Gefördert vom

Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Richtlinien für die Erstellung der deutschen Erkennerwortlisten – SmartKom

Version 1

Nicole Beringer

Ludwig Maximilians Universität München

Technisches Dokument Nr. 9

August 2000

August 2000

Nicole Beringer

Ludwig Maximilans Universität München
Schellingstr. 3
80799 München

Tel.: (089) 2180−5751
FAX: (089) 2800362

E−Mail: beringer@phonetik.uni−muenchen.de

Dieses Technische Dokument gehört zu Teilprojekt 1: Modalitätsspezifische Analysatoren

Das diesem Technischen Dokument zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01 IL 905 gefördert. Die Verantwortung für den Inhalt liegt bei den Autoren.

Inhaltsverzeichnis

1 Vorwort

Auf Grund verschiedenster Quellen für die deutsche Erkennerwortliste ist es notwendig, daß alle Wortlistenproduzenten dieselben orthographischen Richtlinien befolgen. Das vorliegende Technische Dokument gibt eine Übersicht über die relevanten Marker aus den Transliterationskonventionen ([1], [2], [3], [7]) .

Ergänzend dazu gibt es einige Neuerungen bzw. Modifikationen bei den Komposita.

Alle Produzenten sollten sich an vorliegende Richtlinien halten, da sonst ihre deutsche Wortliste nicht verwaltet werden kann.

Das vorliegende Dokument ist eine α– Version.

Die Ergebnisse des Wortlisten–Workshops am 21.08.2000 werden als Update eingebunden.

2 Orthographische Besonderheiten

2.1 Rechtschreibung

Alle Wörter der Wortliste werden nach gängigen Rechtschreibregeln erstellt. Für Deutsch ist die alte Rechtschreibung nach Duden gültig. Fremdsprachige Wörter werden – sofern sie nicht bereits im Duden erscheinen – nach der gültigen Rechtschreibung der jeweiligen Sprache verschriftet und gesondert markiert.

Für die Transliteration deutscher Wörter gilt als orthographische Referenz der "alte" Duden, 20. Auflage.
Die deutsche Rechtschreibreform wird noch nicht beachtet!

Groß– und Kleinschreibung:
Der Duden gilt auch hinsichtlich Groß– und Kleinschreibung gemäß der Wortklasse. Ausnahmen bilden Akzente, wie z. B. in franz. »Aimée und Jaguar«. Diese werden bei der Verschriftung weggelassen.

Vom deutschen Duden etwas abweichend werden behandelt:
Reduzierte Wortformen,
Interjektionen,
Komposita,
Abkürzungen und Buchstabierungen,
Namen,
Zahlen

Dialektale Wörter:
Finden sich in einem Dialog dialektale Wörter,
– die keine dialektale Aussprache von im Duden aufgelisteten Wörtern darstellen,
– aber auch noch nicht in den Duden aufgenommen wurden,
– wie: "ja mei" , "grüß Gott", "moin, moin", "Herrschaftszeiten" , "fei",
so werden diese ohne besondere Kennzeichnung orthographisch verschriftet und optional mit Aussprachekommentar und / oder Kommentar zur Herkunft oder Bedeutung versehen.

Aussprachekommentare werden nicht in die Wortliste aufgenommen.

2.2 ASCII–Kodierung

Die Kodierung der Transliteration erfolgt in **7–bit ASCII**. Für die Umlaute und "ß" wird im Ablieferungsformat für die Wortliste des ersten Demonstrators die **TEX–Schreibweise** verwendet, also "a, "U, "s etc.

z.B. gr "un anstelle von gruen, grün etc.

Diese Schreibweise wird allerdings zugunsten der ISO–Schreibweise bei zukünftigen Wortlisten aufgegeben.

3 Gobale Kommentare

Symbol: **;........**

Beispiele:

```
;Adjektive
```

```
;Komposita
```

Die Wortliste kann optional einen globalen Kommentar enthalten.

Dieser sollte allerdings mit anderen Markern nicht kollidieren und wird daher mit »;« gekennzeichnet.

