsten@)..
CLK089: (@genau <A>)..

HAH008: nein , da mu"s ich zu einem (Besuch@) nach
(Leipzig@)..
TIS009: (@m).. (@aha)..

HAH008: nein, da mu"s ich zu einem (Besuch@) nach
Leipzig..
TIS009: (@m , aha)..

Symbol: (@ bzw. (, @) bzw.)

- sich überlappende Teile werden in runden Klammern verschriftet
- die Zuordnung überlappender Teile erfolgt nur wortweise
- @) steht bei Überlappung mit dem folgendem Beitrag

- bei Überlappung mit dem vorhergehendem Beitrag steht (@ vor dem Überlappendem
- Äußerungen von der Art "mhm", "aha", "m", "ja", "genau", "stimmt", "völlig richtig" usw. werden im Falle des Überlappens mit anderen Beiträgen nur in runde Klammern gesetzt.
- mehrere solcher Äußerungen in einer Sprecherzeile werden als eine Sequenz von Klammerpaaren gesetzt, die durch Punkte voneinander getrennt sind.
- mehrere solcher Äußerungen, die in kurzer Abfolge auftreten, werden durch Kommata voneinander getrennt und in ein Klammerpaar gesetzt
- finden sich in zwei aufeinanderfolgenden Dialogbeiträgen mehrere Überlappungsstellen

zwischen diesen, so gilt die Reihenfolge ihres Auftretens:

- erstes Klammerpaar des ersten Beitrags (...@)
- erstes Klammerpaar des zweiten Beitrags (@....) bilden eine Überlappungsstelle

vor (bzw. (@ darf stehen:

Leerstelle

vor) bzw. @) darf stehen:

:>

/+

/-

<Ger"ausch>..

<#>..

<h"as>

<A>

<P>

<Z>

<%>

<iT>

Großbuchstabe

Kleinbuchstabe

3. Literatur:

K.Kohler , G.Lex , M.Pätzold , M.Scheffers , A.Simpson, W.Thon: Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL -3.0 , Verbmobil-Report, Nr. 11 , September 1994.

A. Batliner, S. Burger, A. Kießling: Außergrammatische Phänomene in der Spontansprache: Gegenstandsbereich, Beschreibung, Merkmalinventar, Verbmobil-Report, Nr. 57, Februar 1994.

S.Burger, E.Kachelrieß: Aussprachevarianten in der VERBMOBIL-Transliteration - Regeln zur konsistenteren Verschriftung, Verbmobil-Memo, Nr. 111, August 1996.

The VERBMOBIL Project August 1995

*Copyright © 1995 Institut für Phonetik und Sprachliche Kommunikation, Universität München
Diese und alle weiteren hier referenzierten Seiten, welche von diesem WWW Server stammen, dürfen kopiert, gedruckt
und an Dritte weitergegeben werden, unter der Bedingung, daß jede Seite vollständig bleibt. Weitergabe von Auszügen
nur mit Erlaubnis des Urhebers.*

Susanne Burger burger@phonetik.uni-muenchen.de

Anhang B – Transliterationslexikon (englisch)

Lexicon of the Conventions for Transliteration of Spontaneous Speech

Susanne Burger

Institut für Phonetik und sprachliche Kommunikation
München

(Lexicon version of the "Handbuch zur Datenaufnahme und Transliteration in TP14 von
VERBMOBIL -3.0"

Kohler, Lex, Pätzold, Scheffers, Simpson, Thon
Kiel, September 1994)

A Postscript-Version of the Handbuch is available at the IPDS Kiel.

Contents

1. General remarks on the transliteration

1.1 Punctuation

1.2 File structure

2. Symbols

3. References

1. General remarks on the transliteration:

All dialogues recorded in **Verbmobil** are transliterated

- in a uniform format
- according to the German Duden (with exceptions, e.g. reduced wordforms)

- with help of defined conventions supporting the transliteration of spontaneous speech.

