Gefördert vom
bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

SmartKom

Mensch-Technik-Interaktion

Transliteration spontansprachlicher Daten – Lexikon der Transliterationskonventionen – SmartKom

Daniela Oppermann
Susanne Burger
Susen Rabold
Nicole Beringer

Ludwig–Maximilians–Universität München

Technisches Dokument Nr. 2, Version6

Dezember 2001

Dezember 2001

Daniela Oppermann

Susanne Burger
Susen Rabold
Nicole Beringer

Ludwig−Maximilans−Universität München
Schellingstr. 3
80799 München

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

E−Mail: {rabold,beringer, daniela, burger}@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

Die für Verbmobil festgelegten Transliterationskonventionen ([1], [2], [3]) wurden für das SmartKom−Projekt in erweiterter und veränderter Form übernommen.

Dies geschah vor allem im Rahmen des Daten−Workshops vom 11.10. − 13.10. 1999 an der Uni München. Die Änderungen zu den Verbmobil−Transliterationskonventionen wurden auf dem Workshop protokollarisch festgehalten. Nachfolgend werden die Änderungen anhand von geeigneten Beispielen erläutert. Die Beschreibung der einzelnen Konventionen sowie die allgemeinen Informationen zu den Transliterationen entstanden in Anlehnung an [3].

Zu den im Transliterationslexikon definierten Konventionen wird es einen Parser und ein Filterprogramm mit diversen Optionen geben.

2 Allgemeines zur Transliteration von Spontansprache in SmartKom

2.1 Spontansprache in SmartKom

Für das Projekt SmartKom werden Versuchspersonen bei der Benutzung einer selbsterklärenden, benutzeradaptiven Mensch−Computer−Schnittstelle aufgenommen. Diese Schnittstelle soll sprachliche und gestische Eingaben interpretieren sowie die Mimik des Nutzers analysieren.

In diesen Mensch−Maschine−Dialogen sollen die Versuchspersonen verschiedene Aufgaben aus verschiedenen Szenarien, etwa die Reiseroute oder das heutige Kinoprogramm, herausfinden.
Das Ergebnis sind einerseits Videodaten,andererseits mehrkanalig aufgezeichnete spontansprachliche Daten, die als Grundlage zu Forschung und Entwicklung im Bereich der Spracherkennung, Sprachsynthese oder des Benutzermodells im SmartKom−Projekt dienen.

2.2 Transliteration von Spontansprache

Diese spontansprachlichen Daten sollen unter Zuhilfenahme der orthographischen Transliteration in einem ersten breiten Verschriftungsschritt allen SmartKom−Partnern symbolisch verfügbar gemacht werden. Das heißt, Projektmitarbeiter hören die Aufnahmen der Dialoge ab und verschriften diese auf Wortebene.

Um spontansprachliche Phänomene wie Satzabbrüche, Korrekturen und Wiederholungen von Äußerungen, Reduktionen oder Häsitationen ebenfalls darzustellen, kommen dabei zusätzlich die Transliterationskonventionen zum Einsatz.
Diese Konventionen regeln auch technische Artefakte, wie technische Aufnahmeabbrüche oder Mikrofongeräusche sowie Sprecherüberlagerungen mit der Sprachausgabe der Maschine in den SmartKom−Dialogen.

2.3 Objekte der Symbolisierung

Folgende Grob−Kategorien bilden die zu verschriftenden Elemente innerhalb eines Dialogbeitrags:

1. lexikalische Einheiten
2. syntaktisch−semantische Gliederung
3. nonverbale artikulatorische Produktionen
4. Geräusche
5. Pausen
6. akustische Überlagerung
7. Kommentare
8. Sonderkommentare
9. Prosodie

2.4 Grundanforderungen

Die Grundanforderungen sind ebenso wie in Verbmobil:

a, automatische Weiterverarbeitung

- einheitliche File–Struktur
- informative und einheitliche Turnnamen
- konsistente Transliterationen
- eindeutige Symbolisierungen
- ASCII–Zeichen
- parsbare Transliterations–Konventionen

b, inhaltliche Anforderungen

- Protokollierung aller hörbaren Dialog–Elemente
- syntaktisch–semantische Marker
- Kennzeichnung von Sprecher– und Geräuschüberlagerungen
- Kennzeichnung bestimmter Wortkategorien (Namen, Zahlen) und Defekte
- möglichst noch lesbare Text–Files

c, Transliterationsvorgang

- einfache Anwendbarkeit der Konventionen
- eingängliche und verständliche Konventionen, auch für Nicht–Experten

2.5 Grenzen der Transliteration

Die hörbaren Ereignisse werden nur orthographisch protokolliert, nicht jedoch genau in ihren lautlichen Besonderheiten beschrieben. D.h. daß die Annotation auf Ebene der breiten Verschriftung
– keine phonologische Verschriftung
– keine phonetische Transkription
– keine zeitliche Zuordnung zu den Signaldaten

liefert.

Für Geräusche und nonverbale Produktionen stehen dem Transliterierer nur Kategorien zur Verfügung.

Besonders auffällige Aussprachen oder Begebenheiten werden durch einen sogenannten Aussprachekommentar oder lokalen Kommentar der Weiterverarbeitung angezeigt, d.h. daß hier etwas Außerordentliches aufgetreten ist.

Aussprachekommentare sind ein Versuch, Aussprachevarianten, Versprecher o.ä. mit Hilfe der Orthographie lautlich zu beschreiben.

2.6 Neues im Vergleich zum alten Transliterations–Lexikon

Die Konventionen aus Verbmobil [1], [2] , [3] decken die Grundanforderungen durchaus ab, sollen aber auf Grund der veränderten Aufnahmesituation vereinfacht werden, um den Transliterationsvorgang zu beschleunigen. Deshalb wurden v.a. die Konventionen zur Geräuschannotation überarbeitet. Außerdem sollen prosodische Marker gesetzt werden in Anlehnung an die M–Grenzenbestimmung [4].
Die neuen Konventionen basieren auf den alten Konventionen und wurden anhand der Wünsche und Vorschläge der Partner, der Erfahrungen aus der Verbmobil–Transliteration

und der Bedingungen, die eine terminierende Grammatik der Transliterations–Konventionen stellt, ergänzt bzw. verändert.

Im Vergleich zu den Transliterations–Konventionen von Verbmobil sind folgende Konventionen neu oder verändert:

1. Symbolisierung technischer Wort– und Turnabbrüche / Turnunterbrechungen
2. neue Konventionen bei prosodischen Markern
3. neue Konventionen für Geräuschüberlagerung.

Im Vergleich zur ersten Version der Transliterations–Konventionen sind folgende Konventionen neu oder verändert:

4. Akronyme
5. Unkonventionelle Komposita
6. Apostrophregelung
7. Header
8. Metatalk

2.7 Benutzung

Das Transliterations–Lexikon basiert auf Erfahrungswerten aus Verbmobil. Es dient beim Lesen und Weiterverarbeiten der Transliterationen zum "Decodieren", soll aber in erster Linie den transliterierenden MitarbeiterInnen als Schulungsunterlage und als Nachschlagewerk zur Verfügung stehen.
Das Lexikon kann abgesehen von manchen Konventionen für deutsche lexikalische Einheiten auch von anderen Sprachen als Standard zur Verschriftung von Spontansprache genutzt werden. Prinzipiell ist die Zielsprache des Lexikons jedoch Deutsch.
Die nur das Deutsche betreffenden Lexikoneinträge sind gesondert mit **deutsch markiert.**

Struktur des Transliterations–Lexikons:

Das Lexikon beschreibt
– die Struktur und das Format einer Transliterationsdatei
– die Konventionen für die Transliteration der einzelnen Turnelemente
– und bietet Übersichtstabellen für alle verwendbaren Symbole.

Die Ordnung der Transliterations–Konventionen richtet sich nach den in die Turnelement–Kategorien eingeteilten Objekten der Symbolisierung.

Lexikoneinträge zu den Turnelementen:
Ein Lexikoneintrag findet sich unter der entsprechenden Element–Kategorie.
Der Eintrag selbst enthält
– Namen des Markers, Symbols oder Ereignisses
– Symbol
– Beispiele
– Definition
– Transliterationskonvention
Die Beispiele stammen entweder direkt aus den SmartKom–Transliterationen bzw. ersatzweise aus den Verbmobil–Transliterationen oder sind entsprechend konstruiert.