Der globale Kommentar steht in einer **neuen Zeile**, die **mit Semikolon** beginnt.
Jede weitere Zeile des globalen Kommentars beginnt ebenfalls mit Semikolon.

3.1 Lexikalische Einheiten

Alle Markierungssymbole werden ohne Leerstelle je nach entsprechender Konvention direkt davor, innerhalb der lexikalischen Einheit oder direkt dahinter transliteriert.
Lexikalische Einheiten können nur einmal bezüglich ihrer Wortklasse (Namen, Zahlen) markiert werden (keine Doppelklassifizierung).

3.1.1 Interjektionen

Symbol:

(nicht markiert)

Beispiele:

```
oh
mm
ne
```

Definition:

Interjektionen sind Ausrufe der Überraschung wie "au", "ah", "oh", "ui", "he",
Bejahung wie "mhm" ,
oder Verneinung "mm",
Einwürfe oder Bestätigungen wie "ne" , "gell", "aha".

Transliterationskonvention:

Interjektionen werden ohne Zusätze in der in der Definition verschrifteten Form in den orthographischen Text eingefügt. Die meisten von ihnen sind bereits im Duden aufgelistet. Für Interjektionen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkung:

Verneinendes "m´m" wird mit "mm" verschriftet, im Gegensatz zu bejahendem "mhm".

Interjektionen sind werden für den ersten Demonstrator in die Wortliste aufgenommen.

3.1.2 Code–Wörter

falls zur Steuerung des SmartKom ein Codewort benötigt wird:

Symbol:

!KEY

Beispiele:

`!KEYMarvin`

`!KEYSmartakus`

`!KEYAladdin`

Definition:

Um das SmartKom per Stimme zu steuern, sei es, daß z.B. Äußerungsanfang und Äußerungsende akustisch gekennzeichnet werden sollen, eine Teiläußerung übersetzt werden oder SmartKom ein Gesprächsprotokoll anfertigen soll, kann ein sogenanntes Codewort benutzt werden, das nicht in die Übersetzung mit eingeht, sondern eine Funktion des SmartKoms aufruft.

Transliterationskonvention:

Wörter, die nur als Systemsteuerelement dienen, werden mit vorangestelltem !KEY markiert.

Code–Wörter werden wie lokale Kommentare behandelt und nicht wie lexikalische Einheiten. Sie können auch ohne Interpunktionszeichen am Turnende stehen.

3.1.3 Komposita

deutsch

Symbol:

– (teilweise mit Bindestrich)

`Acht–Uhr–Maschine`

Definition:

Komposita sind aus mehreren Wörtern gebildete zusammengesetzte Wörter.

Transliterationskonvention: Einfache Komposita werden nach Duden verschriftet.

Bei mehr als Zwei–Wort–Komposita, ungewöhnlichen Zusammensetzungen und Kombinationen mit buchstabierten Einheiten, Namen oder Zahlen stehen Bindestriche zwischen den Kompositateilen.
Acht–Uhr–Maschine
M"anner–Selbsterfahrungs–Gruppe
Hauptbahnhofs–Eingangshalle
Berlin–Video

Abkürzungen:
Bei Abkürzungen, die aus buchstabierten Einheiten bestehen, werden zwischen den buchstabierten Einheiten Bindestriche gesetzt:
~$I–$B–$M
$U–$S–$A
$U–Bahn

Um verschiedene Schreibweisen zu vermeiden, kann in Zweifelsfällen die SmartKom–Wortliste zu Rate gezogen werden.

Für Komposita gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkungen:

Bei Wortunterbrechung zwischen Kompositateilen steht der Bindestrich vor dem wiederaufnehmenden Wortteil (nach der Unterbrechung).

Sonderfall:

Kompositumsergänzung:
Bei zusammengesetzten oder abgeleiteten Wörtern, bei denen ein gemeinsamer Bestandteil nur einmal genannt wurde, steht **doppelter Bindestrich (––)**:
Geld–– und andere Sorgen,
Hin–– und Rückfahrt,
ein–– bis zweimal,
Lederherstellung und ––vertrieb

Das mit Bindestrichen **endende** Wort kann **nicht artikulatorisch oder technisch abgebrochen** sein,
wird aber ansonsten wie eine lexikalische Einheit behandelt.
Das mit Bindestrichen **beginnende** Wort kann **nicht zusätzlich klassifiziert** sein, wird aber ansonsten auch nach den Konventionen lexikalischer Einheiten verschriftet.

deutsch

3.1.4 Unkonventionelle Komposita

Symbol:
+
Definition:
Alle Mehrwortlexeme, wie Filmtitel, Hotelnamen etc.