German umlauts and Sz are transliterated according to the LaTeX-Format.

1.1 Punctuation:

Only **period**, **comma** and **question mark** are allowed. Contrary to Duden conventions, small letters are used after periods and question marks. Nouns are always written with capital letters at the beginning of the word. There is always a blank before and after punctuation marks. In general, punctuation in spontaneous speech is not easy because sentences are rarely grammatically well-formed.

Period

Symbol: " . "

Location: at the end of a sentence; between blanks

Example:

```
... so , guten Tag , mein Name ist <!1 is'> J"ansch .  
<"ah> wir hatten bereits telefoniert<Z> . ...
```

Note:

When in doubt a decision should be based on:

- grammar
- intonation
- pause, breathing
- begin of a new topic

Question mark

Symbol: " ? "

Location: at the end of a sentence; between blanks

Example:

```
... <A> wie schaut 's denn aus , den darauffolgenden  
Sonntag , den <:<#Mikrobe> neunundzwanzigsten:>  
bei Ihnen ? geht 's da ? ...
```

Note:

When in doubt a decision should be based on:

- interrogatives
- syntax
- intonation
- context

Comma

Symbol: ", "

Location: between sentence parts, subordinate clauses; between blanks

Example:

```
... <"ahm> morgen , Freitag , <h"as> wie ich seh' ,
    <"ah> mu"s ich feststellen , da"s ich <"ah>
    "uberhaupt keine Zeit hab' . ...
```

Note:

When in doubt a decision should be based on:

- grammar
- initiative particles at the beginning of the subordinate clause
- intonation
- related topics

e.g.:

- I could , I've always time on Wednesdays .
- I could . <P> <A> I've always time on Wednesdays .

Symbols allowed before blank and punctuation mark:

capital letter

small letter

:>

<A> (only in case of exhalation!)

<Z>

<%>

<!n ..>

<i ..>

1.2 File structure:

A transliteration file consists of a **header**

Example:

```
; Dialog N057K
; zuletzt bearbeitet am 23.5.94
; Tonqualit"at: (allgemeine Kommentare zu
; Sprechern oder zur Aufnahmequalit"at des
; Dialogs)
```

and a **text part**,
which is separated from the header by an empty line.

Every turn of the dialogue is transliterated in the text part. A turn starts with the speaker code, containing three capital letters and a turn number, starting at 000. After the turn number there has to be a colon, followed by a blank.

Now the text part begins. Words are not hyphenated at the end of a line, every new line within a turn begins with eight blanks. There may be a comment referring to the previous turn in the line after a turn. To mark the beginning of each comment, a semicolon is set before the comment. The turns are separated by empty lines.

Example:

```
AAP000: so , guten Tag , mein Name ist <!1 is'>
      J"ansch . <"ah> wir hatten bereits
      telefoniert<Z> , mein Name J<Z>"ansch , $J
      $"A $N $$ $C $H , wegen <:<#Mikrobe> eines
      Arbeitstreffens:> .
```

```
BBP001: gr"u"s Gott , mein Name ist G<Z>"urtner , <A>
      <"ahm> $G $"U $R $T $N $E $R <A> . <A> <"ahm>
;Brummen "uber gesamtem Turn
```

```
AAP002: ja , <:<#Mikrobe> ich kuck' jetzt:> mal nach
      bei mir , wann ich <A> einen <!1 ein'> Termin
      frei h"atte . <A> das <:<#Mikrowind> erste:>
      w"are <Schmatzen> in der Woche oder die Tage
      vom vierzehnten Juli bis zum achtzehnten
      <:<#Mikrobe> Juli:> . <A> ginge das bei <:<#>
      Ihnen:> ?
```

```
BBP003: ...
```