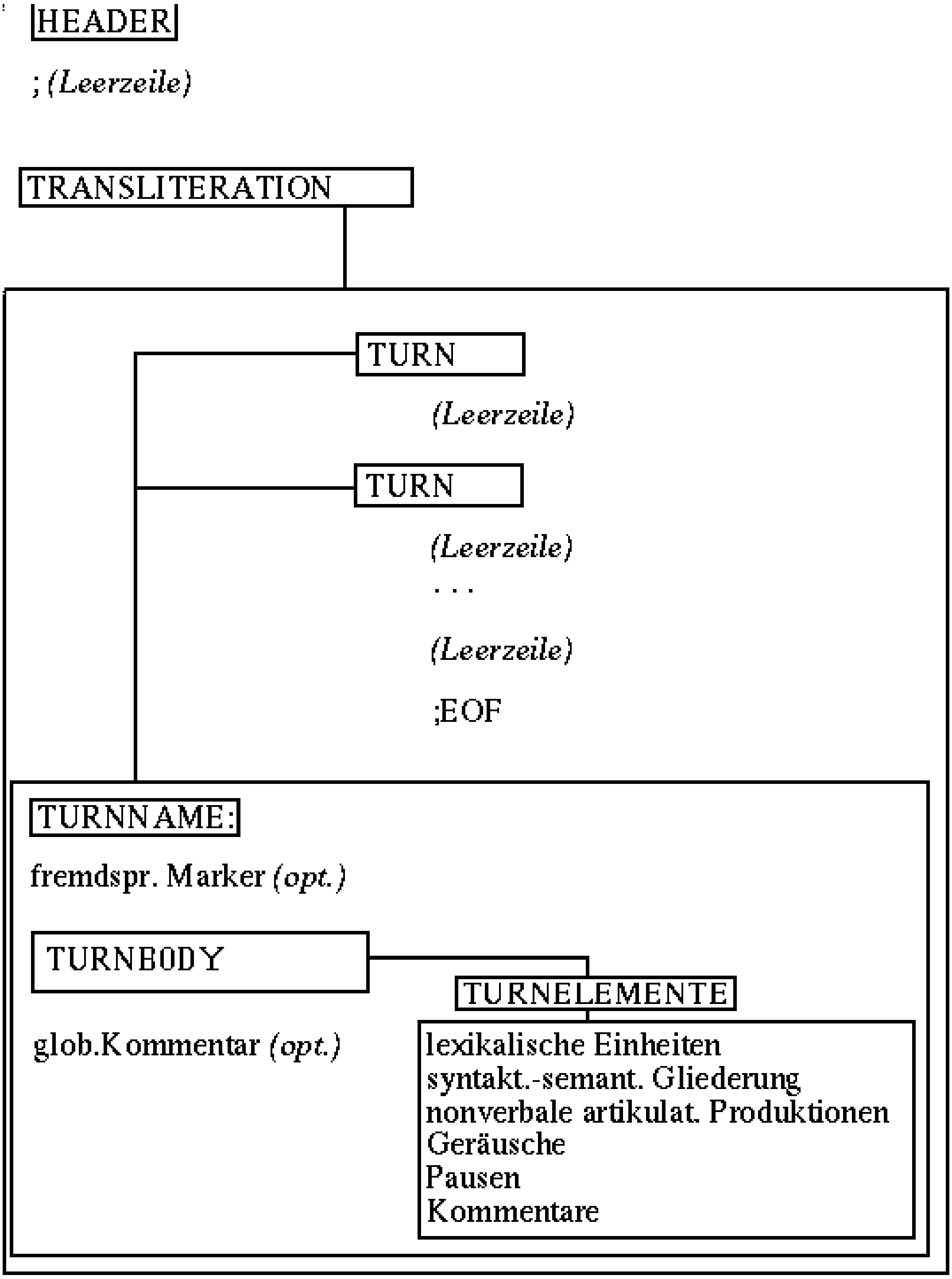
HTML–Format:
Das Lexikon wurde im HTML–Format erstellt. So kann über Links schnell und einfach auf Begriffe, Symbole, deren Bedeutung und Anleitungen zu bestimmten Konventionen zugegriffen werden. Ebenso kann die Liste aller verwendeten Symbole mit der Maus angeklickt werden und so schnell der entsprechende Lexikoneintrag mit Anweisung und Verwendungsbeispiel gefunden werden.
Innerhalb der Lexikoneinträge können sich weitere Links zu entsprechenden Symbolen oder in den Beispielen verwendeten Konventionen befinden.

Alle **Hypertext–Links arbeiten innerhalb des Dokuments**, so daß beim Ausdrucken keine Information verloren geht.

3 Lexikon der Transliterationskonventionen

3.1 Struktur eines Transliterationsfiles

3.1.1 Schematischer Überblick



3.1.2 Globales Fileformat

Ein Transliterations–File besteht aus dem
Header und der **Transliteration der Dialogbeiträge.**
Header und Transliteration sind durch eine **mit Semikolon beginnende Leerzeile** voneinander getrennt.

Die Transliteration ist in **Turns** unterteilt.
Zwischen den einzelnen Dialog–Turns steht **eine Leerzeile** als **Turntrenner.**
Das Ende der Transliteration wird mit einer **Leerzeile nach dem letzten Turn** und einer **mit ;EOF beginnenden Endzeile** angezeigt.

Die Turns bestehen aus
Turnname,
gegebenenfalls einer **Kennung für die Sprache**, in der der Dialogbeitrag gesprochen wurde, **oder dem Kennzeichen für Off–Talk,**
dem **Turnbody** mit den **Turnelementen**
und optional einem **globalen Kommentar** zum Turn.

3.1.2.1 Header

Beispiel:

```
; DVD: 1.0
; Version: 1.0
; Dialog: w001_pk
; ATMO: Demovideo
; ENC: TEX
; zuletzt bearbeitet am: 17.11.99
; VPK: AAA
; Offtalk: kein | wenig | viel
; bearbeitet von: SB
; Tonqualit"at: leise
```

Jede Headerzeile beginnt mit **Semikolon, gefolgt von Leerzeichen**.

In der ersten Headerzeile steht nach "DVD:" und Leerzeichen die DVD Version

```
; DVD:: 12.00
```

in der zweiten Headerzeile steht nach "Version:" und Leerzeichen die aktuellste Version der Transliteration.

```
; Version: 1.0
```

in der dritten Headerzeile steht nach "Dialog:“ und Leerzeichen der Dialogname (=auch Name des entsprechenden Directories, unter dem auf der DVD die Signalfiles des Dialogs zu finden sind.)

```
; Dialog: w001_pk
```

die vierte Zeile gibt das verwendete Encoding an: TEX oder ISO

```
; ENC: TEX
```

in der sechsten Headerzeile wird die Versuchspersonkennung angegeben.

```
; VPK: AAA
```

die siebte Headerzeile gibt die eingespielte Atmosphäre an. Z.B.:

```
; ATMO (Büro, Wohnung, Telefonzelle): Auto
```

die achte Headerzeile gibt an, ob die Transliteration keinen, wenig oder viel Offtalk enthält.

```
; Offtalk: kein | wenig | viel
```

In möglichen weiteren Headerzeilen stehen Kommentare, die sich auf den gesamten Dialog beziehen.

```
; zuletzt bearbeitet am 17.11.99
; bearbeitet von: SB

; Tonqualit"at: ganzer Dialog sehr leise
; .......
```

Die Tonqualität wird subjektiv bestimmt und kann in drei Kategorien (laut , normal, leise) eingeordnet werden. Ist keine Angabe für die Tonqualität angegeben, wird der Defaultwert »normal« eingesetzt.Header und Transliteration sind durch eine **mit Semikolon beginnende Leerzeile** von einander getrennt.

3.1.2.2 Transliteration

Beispiel:

```
w001_pk1_000_SMA: hallo , ich bin !KEYSmartKom . ich
kann Ihnen Auskunft "uber das Wetter und das
Kinoprogramm geben . ich reagiere auf Ihre m"undliche
Anweisung oder auf Eingaben mit dem <*EN>touch<Z>
<*EN>screen . um mich zu aktivieren , verwenden Sie
bitte das Kommandowort !KEYComputer oder !KEYSmartKom
. <#Klopfen>


w001_pkd_001_AAA: !KEYComputer , Wetter .


w001_pkw_002_SMA: f"ur welche Region w"unschen Sie
eine Wettervorhersage ?


w001_pkd_003_AAA: <"ahm> f"ur ~M"unchen . <P>
~M"unchen ? <P> Hilfe .


w001_pkw_004_SMA ...

;EOF
```

Der Transliterationsteil beinhaltet die Niederschrift des gesamten Dialoges.
Jeder Sprecherbeitrag wird dabei in einem gekennzeichneten Turn festgehalten. Da der Smartkom-Assistent aktiv am Dialog beteiligt ist, werden die entsprechenden Turns daher gemäß der Konventionen verschriftet. Zu beachten ist, daß der Assistent in jedem Dialog dasselbe Sprecherkürzel erhält (SMA).