Unter unkonventionellen Komposita fallen alle zusammengesetzten Namen:
Die aus mehreren Wörten bestehenden Namen wie
~Zur+blauen+Traube
~Bu"s–+und+Bettag
~Heilig+Drei+K"onig
werden mit Bindestrichen zusammengeschrieben und als Namen markiert. Die Groß/Kleinschreibung der Namensbestandteile wird in diesem Fall beibehalten. Sowie

Filmtitel etc.

~<*ENG>Eyes+wide+shut

~Agent+Null+Null+Nix
~Aimee+und+Jaguar
~<*ENG>Air+<*ENG>Bud+<*ENG>Champion+auf+vier+Pfoten
~<*ENG>Air+Force+One
~Akte+$X+Der+Film

3.1.5 Buchstabierung und Abkürzungen

Symbol:

$ (bei ausgesprochenen Buchstaben)

Beispiele:

$J

$"A

$N

$S

$C

$H

$U-$S-$A

$A-$B-$R-Filiale

$H $A doppel-$M $E $R

Definition:

Als Buchstabierung gelten ausgesprochene Buchstaben, etwa zur Verdeutlichung der Schreibweise eines Namens oder bei Abkürzungen, bei denen die Buchstaben einzeln ausgesprochen werden (z.B. USA).

Abkürzungen, die als Wort ausgesprochen werden, werden nicht als Buchstabierung behandelt (z.B. OPEC–Länder, CeBIT–Messe).

Transliterationskonvention:

Jedem ausgesprochenen Buchstaben (Buchstabiereinheit) wird $ vorangestellt. Als reine Buchstabiereinheiten gelten Buchstabierungen von Namen oder Wörtern, deren korrekte Schreibweise dem Gesprächspartner verdeutlicht werden soll.

Buchstabiersequenz:
Werden Buchstaben in Abkürzungen oder Komposita ausgesprochen, so wird dem jeweiligen Buchstaben ebenfalls $ vorangestellt. Die zusammengehörigen Buchstabiereinheiten oder Komposita–Teile werden mit Bindestrich aneinandergehängt und wie Komposita behandelt.
$U–$S–$A
$U–Bahn
scharf–$S
$S–$Z
$A–$B–$C–Filiale

3.1.6 Akronyme

Symbol:

&

Definition:

Als Wörter ausgesprochene Abkürzungen werden nach üblicher Schreibweise festgehalten und wie lexikalische Einheiten behandelt.
&DIN
&Benelux
&OPEC
&AStA
&CeBIT

Für Buchstabierungen und Abkürzungen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkungen:

ACHTUNG:
&$F–$A–$Z für gesprochenes "eff a zet"
aber: &FAZ für gesprochenes "Faz"

3.1.7 Namen

Symbol:

~

Beispiele:

```
~Eyes-wide-shut
~Arri
```

Definition:

Markiert werden alle Namen, die nicht in eine anderere Sprache übersetzt werden sollen.

- Personennamen (auch Vornamen)
- Straßennamen
- Hotel– und Restaurantnamen
- Instituts– und Firmennamen
- Ortsnamen

Nicht markiert werden:

- Feiertage, sofern sie keinen eigenen Namen haben (also ~Heilig–Drei–K"onig, aber Maifeiertag
- Namen von Staaten (nicht Bundesstaaten oder Bundesländer), z. B. Frankreich, England
- Kontinente, z. B. Amerika, Afrika
- Namen von Weltmeeren, z. B. Atlantik, Pazifik

Transliterationskonvention:

Namen werden mit vorgestelltem ~ markiert.
Bei Namen, die aus mehreren Wörtern bestehen, werden die Namensbestandteile mit Bindestrichen aneinandergehängt (z.B. ~Eyes–wide–shut).
Ist ein Name nur Teil eines ansonsten aus anderen Wörtern bestehenden Kompositums, wird der Name nicht markiert (z.B. Berlin–Video, Elisabeth–Kirche).
Für Namen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

3.1.8 Zahlen

Symbol:

#

Beispiele:

```
#sechzehn
#neunzehn
```

Definition:

Als Zahlen gelten reine Zahlwörter oder Zahlenkombinationen, auch Ordnungszahlen.