2. Symbols:

About the lexicon items

symbol.....meaning

```
$.....spelling
*.....non-word/mispronunciation
<"ah>.....
<"ahm>.....
<hm>.....
<h"as>.....hesitation/filled pause
<Schmatzen>.....
<Schlucken>.....
<R"auspern>.....
<Husten>.....
<Lachen>.....
<Ger"ausch>.....nonverbal articulatory noises
<Z>.....lengthening
<A>.....breathing
<P>.....pause/silence
li_blank.....word interruption: left boundary
                                  (underscore,blank)
blank_re.....word interruption: right boundary
                                  (blank,underscore)

=.....word fragment
/-.....truncation
+/-.....self-repairs: left boundary
/+/.....self-repairs: right boundary
%.....words which are
                                difficult to understand
<%>.....unintelligible speech
<i..>.....comment
<iT>.....technical break
<! ..>.....variants of pronunciation
<#Klicken>.....
<#Klingeln>.....
<#Klopfen>.....
<#Mikrobe>.....
<#Mikrowind>.....
<#Rascheln>.....
<#Quietschen>.....
<#>.....non-articulatory noises
<:.....beginning of overlay
:>.....end of overlay
(.....beginning of overlapped speech
                                ( and (@
).....end of overlapped speech
                                ) and @)
```

About the lexicon items

There are a lot of special symbols in the text part of the TRL-files. These symbols have been defined for a better transliteration of the phenomena of spontaneous speech.

In the following a description of all used symbols can be found. There are always short examples from the original TRL-files.

A definition of a symbol consists of

- the name of the phenomenon,
- an example,
- the symbol used and instructions for its use,
- a list of symbols, which are permitted to be set immediately before the described symbol.

Spelling:

Example:

```
... mein Name ist J<Z>"ansch , $J $"A $N $S $C $H ...
```

Symbol: \$

Every spelled letter is written as a capital letter and separated by a blank from the following spelled letter.

\$ is set before the capital letter. There is no blank between \$ and the spelled letter.

Symbols allowed before \$:

```
(  
+/  
-  
blank
```

Non-words:

Example:

```
... *haarknapp <:<#Rascheln> um einen <!1 ein'> Tag  
verfehlt ...  
... was *exkursieren Sie denn ? ...
```

Symbol: *

non-words are:

- neologisms,
- mispronunciations,
- words of foreign languages, which not can be found in the German Duden ,
- foreign names.

A non-word is marked with a * . There is no blank between the asterisk and the non-word.

Symbols allowed before * :

(
+/
< ; T >
blank

Hesitations:

Example:

... <"ah> wir hatten bereits telefoniert<Z> ...
... gu<Z>t , <"ahm> wie w"ar' es bei Ihnen am<Z>
neunzehnten Juli ...
... ich denke/- <hm> also bei mir ginge es sehr
gut ...
... w"urd' ich sagen , <A> <h"as> wenn wir ...

Symbol: <"ah> <"ahm> <hm> <h"as>

Hesitations are transliterated between pointed brackets.

- <"ah>: vocalic articulation, independent of the vowel quality
- <"ahm>: vocalic articulation + nasal articulation
- <hm>: nasal articulation
- <h"as>: rare articulations, which can not be categorized by one of the other classes.
e.g. /brrt/ /pf/ /puh/ etc.

Symbols allowed before <h"as> .. :

(
< ; T >
blank

Articulatory noises:

Example:

... Mittagessen <P> <Schmatzen> und<Z> , na ja ...
... ordentlich planen mu"s und <"ah> <Schlucken> wann
w"urde Ihnen ...
... einverstanden . <P> <R"auspern> auf
Wiedersehen ...
... +/je/+ <Husten> je schneller wir das machen ...
... bei <:<Lachen> mir:> terminlich sehr
ung"unstig ...
... dann/- <"ah> <A> <Ger"ausch> das sind sechs
Termine ...

Symbol: <Schmatzen> <Schlucken> <R"auspern> <Husten> <Lachen> <Ger"ausch>

- <Schmatzen>: munching
- <Schlucken>: swallowing
- <R"auspern>: throat clearing
- <Husten>: coughing
- <Lachen>: laughing
- <Ger"ausch>: articulatory noises, which can not be categorized by one of the other classes.