Zwischen den Turns steht eine Leerzeile.

Die Transliteration endet mit einer Leerzeile nach dem letzten Turn und ;EOF als Ende-Markierung.

3.1.2.2.1 Turns

Ein Turn oder Sprecherbeitrag beginnt mit dem Turnnamen, gefolgt von einer Leerstelle. Danach kann eine **Kennung der Hauptsprache** des folgenden Turnbodies stehen (bei mehrsprachigen Dialogen) oder eine **Kennung von Off-Talk** (bei Spracheingaben, die nicht für das SmartKom-system relevant sind), wieder gefolgt von einer Leerstelle.
Im **Turnbody** werden alle hörbaren Ereignisse, syntaktisch-semantische Markierungen und Kommentare in der Regel mit einer Leerstelle voneinander getrennt protokolliert.

Am **Zeilenende innerhalb eines Turns** steht **Zeilenumbruch**, die **neue Zeile** beginnt mit einer **Leerstelle**.

Nach dem **letzten Turnelement im Turn** steht keine **Leerstelle, nur Zeilenumbruch.**

Dem Turn kann ein **globaler Kommentar** folgen. Dieser beginnt in der nächsten Zeile nach dem vorausgegangenen Turn und endet mit Zeilenumbruch.

3.1.2.2.2 Turnname

Beispiel:

```
w001_pkd_001_AAA: !KEYComputer , Wetter .
```

Definition:

Jeder Turn beginnt mit dem Turnnamen. Dieser Name dient als Identifikator, wenn nach einzelnen transliterierten Turns gesucht werden soll.
Im Turnnamen wird der Bezug zum Signalfile und zur Sprecherdatenbank geschaffen.

Transliteration:

Dialogname	Mikro-Art	Turnnummer	Sprecherkürzel
w001_pk	d (= Richtmikrofon)	_001	_AAA

Signalfilename:
Der Signalfilename ist der Name des Signalfiles auf der DVD ohne Extension.
Signalfilenamen bestehen aus:
Dialog-Directory-Namen sowie **Art des verwendeten Mikrofons** bzw. des Wizard-Kanals wird mit Kleinbuchstaben oder Nummer angegeben und **Sprecherkürzel** abgetrennt durch Unterstrich.

Bsp.: w005_pkd_AAA.wav

D-D-N	M	Spk

- Beim **Dialog-Directory-Namen** wird die **Art der Aufnahmen** als einstelliges Kürzel zu Beginn des Dateinamens wiedergegeben. Es wird unterschieden zwischen Biometrieaufnahmen (**b**) und Wizard-Aufnahmen (**w**). Im Anschluß wird eine **3stellige**

Sessionnummer vergeben. Es folgt ein zweistelliges Kürzel zur **Identifikation des Szenarios** abgetrennt durch Unterstrich.

Bsp.: w005_pk

- **Art des verwendeten Mikros**

  Bsp.: d = Raummikrofon

  w = Wizardmikrofon

- **Turnnummer**

  Mit einem Unterstrich getrennt, folgt der Mikrofonkennung die dreistellige Turnnummer. Diese wird über den ganzen Dialog hochgezählt.

- **Sprecherkürzel:**
  Jeder Sprecher wird mit einem individuellen Sprecherkürzel markiert. Dieses Sprecherkürzel besteht aus drei Großbuchstaben (keine Umlaute), die den Sprecher in der Sprecherdatenbank eindeutig identifizieren. Das Sprecherkürzel wird mit Unterstrich abgetrennt.
  Bsp.: AAA

Der Turnname endet mit **Doppelpunkt** und **Leerstelle** (oder Zeilenumbruch−Leerstelle).

3.1.2.2.3 Sprachenkennzeichnung bei mehrsprachigen Dialogen

Symbol: <*tXX>

wobei XX:

EN = englisch
DE = deutsch (bei nicht−deutschen Dialogen oder mehrsprachigen Dialogen)

auch
FR =französisch
IT = italienisch
ES = spanisch
JA = japanisch
etc.

Die Sprachkennung folgt den Normen der ISO−Language−Codes.

(http://www.ics.uci.edu/pub/ietf/http/related/iso639.txt)

Beispiel:

```
w123_ptc_000_AAP: <*tEN> good
 morning , ~John . how are you ?
w123_ptc_000_AAP: <*tDE>
 guten Tag , Herr ~Miller . danke , es geht mir gut .
```

Definition:

Bei mehrsprachigen Dialogen wird nach jedem Turnnamen eine Kennzeichnung für die Sprache, in der der folgende Dialogbeitrag geführt wurde, gesetzt.

Transliteration:

Nach der Leerstelle des Turnnamens steht <*tXX> und eine weitere Leerstelle (oder Zeilenumbruch–Leerstelle).

Anmerkung:

Für den Dialogbeitrag gelten die orthographischen Regeln der jeweiligen Sprache.

3.1.2.2.4 Kennung von Off–Talk

Symbol: <*tOOT> und <*tROT>

Ebenso wie bei der Markierung von Fremdsprachen, kann ein Turn mit einem Off–Tal–Marker versehen sein. <*tOOT> wird eingesetzt, wenn ein ganzer Turn im Other–Off–Talkgesprochen wird, <*tROT>wenn Text vom Bildschirm abgelesen wird, also bei Read–Off–Talk.

Beispiel:

w002_pkd_010_AAA: <t*ROT> <Ger"ausch> Kino ~Schlo"s , ~Hauptstra"se , und Kino<Z> <P> ~Hoelldobler .

Transliteration:

Nach der Leerstelle des Turnnamens steht <*tROT> oder <*tOOT> und eine weitere Leerstelle (oder Zeilenumbruch–Leerstelle).

3.1.2.2.5 Turnbody

Im Turnbody werden nach dem Turnnamen alle hörbaren Ereignisse, wie lexikalische Einheiten und Geräusche, – die **Turnelemente** – und zusätzliche Marker, wie syntaktische Markierungen oder Kommentare, transliteriert.

Format des Turnbodies:

a, ASCII–Kodierung
Die Kodierung der Transliteration erfolgt in **7–bit ASCII**. Für die Umlaute und "ß" wird im Verschriftungsformat die **TeX–Schreibweise** verwendet, also "a, "U, "s etc.

Die TeX–Schreibweise wird im Ablieferungsformat nach ISO konvertiert.

b, Element–Trennung
Zwischen den Turnelementen steht ein **Leerzeichen** oder **am Zeilenende** ein **Zeilenumbruch** gefolgt von einem **Leerzeichen.**

c, Worttrennung
Worttrennungen werden **nicht** durchgeführt; es wird immer das ganze Wort in die nächste Zeile geschrieben.

Turnende

Enthält ein Turn eine lexikalische Einheit, muß er durch die Satzzeichen "Punkt" oder "Fragezeichen" oder durch eine Turnabbruchmarkierung (<*T>t) beendet werden. Nach dem letzten Satzzeichen dürfen noch technische Geräusche oder nonverbale artikulatorische Produktionen verschriftet werden. Die Turnabbruchmarkierung muß dagegen die letzte Einheit des Turns bilden.
Enthält ein Turn keine lexikalische Einheit (vgl. minimaler Turn), so wird kein Satzzeichen am Ende gesetzt. Eine Turnabbruchmarkierung kann jedoch – je nach Höreindruck – gesetzt werden.

Minimaler Turn

Ein Turn muß mindestens enthalten:

- Entweder eine lexikalische Einheit
- oder eine Häsitation
- oder das Symbol für eine unverständliche Äußerung (<%>).