Transliterationskonvention:

Zahlen und zusammengesetzte Zahlen werden mit vorangestelltem # markiert.

Zusammengesetzte Zahlen
Die Zahlen von 13 bis einschließlich 99 (zweistellige Zahlen) und
die von "ein" bis "neunzehn" gezählten Hunderter (d.h. einschließlich Jahreszahlen wie neun−zehn−hundert), Tausender, usw.
werden nicht mehr wie in Verbmobil mit Bindestrichen zusammengeschrieben.
#zweiundzwanzig
#dreizehnter
#einhundert #f"unfundzwanzig

Alle übrigen Zahlenkombinationen werden mit Leerzeichen (oder Zeilenumbruch−Leerzeichen) getrennt (dreistellige Zahlen und größer):
#neunzehnhundert #dreiundneunzig
#drei #Millionen #neunundvierzig #tausend #sechshundert #vierzehn
#dreihundert und #neunundvierzig

Ist eine Zahl nur Teil eines ansonsten aus anderen Wörtern bestehendem Kompositums, wird die Zahl **nicht** markiert.
Acht−Uhr−Maschine
Fünf−Tages−Seminar

Für Zahlen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

3.1.9 Neologismen

Symbol:

*****

Beispiele:

`*haarknapp`

`*verschraubel`

`*exkursieren`

`*Diaabend−Weintrink−Revisionstreffen`

Definition:

Als Nichtwort gelten:

- Neologismen (Wortneuschöpfungen)
- unsinnige Wortkombinationen
- Versprecher, die zu Neologismen führen

Transliterationskonvention:

Einem Neologismus wird * ohne Leerstelle vorangestellt.
Unwörter können **nicht** artikulatorisch abgebrochen sein.

Für Neologismen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkungen:

Neologismen sollten nur im Rahmen der Transliterationswortliste in die allgemeine Wortliste eingehen.

3.1.10 Fremdsprachige Wörter

Symbol:

<*XXX>

wobei XXX:
ENG = englisch
DEU = deutsch (bei nicht–deutschen Dialogen oder mehrsprachigen Dialogen)

auch
FRA =französisch
ITA = italienisch
SPA = spanisch
JAP = japanisch
etc.

Beispiele:

```
<*ENG>touch
<*ENG>screen
<*JAP>sayonara
<*FRA>bien
<*FRA>c´est
<*FRA>la
<*FRA>vie
<*ITA>bella
```

Definition:

Fremdsprachige Wörter sind Wörter aus einer anderen Sprache als der im Dialogbeitrag mehrheitlich verwendeten.

Transliterationskonvention:

Fremdsprachige Wörter werden mit vorangestelltem <*ENG> (<*JAP>, <*ITA>, <*FRA>,<*GER>, ...) markiert, es sei denn, es handelt sich um Namen.

Ausnahmen: Fremdsprachige Namen werden nicht extra markiert!

Für fremdsprachige Wörter gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkungen:

ABER: Manche fremdsprachigen Wörter stehen z.B. im deutschen Duden und brauchen im Fall deutscher Dialoge dann nicht als fremdsprachig markiert werden.

tschau
Meeting

deutsch

ABER: fremdsprachige Wörter, die mit Apostroph geschrieben werden, sollten, auch wenn sie im deutschen Duden zu finden sind, als fremdsprachig markiert werden, weil sonst Konflikte mit Text−Filtern für deutsche Wort−Reduktionen auftreten können.
<*FRA>d´accord

deutsch

3.2 Transliterations−Wortliste

Wie bereits im Abschnitt 3.1.2.8. angedeutet wurde, gibt es »Wörter« in den Wortlisten, die weitere Marker erfordern.
Dazu gehören neben Neologismen auch schwerverständliche Wörter und abgebrochene lexikalische Einheiten.