Articulatory noises are transliterated at the time of their production between blanks.
Other verbal and nonverbal productions can be overlaid with articulatory noises.

Symbols allowed before <Ger"ausch> .. :

(
<:
blank

Lengthening:

Example:

... ich h"atte Zeit +/am<Z>/+ <A> <Schmatzen> ab
Dienstag ...
... ich dachte ger<Z>ade ...

Symbol: <Z>

<Z> is added immediately after the letter(s) representing the lengthened sound, regardless of the place within the word where the lengthening occurred.

Symbols allowed before <Z> :

small letter
capital letter

Breathing:

Example:

... ist G<Z>"urtner , <A> <"ahm> \$G \$"U \$R \$T \$N
\$E \$R <A> . <A> ...

Symbol: <A>

Most of the time <A> marks inhalation and is thus transcribed after the punctuation mark. In case of clear exhalation at the end of a sentence or sentence part, breathing also can be marked before the punctuation mark.

Symbols allowed before <A> :

(
blank

Pauses:

Example:

... ich habe <P> M"oglichkeiten dazu ...

Symbol: <P>

In case of a coincidence of <P> and punctuation mark, the pause is set after the punctuation mark.

Symbols allowed before <P> :

(
blank

Word interruption:

Example:

... Ver_ <A> _pflichtungen ...
... statt_ +/f=/+ <h"as> _findet ...

Symbol: _

Sometimes words are interrupted.

Here an underscore should be set immediately after the interruption without a blank in between, and also an underscore immediately before the continuation.

All events occurring during the interruption are transliterated between blanks and between the word parts.

Symbols allowed before li_blank (left underscore) :

-
<Z>
capital letter
small letter

Symbols allowed before blank_re (right underscore) :

blank

Word fragments:

Example:

... +/sieb=/+ siebzehnter ...
... im Ja=/- also ich sag' Ihnen jetzt ...

Symbol: =

In case of a complete break-off of a word without continuation the break-off position is marked with an = added at the break-off. There is no blank between the break-off and the equals sign.

Symbols allowed before = :

<Z>
capital letter
small letter

Truncation:

Example:

... ab dem dritten August <A> bis zum/- <P> Moment ,
ich ...
... ja , ich hab' da eigentlich/- also ich bin vom
neunzehnten bis ...

Symbol: /-

Truncation refers to the case in which a topic has been broken off while speaking and a new topic is initiated after the break-off.

The break-off position is marked with /- and without a preceding blank.

There is no punctuation mark after /- .

Symbols allowed before /- :

=

<A> (only in case of exhalation!)

<Z>

<;T>

capital letter

small letter

Self-repairs:

Example:

... +/am/+ <: <#Mikrobe> am Donnerstag:> kann ich
erst ...
... +/im Sep=/+ im September ...
... die Woche +/von/+ <"ah> mit Freitag ...
... also +/das/+ das/+ das zweite ...

Symbol: +/./+

In case of self-repairs after the break-off the reparandum is either repeated or corrected.

The break-off position is marked with /+ and without blank in between .

The utterance which is to be repeated/corrected after the break off starts with +/ and without a following blank, so that +/./+ put the reparandum in brackets.

Symbols allowed before +/ :

(

blank

Symbols allowed before /+ :

=
<A> (only in case of exhalation!)
<Z>
<;T>
capital letter
small letter

Words which are difficult to understand:

Example:

... %eins , %zwei , %drei , %vier ...
... wann h"atten Sie %da bitte Zeit ...

Symbol: %

Words which are difficult to understand have a % before without a blank in between.

Symbols allowed before % :

(
+/
<;T>
blank

Unintelligible speech:

Example:

... aber <A> <"ah> <%> <"ahm> wie w"ar's denn ...
... <%> fr"uher geht 's leider nicht ...

Symbol: <%>

unintelligible items are transliterated with <%> between blanks.