3.1.2.2.6 Gobale Kommentare

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

Beispiele:

```
w123_ptc_001_BBP: gr"u"s Gott , mein Name ist
~G<Z>"urtner , <A> <"ahm> $G $"U $R $T $N $E $R
 . <A> <"ahm>
;Brummen "uber gesamtem Turn
```

```
w123_ptc_001_BBP: hallo , Herr
 ~Meier , wir m"ussen einen Termin ausmachen.
;Sprecher ist heiser.
```

Definition:

Dem Turn kann optional ein globaler Kommentar folgen.
Im globalen Kommentar werden Ereignisse oder Besonderheiten, die während des gesamten Turns stattgefunden haben, protokolliert, wie Geräusche, die durchgehend zu hören waren, artikulatorische Auffälligkeiten des Sprechers oder sonstige Vorkommnisse, die der Transliterierer für wichtig hält und vermerken möchte.

Transliteration:

Der globale Kommentar steht nach dem letzten Turnelement in einer **neuen Zeile**, die **mit Semikolon** beginnt.
Jede weitere Zeile des globalen Kommentars beginnt ebenfalls mit Semikolon.
Nach dem Kommentar folgt dann die Turntrennung mit einer Leerzeile.

3.2 Transliteration der Turnelemente

3.2.1 Lexikalische Einheiten

Definition der Elementklasse:

Lexikalische Einheiten sind

– Wörter (entsprechend einem Dictionary der jeweiligen Sprache, z.B. deutsch: Duden)

– Interjektionen

– reguläre reduzierte Wortformen

– Komposita

– klassifizierte Wörter (Wörter, die bei der Weiterverarbeitung der Transliterationen gesondert behandelt werden, wie Namen, Zahlen)

– Wörter mit artikulatorischen Eigenheiten (Reduktion, Abbruch, starke Dehnung)

Verschriftungsregeln:

Leerstelle:
Vor und nach der lexikalischen Einheit steht eine Leerstelle oder Zeilenumbruch + Leerstelle.

Ausnahme:
– Markierung für agrammatische Phrase

– Markierung von akustischer Überlagerung

– Markierung von Unverständlichkeit

– Markierung von Off–Talk

Orthographie:
Entsprechend der Sprache, in der der Dialogbeitrag geführt wurde, gelten bei der Transliteration die jeweiligen für diese Sprache gültigen orthographischen Regeln. Das gilt auch für die in der Transliteration gesondert markierten Wörter, soweit sie noch als der entsprechenden Sprache zugehörig erkennbar sind.

Zusätzlich gilt für die SmartKom–Transliterationen eine Wortliste, in der bei verschiedenen möglichen Schreibweisen eine einheitliche festgelegt wird.

Unabhängig von der Orthographie kann nach gezögerten Lauten <Z> für Zögerung an die entsprechende Position des Wortes gesetzt werden.

deutsch

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.
ABER: Am Äußerungsanfang werden nur Substantive großgeschrieben. In allen anderen Fällen wird die Dudenkonvention zur Großschreibung am Anfang eines Satzes nicht beachtet.
Sowohl die nominalen Formen der Anredepronomina "Sie" und "Ihnen", als auch die possessive Form "Ihr" (mit allen Flexionen) werden groß geschrieben, um sie von den Pronomina für die dritte Person zu trennen. Da eine solche Trennung bei "du" nicht erforderlich ist, wird es immer klein geschrieben.

deutsch

Sondermarkierungen:
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 markiert werden.

Ausnahme: Fremdsprachliche Markierung. Diese darf zusätzlich nach der Markierung, vor der lexikalischen Einheit stehen.
Wenn artikulatorische Besonderheiten symbolisiert werden (Abbrüche, Zögerung), können diese Markierungen auch an klassifizierte lexikalische Einheiten angehängt werden, solange noch erkennbar ist, um welche lexikalische Einheit es sich handelt.

Elementumgebung:

Lexikalische Einheiten können von Geräuschen oder von einem anderen Sprecher überlagert sein.

Lexikalischen Einheiten kann ein Aussprachekommentar folgen.

Agrammatische Phänomene treten nur in Verbindung mit lexikalischen Einheiten auf.

Die Interpunktion kann nur nach lexikalischen Einheiten (oder kommentierten lexikalischen Einheiten) stehen.

Turnende:
Nach einer lexikalischen Einheit (oder lexikalischen Einheit mit Kommentar) am Turnende **muß** stehen:
Punkt oder
Fragezeichen oder
Turnabbruch (bei technischem Abbruch des Turns oder wenn der Sprecher seinen Dialogbeitrag selbst abbricht)

3.2.1.1 Wörter eines Dictionaries

Symbol:

(nicht markiert)

Beispiele:

```
w001_pk1_008_SMA: verwenden Sie bitte f"ur eine
Eingabe das Kommandowort !KEYSmartKom oder
!KEYComputer .
```

```
w001_pk1_009_AAA: !KEYComputer , ~M"unchen .
```

```
w001_pk1_010_SMA: heute herrscht unbest"andiges
Herbstwetter . Es dominieren dichte Wolkenfelder, die
zeitweise Regenschauer bringen . die Temperaturen
liegen zwischen #dreizehn und #siebzehn Grad .
w001_pk1_011_AAA: !KEYComputer , Temperatur noch mal .
w001_pk1_012_SMA: die Temperaturen liegen zwischen
#dreizehn und #siebzehn Grad .w001_pk1_013_AAA: gut ,
danke sch"on . <"ahm> <P> !KEYComputer , Wetter .
```

<u>Definition:</u>

Wörter eines Dictonaries sind lexikalische Einheiten,
– die in der transliterierten Form in die SmartKom−Wortliste eingehen und in eine andere Sprache übersetzt werden können,
– die gut verständlich und nicht verstümmelt, nicht neu erfunden oder nicht fremdsprachlich sind.

<u>Transliterationskonvention:</u>

Entsprechend der Sprache, in der der Dialogbeitrag geführt wurde, gelten bei der Transliteration die jeweiligen für diese Sprache gültigen orthographischen Regeln.
Für Wörter eines Dictionaries gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Anmerkungen:

deutsch

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.

deutsch

3.2.1.2 Interjektionen

<u>Symbol:</u>

(nicht markiert)

Beispiele:

```
... oh, das pa"st mir gar nicht . mm . ja nun , ne ,
 dann brauchen wir einen anderen Termin . ....
```

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".

3.2.1.3 Wortreduktionen

deutsch

Symbol:

´ (teilweise mit Apostroph)

Beispiele:

```
... okay , nehmen <!1 nehm'> wir <!1 ma> mal an , das
das <!1 des> morgen is' ...
```

Definition:

Wortreduktionen, die in die Transliteration mit eingehen, sind:
- End−e Reduzierung
- End−t Reduzierung (nur bei "ist" −> "is'" und "nicht" −> "nich'")
- Reduktion unbestimmter Artikel
- Verschmelzung von Präposition und reduziertem bestimmten Artikel
- reduziertes Pronomen der 3. Pers. Sing. (es)

Transliterationskonvention:

a, 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.
(ich) hab´
m"u"st´
wollt´
heut’
(aber: wir haben <!1 hab´> da keine Zeit)

End−t Reduzierungen:
is’
nich’
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 im Text 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>.

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

b, Sonstige Reduktionen

Die Verschmelzung von Präposition mit bestimmtem Artikel wird durch einen Aussprachekommentar angezeigt, sofern die Präposition ebenfalls reduziert ist. (Siehe auch 3.2.1.7.)

mit dem <!2 mi´m>

aber:
nach dem <!1 ´m>..

Für Wortreduktionen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

deutsch

3.2.1.4 Komposita

deutsch

Symbol:

– (teilweise mit Bindestrich)

oder auch

–– (Doppelbindestrich bei Kompositumsergänzung)

Beispiele:

```
... wenn wir die Acht-Uhr-Maschine noch erreichen ...
...
```

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, wobei die Groß– und Kleinschreibung der einzelnen Bestandteile erhalten bleibt.
Acht–Uhr–fünf–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 nach dem ersten Wortteil.