3.2.1 Wortreduktionen

deutsch

Symbol:

´ (teilweise mit Apostroph)

Definition:

Wortreduktionen, die in die Transliteration mit eingehen, sind:
− End−e Reduzierung
− Reduktion unbestimmter Artikel
− Verschmelzung von Präposition und reduziertem bestimmten Artikel
− reduziertes Pronomen der 3. Pers. Sing. (es)

− reduzierte Flexionsendung

Transliterationskonvention:

Mit Apostroph werden transliteriert:

End−e Reduzierungen:
Für fehlendes End−e steht Apostroph.
Aber: wurde die Endsilbe −en nicht gesprochen, wird sie trotzdem verschriftet und ihr Fehlen durch einen Aussprachekommentar angezeigt.

is´
hab´
(aber: wir haben <!1 hab´> da keine Zeit)

m"u"st´
wollt´

unbestimmte Artikel:
Bei reduzierten unbestimmten Artikeln wird mit Apostroph verschriftet.
Aber: wurde z.B. statt einen nur "´n" gesagt, steht " einen <!1 ´n> ", d.h. der

Apostroph steht immer nur für fehlendes /ei/.
´nen
´ner
´n
´nem
´ne
´nes..

reduziertes Pronomen der 3. Pers. Sing. : es
´s

Nach einer transliterierten Reduktion kann kein Aussprachekommentar folgen. Ist die lexikalische Einheit so verändert, daß über die Reduktion hinaus ein Aussprachekommentar erforderlich ist, steht in der Transliteration die Vollform der betreffenden Einheit." Beispiel: FALSCH: hab' <!1 hob> RICHTIG: habe <!1 hob>.

Bei Lautveränderungen fällt der Reduktionsmarker weg.

Diese Wortformen stehen als jeweils eigene lexikalische Einheiten zwischen Leerstellen, **auch wenn zwei Apostrophe aufeinandertreffen.**
w"ar´ ´s
hab´ ´nen ..

3.2.2 Schwerverständliche Wörter

Symbol:

%

Beispiele:

```
#vier%
auch%
```

Definition:

Schwerverständlich sind alle Wörter, die nicht eindeutig verstanden werden können, weil sie durch Geräusche oder Artikulation akustisch schlecht gehört werden.

Dazu gehören auch dialektale oder artikulatorische Varianten, bei denen nur schwer auf die hochsprachige Version zurückgeschlossen werden kann.

Transliterationskonvention:

Verschriftet wird ein Wort, das der schwerverständlichen Äußerung ähnlich klingt und auch im Kontext plausibel erscheint. Dem in der üblichen Orthographie transliterierten schwerverständlichen Wort wird % ohne Leerstelle angehängt.

Für schwerverständliche Wörter gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Da diese »Wörter« innerhalb des Erkennerprozesses auch erkannt werden sollten, werden sie in die Erkennerwortliste aufgenommen, jedoch markiert.

Anmerkungen:

Schwerverständliche Wörter können nicht am Wortende abgebrochen sein.
Werden Wörter wegen dialektaler oder aussprachebdingter Variation schlecht verstanden, werden sie nach den Konventionen für schwerverständlich verschriftet. Ein Aussprachekommentar ist hier nicht nötig.
Wörter, die keinerlei Rückschlüsse auf hochsprachliche Formen zulassen und nicht als einem Dialekt zugehörig identifiziert werden können, werden nach der Konvention für unverständliche Sprachproduktionen verschriftet.

3.2.3 Abgebrochene lexikalische Einheiten

Definition:

Lexikalische Einheiten können artikulatorisch oder technisch abgebrochen oder unterbrochen sein und gehen als markierte Wortfragmente in die Transliteration mit ein.

Transliteration:

Die Abbruchstelle wird markiert, das verbleibende Wortfragment wird, solange sein Ursprung klar ist, in üblicher Orthographie nach den Konventionen für lexikalische Einheiten verschriftet.
Klassifizierungssymbole bleiben erhalten (z.B. bei abgebrochene Namen oder Zahlen).