Symbols allowed before <%> :

(
blank

Comments:

Example:

```
; Tonqualit"at: viele Nebenger"ausche ...
;gesamter Turn verrauscht ...
... zum Beispiel <;"ubersteuert> ...
... einen Termin vereinbaren <;heiser> ...
```

Symbol: ;

A comment relating to the whole dialogue is written in front of the dialogue in the header. A comment line starts with a semicolon. After the comment follows an empty line.

A comment about a turn is prefixed with a semicolon at the beginning of the line and added after the contribution without an empty line.

Symbol: <;..>

Local comments are inserted in the text after the concerning position and after a blank, prefixed with a semicolon and set in pointed brackets.

Symbols allowed before <;..> :

blank

Technical break:

Example:

```
... ab ein U<;T> ...
... <;T>neunzehnten is<Z>t etwas schwierig ...
... erste Woche ginge <A> <;T> <#Klicken> <A>
  <"ah> Montag ...
... unsere Termine alle untergebra<;T> <;T>ke ...
```

Symbol: <;T>

Depending on when the recording button is activated a speech recording may be truncated at the beginning or at the end, but may be also interrupted during a dialogue contribution.

In case of a technical disturbance during a lexical unit, <;T> is added before or after the unit without a blank in between. A blank is placed between two consecutive <;T>.

If no lexical unit is involved <;T> is written between blanks.

Symbols allowed before <;T> :

< :
+/
<h"as>
<Z>
capital letter
small letter
blank

Pronunciation variants:

Example:

... damit w"ar' das <!1 des> eigentlich klar ...
... dann kommen Sie <!2 komm' Se> doch ...
... wenn wir <!2 wemma> 's die Woche noch machen ...

Symbol: <!n ..>

Dialectal variants, different speaking styles or other deviations are transliterated in correct orthography according to the German Duden. The actually produced utterance is transliterated after the correct version and after a blank. This utterance is placed in pointed brackets. An exclamation mark and a following digit giving the number of lexical units concerned and a further blank initiates the pronunciation variant.

A more detailed description of the regularities for the transliteration of pronunciation variants at the orthographic level can be found in:

Verbmobil-Memo 111: Aussprachevarianten in der VERBMOBIL-Transliteration - Regeln zur konsistenteren Verschriftung (Burger, Kachelrieß, München, August 1996).
(<http://www.phonetik.uni-muenchen.de/VMTechDoks.html>)

Symbols allowed before <!n ..> :

blank

Non-articulatory noises:

Example:

... de<Z>m <:<#Klicken> neunundzwanzigsten:>
August ...
... <:<#Klingeln> am Mittwoch st"anden:> ...
... <:<#Klopfen> gern so machen:> ...
... <#Mikrobe> das tut mir leid ...
... ich k"onnte Ihnen <:<#Mikrowind> vorschlagen:> ...

... nach Berlin <P> ausmachen . <#Rascheln> <"ahm> ...
... der gesamte <:<#Quietschen> Rest des Mais:> ...
... k"ame mir gelegen , <A> <#> allerdings ...

Symbol: <#Klicken> <#Klingeln> <#Klopfen> <#Mikrobe> <#Mikrowind>
<#Rascheln> <#Quietschen> <#>

<#Klicken>: the noise of pushing the record button

<#Klingeln>: telephone ring

<#Klopfen>: e.g. knocking on the table

<#Mikrobe>: touching the microphone

<#Mikrowind>: blowing into the microphone

<#Rascheln>: e.g. paper rustle

<#Quietschen>: e.g. chair squeaking

<#>: non-articulatory noises, which can not be categorized by one of the other classes.

Non-articulatory noises are transliterated with a prefixed # and in pointed brackets.

Other verbal and nonverbal productions can be overlaid with non-articulatory noises.