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

3.2.1.5 Klassifizierte lexikalische Einheiten

3.2.1.5.1 Buchstabierung

Symbol:

$ (bei ausgesprochenen Buchstaben)

Beispiele:

```
... mein Name ist ~J<Z>"ansch , $J , $"A , $N , $S ,
$C , $H ...
... in die $U-$S-$A fahren ...
... die $A-$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. Die Buchstabiereinheiten sind hier durch Kommata getrennt und werden jeweils wie lexikalische Einheiten behandelt.

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

3.2.1.5.2 Akronyme

deutsch

Symbol:

&

Beispiele:

```
... <Schmatzen> <A> <"ahm> <P> &$I-$B-$M . ...
... wenn wir die &Soko benachrichtigen ...
```

Definition:

Offizielle Buchstabenfolgen, die stellvertretend für Wörter stehen können.

Als Wörter ausgesprochene Abkürzungen werden nach üblicher Schreibweise festgehalten und wie lexikalische Einheiten behandelt.
&DIN
&Benelux
&OPEC
&AStA
&CeBIT
&$A–$R–$D
&Soko
&<*EN>$B–<*EN>$B–<*EN>$C

Transliteration:

Akronyme werden mit vorgestelltem & amrkiert. 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"

3.2.1.5.3 Namen

Symbol:

~

Beispiele:

```
... auf dem Display sehen Sie die heutigen
Vorstellungszeiten f"ur ~<*EN>Eyes+Wide+Shut im ~Arri
. f"ur welche Uhrzeit soll ich Karten bestellen ? ...
```

Definition:

Markiert werden:

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

Nicht markiert werden:

- Feiertagsnamen, sofern sie keinen eigenen Namen haben oder Teil eines Kompositums sind (also ~Heilig+Drei+K"onig, aber Maifeiertag oder Weihnachtswoche)
- Namen von Staaten , z. B. Frankreich, England (aber: Bundesstaaten oder Bundesländer)
- Kontinente, z. B. Amerika, Afrika
- Namen von Weltmeeren, z. B. Atlantik, Pazifik

Transliterationskonvention:

Namen werden mit vorgestelltem ~ markiert.
Ist ein Name nur Teil eines ansonsten aus anderen Wörtern bestehenden Kompositums, wird der Name nicht markiert (z.B. Berlin–Video).
Für Namen gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Bei Zusammengesetzte Namen, die in normaler Orthographie durch Leerzeichen getrennt sind, werden die Namensbestandteile mit Pluszeichen aneinandergehängt (z.B. ~<*EN>Eyes+Wide+Shut).

Beispiel:
~Zur+blauen+Traube
~Bu"s+–und+Bettag
~Heilig+Drei+K"onig
Diese werden mit Pluszeichen zusammengeschrieben und als Namen markiert. Die Groß/Kleinschreibung der Namensbestandteile wird auch in diesem Fall beibehalten. Angaben zur Sprache werden ebenfalls markiert.

Zusammengesetzte Namen sind Kinonamen, Hotel–, Restaurantnamen, Filmtitel.

3.2.1.5.4 Zahlen

Symbol:

#

Beispiele:

```
... st"urmisch, dabei aufgelockert bew"olkt mit
sonnigen Phasen . Temperaturen liegen zwischen #zehn
und #sechzehn Grad ...
```

... Temperaturen liegen tags"uber um die #neunzehn Grad ...

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 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.

3.2.1.5.5 Neologismen

Symbol:

*

Beispiele:

... *haarknapp <:<#Rascheln> um:> einen <!1 ein’> Tag verfehlt ...
... ich *verschraubel das jetzt mal
... was *exkursieren Sie denn ? ...
... *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:

Kleinere Versprechern sollten in der richtigen Version mit Aussprachekommentar, völlig unsinnige Lautkombinationen als unverständliche Sprachproduktion verschriftet werden.

3.2.1.5.6 Fremdsprachige Wörter

Symbol:

<*XX>

wobei XX:
EN = englisch
DE = deutsch (bei nicht−deutschen Dialogen oder mehrsprachigen Dialogen)

auch
FR =französisch
IT = italienisch
ES = spanisch
JA = japanisch
etc.

Beispiele:

```
... ich reagiere auf Ihre m"undliche Anweisung oder
auf Eingaben mit dem <*EN>touch<Z> <*EN>screen...
... <*JA>sayonara , Herr ~Fujisaki . ...
... <*FR>bien . <*FR>c´est <*FR>la <*FR>vie w"urde ich
sagen ...
... tschau , <*IT>bella ...
```

Definition:

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

Transliterationskonvention:

Fremdsprachige Wörter werden mit vorangestelltem <*EN> (<*JA>, <*IT>, <*FR>,<*DE>, ...) markiert. Die Abkürzungen für die Sprache richtet sich nach dem ISO−Language−Code 639.

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

Anmerkungen:

ABER: Manche fremdsprachigen Wörter stehen 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.
<*FR>d´accord

deutsch

3.2.1.5.7 Kommando–Wörter

Symbol:

!KEY

Beispiele:

```
.. um mich zu aktivieren , verwenden Sie bitte das
Kommandowort !KEYComputer oder !KEYSmartKom ...
```

Definition:

Kommandowörter sind Wörter, mit denen das System metasprachlich bedient werden kann.

Transliterationskonvention:

Kommandowörter werden mit einem vorangestellten !KEY ohne Leerstellte verschriftet und ansosnten wie normale lexikalische Einheiten behandelt.

3.2.1.6 Lexikalische Einheiten mit artikulatorischen Besonderheiten oder Defekten

3.2.1.6.1 Zögerung

Symbol:

<Z>

Beispiele:

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

Bei vokalisiertem R: zur<Z>
```

Definition:

Werden Laute innerhalb einer lexikalischen Einheit verhältnismäßig lang gedehnt, z.B. gedehnte Laute vor Phrasengrenzen ("Prefinal Lengthening") oder im Sinne einer Häsitation, wird dies als Zögerung eines Lautes protokolliert.
Das gilt auch für Plosive mit überlanger Verschlußphase, starker oder andauernder Aspiration.

Transliterationskonvention:

<Z> wird ohne Leerstelle direkt an den gedehnten Laut angehängt. Bei Zögerung innerhalb eines Wortes erfolgt die Verschriftung des Wortendes direkt im Anschluß an <Z>.

Für Wörter mit gezögerten Lauten gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

3.2.1.6.2 Schwerverständliche Wörter

Symbol:

%

Beispiele:

```
... #eins% , #zwei% , #drei% , #vier% ...
... Sie k"onnen den Ort auch% auf dem Display eingeben
...
```

Definition:

Schwerverständlich sind alle Wörter, die nicht eindeutig verstanden werden können, weil sie durch Geräusche oder Artikulation akustisch schlecht gehört werden, oder bei denen nur schwer auf die hochsprachige Version rückgeschlossen werden kann, (bedingt durch dialektale oder artikulatorische Varianten).

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.

Anmerkungen:

Schwerverständliche Wörter können nicht am Wortende abgebrochen sein.
Werden Wörter wegen dialektaler oder aussprachebedingter 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.

3.2.1.6.3 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.

3.2.1.6.3.1 Artikulatorischer Abbruch

Symbol:

=

Beispiele:

```
... wie gebe ich +/das Wetter/+ +/da=/+ den Ort auf
dem Display ein...
... -/im Ja=/- also , ich sag' Ihnen jetzt ...
```

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

3.2.1.6.3.2 Artikulatorische Unterbrechung lexikalischer Einheiten

Symbol:

_ (Unterstrich)

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 _ (Unterstrich) 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 Unterstrich 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 Unterstrich.

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 oder ein Plus bei zusammengesetzten Namen **nach** dem ersten Wortteil, aber **vor** dem ersten Unterstrich.
z.B. Wort−_ <"ah> _Unterbrechung

3.2.1.6.3.3 Technischer Abbruch

Symbol:

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

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

Beispiele:

```
... auf Wie<_T> <*T>t
... ich h"atte am #vier<_T> <*T> <T_>wanzigsten Zeit
...
... danke , wi<_T> <*T> <T_>ffen uns ...
```

```
... k"onnten wir uns <*T> oder #acht-zehnten M"arz
treffen ? ...
... dann bis <*T>t
... <T_>"u"s Gott <*T> <T_>ber , wir <T_>ssen noch
einen Termin ausmachen ...
... danke , wi<_T> <*T>t
```

Definition:

Wackelkontakte im Aufnahmeequippment oder Bedienungsfehler führen zu technischen Unterbrechungen oder Abbrüchen der Dialogaufnahme.
Im Signalfile ist an diesen Stellen kein Amplitudenausschlag 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 Turnabbruch und Interpunktionszeichen, fällt das Interpunktionszeichen immer weg.

3.2.1.7 Aussprachekommentare

Symbol:

<!n ..>

wobei n = Anzahl der betroffenen lexikalischen Einheiten.