Anmerkung:

Abgebrochene lexikalische Einheiten können nicht zusätzlich als schwerverständlich markiert werden.

3.2.3.1 Artikulatorischer Abbruch

Symbol:

=

Beispiele:

`da=`

`Ja=`

Definition:

Artikulatorisch abgebrochen heißt, der Sprecher selbst bricht eine lexikalische Einheit ab, meist um sich dann zu korrigieren.

Transliterationskonvention:

An ein artikulatorisch abgebrochenes Wort wird = ohne Leerstelle angehängt. Das Wortfragment wird bis zur Abbruchsstelle orthographisch dem wahrscheinlich zugrunde liegenden Wort entsprechend verschriftet.

Bei artikulatorischem Abbruch am Turnende wird statt des in diesem Fall unsinnigen Interpunktionszeichen ein Turnabbruch protokolliert, auch wenn der Turn nicht aus technischen Gründen abgebrochen wurde.
z.B. .. –/gr"u"s Gott , Herr<Z>/– wie war der Na= <*T>t

3.2.3.2 Artikulatorische Unterbrechung lexikalischer Einheiten

Symbol:

_ (Subline)

Beispiele:

```
... <:<#> Ver_:> <A> <:<#> _pflichtungen:> ...
... statt_ +/f=/+ <h"as> _findet ...
... Acht-Uhr_ <"ah> _-Maschine ..
... ~Zur-blauen_ +/Treb=/+ _-Traube ...
```

Definition:

Ein Wort kann durch Versprecher, eine Pause, Atmen oder durch Häsitationen unterbrochen sein. Nach den unterbrechenden Elementen wird das unterbrochene Wort fortgesetzt.

Transliterationskonvention:

An der Unterbrechungsstelle wird _ (Subline) ohne Leerstelle an das Wortfragment angehängt.
Danach werden nach einer Leerstelle (oder Zeilenumbruch–Leerstelle) die unterbrechenden Elemente transliteriert.
Nach dem letzten dieser Elemente folgt wieder Leerstelle, nach einem weiteren Subline wird ohne Leerstelle die Fortsetzung des unterbrochenen Wortes verschriftet.

Reihenfolge der Markierungen:
Wortklassensymbole bleiben vor dem ersten Wortteil erhalten.
Beide Wortteile ("vorn_" und "_hinten") können jeweils sprecherüberlagert und/oder geräuschüberlagert sein. Die Überlagerungsmarkierungen stehen gegebenenfalls nach (vorderer Teil) bzw. vor (hinterer Teil) dem Subline.

Anmerkungen:

Falls unbedingt erforderlich, könnte im Falle unterbrochener Wörter der letzte Wortteil (die Fortsetzung des unterbrochenen Wortes) als schwerverständlich oder abgebrochen markiert werden.

Ist ein Kompositum zwischen den Kompositumsteilen unterbrochen, steht ein möglicher Bindestrich nach dem zweiten Subline vor dem zweiten Wortteil.
z.B. Wort_ <"ah> _–Unterbrechung

3.2.3.3 Technischer Abbruch

Symbol:

<*T> (technische Turnunterbrechung)
<*T>t (Turnabbruch)

<T_>.. (Wortanfang fehlt)
..<_T> (Wortende fehlt)

Beispiele:

```
Wie<_T>
#vier<_T>
```

```
<T_>wanzigsten
wi<_T>
```

```
<T_>ffen
<T_>"u"s
```

```
<T_>ber
```

```
wir <T_>ssen
<T_>zehnten
```

```
wi<_T> <*T>t
```

Definition:

Wackelkontakte im Aufnahmeequipment oder Bedienungsfehler führen zu technischen Unterbrechungen oder Abbrüchen der Dialogaufnahme.
Im Signalfile ist an diesen Stellen kein Amplitudenausschlag mehr zu sehen.
Aufnahmestörungen treten am Turnanfang, während des Turns und am Turnende auf.
Lexikalische Einheiten können davon direkt betroffen sein, wenn Anfang oder Ende abgeschnitten wurden.