Symbols allowed before <#>.. :

(

<:

blank

Overlay:

Example:

... de<Z>m <:<#Klicken> neunundzwanzigsten:>
August ...
... bei <:<#Lachen> mir:> terminlich sehr
ung"unstig ...
... <:<#Klopfen> <Lachen> k"amen mir die Monate:>
April ...
... <:<#Lachen> ich habe:> <:<#Klopfen> <Lachen>
heute:> keine Zeit ...

Symbol: <:...:>

Non-articulatory noises or nonverbal articulatory productions overlaying lexical units or other events are set in brackets together with the concerned lexical units or the concerned events.

Behind <: follows the overlay, a blank and the overlaid production. :> closes the overlaid utterance. There is no blank between the overlaid production and the closing bracket.

In case of several overlaias the following order of events has to be obeyed:

- non-articulatory noises,

- articulatory noises,
- lexical units.

Symbols allowed before <: :

(
blank

Symbols allowed before :> :

/+
/-
-
<.. >
<#.. >
<h"as>
<A>
<Z>
<%>
<;T>
capital letter
small letter

Overlapping speech (only for dialogues without button pushing):

Example:

PEG004: wollen wir mal die Termine ausmachen ,
(f"ur die n"achsten@) Monate ?

CLK005: (@ja gerne).. gerne.

CLK087: nee , nicht den elften , den<Z>
(f"unfzehnten <A>@)..

PEG088: (@f"unfzehnten) und den
(achtundzwanzigsten@)..

CLK089: (@genau <A>)..

HAH008: nein , da mu"s ich zu einem (Besuch@) nach
(Leipzig@)..

TIS009: (@m).. (@aha)..

HAH008: nein, da mu"s ich zu einem (Besuch@) nach
Leipzig..

TIS009: (@m , aha)..

Symbol: (@ bzw. (, @) bzw.)

- overlapping speech is transliterated in parentheses
- the assignment of overlapping parts occurs word for word
- @) marks the overlap with the following speaker contribution
- an overlap with the preceding speaker contribution is marked with (@
- productions like "mhm", "aha", "m", "ja", "genau", "stimmt", "völlig richtig" etc are only put in parentheses in case of overlap with other contributions.
- several productions of those in one speaker line are transliterated as a sequence of parentheses which are separated from each other with periods.
- several productions of those occurring in a short time course are separated with commas from each other and combined in one pair of parentheses.
- if several overlaps are found in two consecutive dialogue contributions, the following order should be obeyed:
 - first pair of parentheses of the first dialogue contribution (...@)
 - and first pair of parentheses of the second dialog contribution (@....) mark an overlap.

Symbols allowed before (resp. (@ :

blank

Symbols allowed before) resp. @) :

:>

/+

/-

<Ger"ausch>..

<#>..

<h"as>

<A>

<P>

<Z>

<%>

< ;T>

capital letter

small letter

3. References:

K.Kohler, G.Lex, M.Pätzold, M.Scheffers, A.Simpson, W.Thon: Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL -3.0 , Verbmobil Report , Nr. 11 , September 1994.

A.Batliner, S.Burger, A.Kießling: Außergrammatische Phänomene in der Spontansprache: Gegenstandsbereich, Beschreibung, Merkmalinventar, Verbmobil-Report , Nr. 57 , Februar 1994.

S.Burger, E.Kachelrieß: Aussprachevarianten in der VERBMOBIL-Transliteration - Regeln zur konsistenteren Verschriftung, Verbmobil-Memo, Nr. 111, August 1996.

The VERBMOBIL Project August 1995

*Copyright © 1995 Institut für Phonetik und Sprachliche Kommunikation, Universität München.
Parts of this document are Copyright © Institut für Phonetik und digitale Sprachverarbeitung der
Christian-Albrechts-Universität zu Kiel.*

*This page and all other pages which belonged to this server may be copied, printed and distributed to other parties,
under the condition that the pages are distributed as shown here. Parts of pages or extended pages may not be
distributed further without permission of the author.*

Susanne Burger burger@phonetik.uni-muenchen.de