Beispiele:

```
... Donnerstag <!1 Donnaschag> ...
... nat"urlich <!1 ´t"urlich> ...
... irgendwie <!1 irgen´wie> ...
... und dann <!2 un´a´> ...
... #siebenundzwanzig <!1 sienzanzesch> ...
... k"onnen wir <!2 k"omma> ...
... kannst du <!2 kannste> ....
... das <!1 des> is’ aber das Wetter von heute ...
... dann kommen <!1 komm´> Sie <!1 Se> doch ...
... wenn wir <!2 wemma> ’s die Woche noch machen ...
```

Definition:

Dialektaussprachen, andere Stilformen, Versprecher oder sonstige Abweichungen von der üblichen Aussprache werden in korrekter Form nach Duden verschriftet. Im Aussprachekommentar wird versucht, mittels Orthographie und Apostrophregelung die Abweichungen so zu protokollieren, daß schon aus der Transliteration erste Informationen über Aussprachevarianten gezogen werden können und für tiefere Analysen die interessanten Stellen markiert sind.

Transliterationskonvention:

Nach einer Leerstelle (oder Zeilenumbruch−Leerstelle) folgt der betroffenen / den betroffenen lexikalischen Einheit/en der Aussprachekommentar.
Der Aussprachekommentar steht nach <! , einer Zahl, die die Anzahl der betroffenen lexikalischen Einheiten bezeichnet und einer Leerstelle. Der Aussprachekommentar schließt mit >.

Innerhalb der Kommentarklammern gelten in Zweifelsfällen die orthographischen Regeln. (Groß−und Kleinschreibung wird beibehalten, Komposita−Bindestriche und Namens−Plus, ebenso doppel−s− oder ie−Schreibung vorausgesetzt, die Aussprache legt nichts anderes nahe.)

Ansonsten kann die Orthographie zur Verdeutlichung bestimmter Varianten benutzt werden (besonders langes /i/ statt kurzem /i/ wird dann z.B. mit "ie" transliteriert).

Positionen von Lautelisionen werden mit **einem** Apostroph gekennzeichnet.

Entfällt bei einem Wort der Endlaut/Endlaute , beim folgenden der Anfangslaut /Anfangslaute, so steht auch hier für die Elisionsposition **nur ein** Apostroph für die Elisionsstelle und keine Leerstelle zwischen den Wörtern.

Werden in der Aussprachevariante Laute durch andere Laute ersetzt oder Laute hinzugefügt, dann wird versucht, mittels geeigneter Buchstaben die veränderten Laute zu beschreiben. Fehlen in diesen Wörtern zusätzlich Laute, steht in diesem Fall **kein** Apostroph. Das veränderte Wort wird so transliteriert, als ob es sich um ein neues Wort handeln würde.

Zwischen enklitisierten Varianten steht kein Apostroph. Diese Wörter werden zusammengeschrieben.

Anmerkungen:

Die Zusammenfassung mehrerer variierter lexikalischer Einheiten in einen Aussprachekommentar sollte eher vermieden werden. Nur bei stark zusammengezogenen oder enklitisierten Einheiten bezieht sich der Kommentar notwendigerweise auf mehrere Einheiten.

deutsch

Aussprachevarianten wie /könig/ vs /könich/ brauchen nicht kommentiert werden, da beide Versionen im Deutschen üblich sind.

deutsch

3.2.2 Syntaktisch–semantische Gliederung

Definition der Elementklasse:

Syntaktisch–semantische Gliederungen sind nicht–zeitintensive Markierungen zur Strukturierung des Satzflusses.
Soweit in der Spontansprache möglich wird versucht, reguläre Satz– und Nebensatzstrukturen mittels eines Subsets an Interpunktionszeichen zu markieren. Grammatisch irreguläre Phänomene wie Korrekturen oder Satzabbrüche werden so markiert, daß sie mit Hilfe eines geeigneten Filters so aus dem Text genommen werden können, daß sinnvolle syntaktisch–semantische Strukturen entstehen.

3.2.2.1 Interpunktion

Symbol:

.
?
,

Definition:

Punkt, Fragezeichen und Komma stehen als Interpunktions–Subset zur syntaktischen Markierung regulärer Satzteile zur Verfügung. Generell ist aber die Zeichensetzung bei Spontansprache mit Schwierigkeiten verbunden, da grammatikalisch "richtige" Sätze häufig fehlen.

Transliterationskonvention:

Die Interpunktionszeichen stehen wie alle Turnelemente zwischen Leerzeichen (oder Zeilenumbruch–Leerzeichen).
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. Die Zeichensetzung erfolgt nach den Regeln der für die verwendete Sprache gültigen Grammatik, soweit möglich.

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

- Grammatik
- Intonation

- Pause, Atmen
- Beginn eines neuen Gedankens

Beispiel:

```
... gut , danke sch"on . <"ahm> <P> !KEYComputer ,
Wetter ...
```

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

- Fragewörter
- Satzbau
- Intonation

Kontext Beispiel:

```
... ach so . <A> <"ahm> <P> und wie ? !KEYComputer ,
wie gebe ich +/das Wetter/+ +/da=/+ den Ort auf dem
Display ein ? ...
```

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

- Grammatik
- einen Nebensatz einleitende Partikel
- Intonation
- zusammengehörende Gedanken
- Phrasenmarkierungen wie Atmen oder Pause

z.B. :
a, ich könnte schon , mittwochs hab´ ich immer Zeit .
b, ich könnte schon . <A> <P> mittwochs hab´ ich immer Zeit .

Beispiel:

```
... wolkenreich , aber wenig schaueranf"allig .
starker nordwestlicher Wind . ...
```

Anmerkungen:
Vor den Interpunktionszeichen finden sich nur lexikalische Einheiten, lokale Kommentare, Ausprachekommentare oder prosodische Marker. Alle anderen Ereignisse, auch Atmen, Häsitationen oder Unverständliches, werden immer nach den Interpunktionszeichen verschriftet.
Komma kann nicht am Turnende stehen.
Nach Abbruchsmarkierungen steht keine Interpunktion (Wortabbruch genauso wie Satzabbruch oder technische Unterbrechung).

3.2.2.2 Agrammatische Phänomene

Definition:

Agrammatische Phänomene treten dann auf, wenn ein Sprecher sich innerhalb einer Äußerungsphrase unterbricht. Nach der Abbruchstelle werden Teile der Äußerungsphrasen wiederholt oder korrigiert (Wiederholung/Korrektur) oder es erfolgt keine Wiederaufnahme und der Sprecher beginnt einen neuen Gedankengang (False Start). Markiert werden der Anfang der agrammatischen Phrase und die Abbruchstelle.

Siehe auch [5].

Transliteration:

Anfangs– und Endklammer der agrammatischen Phrase werden immer ohne Leerstelle vor (Anfang) oder nach (Ende) der ersten bzw. letzten lexikalischen Einheit verschriftet. Klassifizierungssymbole und Abbruchmarkierungen gehören zur lexikalischen Einheit.
+/~Huber/+
+/da=/+
Ist eine der zu markierenden lexikalischen Einheiten geräusch– oder sprecherüberlagert, dann steht die Phrasenmarkierung vor bzw. nach der Überlagerungsmarkierung.
+/@1Kartoffel/+
+/Kartoffel1@/+
+/<:<#> Kartoffel:>/+
Kommentare zu den betroffenen lexikalischen Einheiten stehen nach einer Leerstelle und der Phrasenmarkierung, es sei denn, es folgt eine weitere lexikalische Einheit innerhalb der Klammerung (<–/sind <!1 sin’> wir/– vs. –/sind wir/– <!1 wa>).

3.2.2.2.1 Wiederholung / Korrektur

Symbol:

+/.. ../+

Beispiele:

```
.... wie gebe ich +/das Wetter/+ +/da=/+ den Ort auf
dem Display ein...
... +/im Sep=/+ im September ...
... die Woche +/von/+ <"ah> mit Freitag ...
... also +/das/+ +/das/+ das #zweite ...
```

Definition:

Bei Wiederholung/Korrektur werden Teile von Äußerungsphrasen wiederholt oder korrigiert. Dabei muss der wiederholte Teil nicht wortwörtlich sein, sondern auch syntaktisch ähnliche Strukturen werden als wiederholt oder korrigiert markiert.
Markiert werden
– der Anfang des im folgenden Wiederholten/Korrigierten (Reparandum)
– die Abbruchstelle, nach der dann wiederholt oder korrigiert wird (Reparatum).