Die Turnunterbrechungs–Markierungen werden nur bei technischen Störungen verwendet. Turnabbrüche können dagegen auch nicht–technische Ursachen haben. Die Turnabbruch–Markierung wird auch verwendet, wenn ein Sprecher seinen Turn selbst abbricht, weil z.B. sein Gegenüber zu sprechen begonnen hat.

Transliterationskonvention:

Turnanfang:
Setzt die Aufnahme erst während eines Dialogbeitrags ein, so wird <T_> ohne Leerstelle vor die lexikalische Einheit oder abgeschnittene lexikalische Einheit gesetzt.

Während des Turns:
Bricht die Aufnahme während einer lexikalischen Einheit ab, dann wird <_T> ohne Leerstelle an das Fragment gehängt.
Fehlen durch die Aufnahmeunterbrechung Teile des Dialogbeitrags, steht anstelle dieser Elemente <*T> zwischen Leerzeichen für Turnunterbrechung.
Setzt die Aufnahme während einer lexikalischen Einheit wieder ein, wird für den fehlenden Wortanfang <T_> ohne Leerstelle vor das Fragment gesetzt.

Turnende:
Ist die Aufnahme am Turnende während einer lexikalischen Einheit abgebrochen, so wird an das Wortfragment <_T> ohne Leerstelle angehängt. Nach einer Leerstelle wird mit <*T>t der Turnabbruch signalisiert.
Wenn keine lexikalische Einheit direkt vom Abbruch betroffen ist, aber eine Äußerungsphrase eindeutig abgebrochen wurde, steht nach dem letzten Wort und einer Leerstelle <*T>t für Turnabbruch.
Bei einem Turnabbruch steht kein Interpunktionszeichen am Ende.

Anmerkungen:

Beim Zusammentreffen von Turnunterbrechung oder technischem Turnabbruch und Interpunktionszeichen, fällt das Interpunktionszeichen immer weg.

4 Liste aller verwendbarer Symbole innerhalb von Wortlisten

;..	globaler Kommentar
..´..	Apostroph (Wort−Reduktion)
..−.. (−−)	Bindestrich (Komposita)
$..	Buchstabierung und Buchstabiersequenzen
~..	Namen
&_	Akronyme
#..	Zahlen
*..	Neologismus
<*XXX>..	fremdsprachiges Wort (FRA,ITA, ..)
..%	schwerverständliches Wort
..=	artikulatorischer Wortabbruch
.._	Wortunterbrechung, linkes Fragment
_..	Wortunterbrechung, rechtes Fragment
<T_>..	technischer Wortabbruch, vorne
..<_T>	technischer Wortabbruch, hinten
<*T>	technische Turn−Unterbrechung
<*T>t	Turn−Abbruch
<%>	unverständliche Sprachproduktion
!KEY..	Codewort

5 Literatur

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

[2] S. Burger: Transliterationslexikon. Verbmobil Technisches Dokument, Nr. 36. München. Oktober 1995

[3] S. Burger: Transliteration spontanprachlicher Daten – Lexikon der Transliterationskonventionen – VERBMOBIL II Verbmobil Technisches Dokument 56.Universität München. April 1997.

[4]A. Batliner: M specified: A revision of the syntactic–prosodic labelling system for large spontaneous speech databases. Verbmobil–Memo 124 F.–A.–Universität Erlangen–Nürnberg. August 1997

[5] A. Batliner, S. Burger, A. Kießling: Außergrammatische Phänomene in der Spontansprache: Gegenstandsbereich, Beschreibung, Merkmalinventar. Verbmobil–Report, Nr. 5. München, Erlangen. Februar 1994.

[6] A. Batliner, A. Kiessling, S. Burger, E. Noeth: Filled Pauses in Spontaneous Speech. Verbmobil–Report, Nr. 88. München, Erlangen. Juli 1995.

[7] Nicole Beringer, Daniela Oppermann, Susanne Burger: Transliteration spontanprachlicher Daten–Lexikon der Transliterationskonventionen–SmartKom (Version1). SmartKom Technisches Dokument Nr. 2, Februar 2000.