Das "Reparandum" wird so geklammert, daß nach dessen Herausnahme zusammen mit der verbleibenden Wiederholung/Korrektur ein nahezu korrektes Satzgefüge entsteht.

Transliterationskonvention:

Direkt vor die erste lexikalische Einheit des Reparandums wird ohne Leerstelle die Anfangsklammer (+/) gesetzt.
Direkt an die letzte lexikalische Einheit des Reparandums, an der Satzgefüge–Abbruchstelle, folgt ohne Leerstelle die Endklammer (/+) .

Wird die Korrektur/Wiederholung noch einmal korrigiert/wiederholt, dann wird wieder Anfang und Ende des neuen Reparandums geklammert.

3.2.2.2.2 False Starts (oder Neustart)

Symbol:

–/.. ../–

Beispiele:

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

Definition:

Der Sprecher beginnt eine Äußerung und bricht diese ab (False Start). Nach der Abbruchstelle erfolgt keine Wiederaufnahme, der Sprecher beginnt einen neuen Gedankengang (Neustart). Markiert wird der gesamte False Start, also der Anfang des abgebrochenenen Satzgefüges, meistens nach einem Interpunktionszeichen, und die Abbruchstelle.

Transliterationskonvention:

Direkt vor die erste lexikalische Einheit des abgebrochenen Satzgefüges wird ohne Leerstelle die Anfangsklammer(–/) gesetzt.
Direkt an die letzte lexikalische Einheit des abgebrochenen Satzgefüges, an der Satzgefüge–Abbruchstelle, folgt ohne Leerstelle die Endklammer (/–) .

Anmerkungen:

False Starts können nicht am Turnende stehen, da kein neuer Gedanke mehr folgt. In diesem Falle wird ein Turnabbruch (<*T>t) protokolliert.

3.2.3 Nonverbale artikulatorische Produktionen

Definition der Elementklasse:

Artikulatorische Produktionen des Sprechers ohne erkennbaren semantischen Gehalt heißen nonverbale artikulatorische Produktionen.

Hierunter fallen:

1. Atmen
2. Häsitationen

3. vollkommen unverständliche Äußerungen
4. artikulatorische Geräusche wie Lachen

Verschriftungsregeln:

Alle nonverbalen artikulatorischen Produktionen stehen zwischen spitzen Klammern (<>). Sie können als zeitintensive Turnelemente sprecherüberlagert sein, Häsitationen oder Unverständliches können auch aktiv Beiträge des anderen Sprechers überlagern. Überlagernde artikulatorischen Geräusche werden nach den Konventionen für Geräuschüberlagerung protokolliert.

Elementumgebung:

Diese Turnelemente können **nicht vor** Interpunktionszeichen stehen.

3.2.3.1 Atmen

Symbol:

<A>

Beispiele:

```
... !KEYComputer , <A> Kinoprogramm .
```

Definition:

Atmen als einzelnes Turnelement wird bei deutlich hörbarem Ein– oder Ausatmen protokolliert.

Transliterationskonvention:

Atmen wird mit <A> transliteriert.

Anmerkungen:

Ausatmen nach einem Plosiv wird nicht als Atmen, sondern als gezögerte Plosion markiert (<Z>).
Treffen Atmen und Interpunktionszeichen zusammen, dann steht <A> immer nach dem Interpunktionszeichen.

3.2.3.2 Häsitationen

Symbol:

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

Beispiele:

```
... <"ah> f"ur ~Bayern ...
... <"ahm> f"ur ~M"unchen...
... -/ich denke/- <hm> also , bei mir ginge es sehr
gut ...
... <Ger"ausch> <A> <h"as> f"ur ~M"un<Z>chen...
```

Definition:

Häsitationen (auch gefüllte Pausen genannt) sind Zögerungen, die als Einzelereignis zwischen Äußerungen zu finden sind (im Gegensatz zu den gezögerten Lauten innerhalb eines Wortes <Z>).

Transliterationskonvention:

Häsitationen stehen zwischen spitzen Klammern.
Die Vielzahl der möglichen Häsitionen, die mittels der Orthographie nicht entsprechend repräsentiert werden kann, wird mit einer der vier Häsitionsklassen transliteriert:

- <"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.

Achtung:

Steht eine Häsitation am Turnende oder besteht ein Turn nur aus Häsitation (= minimaler Turn), steht danach kein Interpunktionszeichen, da es sich nicht um ein lexikalisches Element handelt.

UnverständlicheSprachproduktionen

Symbol:

<%>

Beispiele:

... aber <A> <"ah> <%> <"ahm> wie w"ar' 's denn ...

... <%> fr"uher geht 's leider nicht ...

... ich k"onnte am <%> geht% das bei Ihnen ?

Definition:

Wird eine Äußerung eines Sprechers überhaupt nicht verstanden, wegen akustischer Überlagerung, oder unverständlicher Artikulation, kann also auch durch den Kontext nicht identifiziert werden (im Gegensatz zu schwerverständlichen Äußerungen), dann wird an Stelle dieser Äußerung das Symbol für vollkommen unverständliche Sprachproduktion transliteriert.

Transliterationskonvention:

Für Unverständliches steht <%> .

Anmerkungen:

<%> steht gegebenenfalls immer nach einem Interpunktionszeichen. Im Zweifelsfall muß auf das Interpunktionszeichen verzichtet werden.

3.2.3.4 Nonverbale artikulatorische Geräusche

Symbol:

<Lachen>
<Ger"ausch> (= sonstiges artikulatorisches oder nicht−identifizierbares artikulatorisches Geräusch, wie Schmatzen, Räuspern etc.)

Beispiele:

```
... <Ger"ausch> <A> <h"as> f"ur ~M"un<Z>chen ...
... <Ger"ausch> <A> <"ah> <P> Kinoprogramm .
... <Ger"ausch> <A> !KEYComputer , Kinoprogramm ...
... einverstanden . <P> <Ger"ausch> auf Wiedersehen
...
... <Ger"ausch> <A> <"ahm> <P> ~<*EN>Eyes+Wide+Shut .
...
... <Ger"ausch> kann ich die Uhrzeit noch haben ? ...
... <Lachen> richtig .<"ahm> <P> !KEYComputer , Ende .
...
```

Definition:

Nonverbale artikulatorische Geräusche bezeichnen artikulatorische Produktionen des Sprechers, die den Sprachfluß unterbrechen oder Äußerungen überlagern.

Transliterationskonvention:

Alle nonverbalen artikulatorischen Geräusche stehen zwischen spitzen Klammern.
Um die Vielzahl der artikulatorischen Geräusche einzuschränken (Husten, Räuspern, Schmatzen), werden alle außer **<Lachen>** mit der Allgemeinklasse <Ger"ausch> verschriftet.

Anmerkungen:

Falls andere Personen im Hintergrund ein artikulatorisches Geräusch verursachen, wird dies auch als artikulatorisches Geräusch, unter Umständen als Geräuschüberlagerung, in die Transliteration mit übernommen.
Hintergrundreden wird mit <Ger"ausch> verschriftet.
In diesem Fall wird aber nicht Sprecherüberlagerung, sondern Geräuschüberlagerung protokolliert.

Artikulatorische Geräusche stehen gegebenenfalls nach einem Interpunktionszeichen.

3.2.4 Technische Geräusche und Artefakte

Symbol:

<#>

Beispiele:

```
... de<Z>m <:<#> #neunundzwanzigsten:> August ...
... f"ur welche Region w"unschen Sie eine
Wettervorhersage ? <#> ...
... um mich zu aktivieren , verwenden Sie bitte das
Kommandowort !KEYComputer oder !KEYSmartKom . <#> ...
... <#> das tut mir leid ...
... ich k"onnte Ihnen <:<#> vorschlagen:> ...
```

Definition:

Bei technischen Geräuschen und Artefakten handelt es sich um diejenigen hörbaren Turnelemente, die nicht durch die Artikulation eines Sprechers produziert wurden.
Diese Ereignisse treten zum Teil im Zusammenhang mit der Bandaufnahme des Dialogs auf (Mikrofonberührungen, Klicken bei Knopfdruckaufnahmen) oder werden durch Gegenstände im Hintergrund verursacht (Rascheln, Klopfen, Klingeln). Sie sind entweder während Sprechpausen zu hören oder überlagern lexikalische Einheiten.

Transliterationskonvention:

Technische Geräusche werden in spitzen Klammern und # verschriftet.
Sie können als einzelne Ereignisse auftreten oder lexikalische Einheiten überlagern.

Anmerkungen:

Technische Geräusche stehen immer nach Interpunktionszeichen.

3.2.5 Sprechpausen

Symbol:

<P>

Beispiele:

```
... <"ah> f"ur ~Bayern . <P> wiederholen ...
... gut , danke sch"on . <"ahm> <P> !KEYComputer ,
Wetter ...
```

Definition:

Wenn ein Sprecher seinen Sprachfluß kurz stoppt, um beispielsweise Grenzen zu markieren oder kurz nachzudenken, wird eine Pause protokolliert. Während einer Pause ist im Signalfile noch ein kleiner Amplituden–Ausschlag bedingt durch normales Hintergrundrauschen oder Aufnahmerauschen zu sehen (im Gegensatz zur technischen Aufnahmeunterbrechung).

Transliterationskonvention:

Pausen werden mit <P> transliteriert.
Vom Sprecher selbst verursachte Ereignisse, die während einer Sprechpause stattfinden, wie Atmen oder artikulatorische Geräusche, überlagern nicht die Pause, sondern werden als eigene Elemente transliteriert. In diesen Fällen wird dann keine Pause markiert. Anders

verhält es sich bei technischen Geräuschen. Treten sie in Sprechpausen auf, wird die Sprechpause ebenfalls markiert.

Beispiel:

```
... wir kommen am <P> <Geräusch> <A> <h"as> #dritten
...
```

Anmerkungen:

Sprechpausen am Anfang oder Ende eines Turns werden **nicht** transliteriert.
Pausen, die mit Interpunktionszeichen zusammentreffen, stehen nach dem Interpunktionszeichen.

Pausen können als zeitintensives Turnelement passiv sprecherüberlagert sein. Das heißt, der Dialogpartner kann in eine Sprechpause des anderen Partners hineinsprechen, überlagert aber damit den Beitrag des Partners, weil dieser noch nicht zu Ende ist.

3.2.6 Szenario–bedingte Pause

Symbol:

<PP>

Definition:
Diese Pausen unterscheiden sich von den normalen Sprechpausen, weil sie durch die Aufnahmesituation bedingt künstlich erzeugt werden und nicht wie Sprechpausen syntaktische oder semantische Funktionen im Sprachfluß wahrnehmen.

Transliterationskonvention:

Künstliche Pausen werden mit <PP> transliteriert. Ansosnten werden sie wir normale Pausen <P> behandelt.

3.2.7 Akustische Überlagerung

3.2.7.1 Sprecherüberlagerung

Symbol:

..n@ (passive Sprecherüberlagerung lexikalischer Einheiten)

@n.. (aktive Sprecherüberlagerung lexikalischer Einheiten)

..n@> (passive Sprecherüberlagerung sonstiger Ereignisse)

<@n.. (aktive Sprecherüberlagerung sonstiger Ereignisse)

Beispiel (Verbmobil):

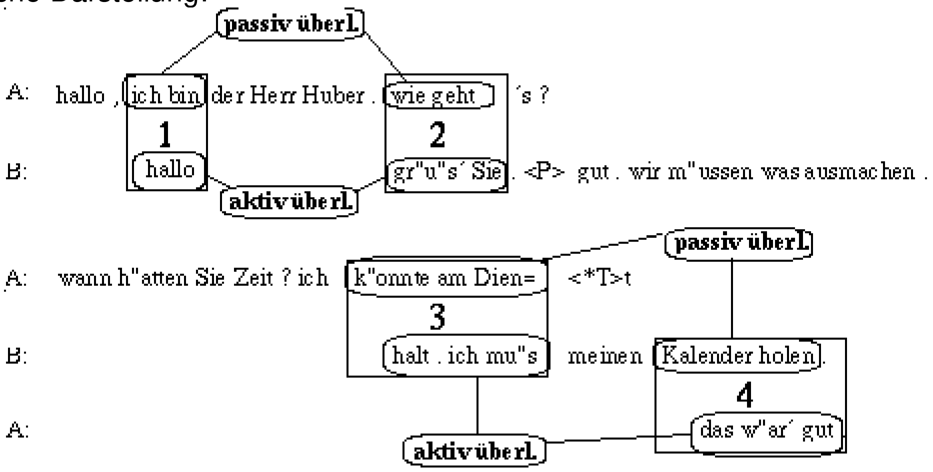
```
Spr A: hallo1@ , <"ah>1@> ich bin der Herr ~Huber2@.
Spr B: @1hallo . <P> @2ah @2ja . wie geht es
<:<#>Ihnen3@:> denn3@ ?
```

```
Spr A: @3gu<Z>t . <@3<A> wir m"ussen noch einen Termin
aus_ <"ah> _machen4@ . wann4@ k"onnen Sie ?
Spr B: @4ah @4ja . ich k"onnte am ...
```

<u>Definition:</u>

Bei Dialogen, bei denen die Sprecher nicht durch technische Gegebenheiten dazu gezwungen sind, nur dann zu sprechen, wenn der Gesprächspartner gerade nicht spricht (z.B. bei Knopfdruckdialogen), kann es zur gegenseitigen Überlagerung von Gesprächsbeitragselementen kommen (bestätigende Interjektionen, "dazwischenreden", unterbrechen usw.).

schematische Darstellung:



Sprecher B überlagert das "ich bin" im ersten Beitrag von Sprecher A mit "hallo" und das "wie geht ´s" mit "gr"u"s Sie".
Sprecher B unterbricht den zweiten Beitrag von Sprecher A, wobei sich die Teile "k"onnte am Dien=" (Sprecher A) und "halt, ich mu"s" (Sprecher B) überlappen. Sprecher A bricht hier seinen Dialogbeitrag ab.
Sprecher B wird dann in der Fortführung seiner Äußerung bei "Kalender holen" von "das w"are gut" (Sprecher A) überlagert.

Aus der Darstellung wird klar,
– daß jeder Sprecherbeitrag auch bei zeitlicher Überlappung in einem eigenen Turn steht
– daß passive (das Überlagerte) und aktive (das Überlagernde) Überlagerung durch gleichzeitiges Auftreten ein Paar bilden.

<u>Transliteration:</u>

Bei einer Sprecherüberlagerung werden zeitintensive Turnelemente, die gleichzeitig stattfinden, markiert.

sprecherüberlagerte /überlagernde Turnelemente können sein:
Passiv überlagerte Elemente, die als sprecherüberlagert markiert werden:
– lexikalische Einheiten
– Atmen
– Sprechpausen
– Häsitationen
– unverständliche artikulatorische Produktionen (<%>)

Aktiv überlagernde Elemente, die als sprecherüberlagernd markiert werden:
– lexikalische Einheiten
– Häsitationen
– unverständliche artikulatorische Produktionen (<%>)

ABER: artikulatorische und technische **Geräusche, die aktiv Sprachproduktionen überlagern**, werden als **Geräuschüberlagerung** im überlagerten Turn transliteriert!

Bei kurz aufeinanderfolgenden Überlagerungen stellt sich manchmal die Frage, ob die Numerierung hochgezählt wird oder nicht.
Hier gilt: Die Numerierung ändert sich, wenn **beim passiven Sprecher** folgende Einheiten zwischen die Überlagerung treten:
1. lexikalische Einheiten
2. Häsitationen
3. Unverständliches
(entspricht den Einheiten, die aktiv überlagern können.)
Außerdem ändert sich die Numerierung bei Turnwechseln.

Beispiele:

A: am Montag1@ oder Dienstag2@ .
B: @1Mittwoch , <P> <Ger"ausch> @2Donnerstag , Freitag .

A: am Montag1@ <"ahm> oder2@ Dienstag2@ .
B: @1Mittwoch , <P> <Ger"ausch> @2Donnerstag , Freitag .

A: am Montag1@ <A> <Schmatzen> <Lachen> oder1@ Dienstag1@ .
B: @1Mittwoch , <P> <Ger"ausch> @1Donnerstag , Freitag .

Turnwechsel:

A: okay1@ .
B: @1gut .
A: dann2@ auf2@ Wiedersehen .
B: @2tsch"u"s .

Aber:

A: okay1@ , dann1@ auf1@ Wiedersehen .
B: @1gut , @1tsch"u"s .

Kennzeichnung:
Bei jedem passiv überlagerten Element wird die Überlagerungsmarkierung ohne Leerstelle an das Element hinten angehängt (unabhängig von davorstehenden Markierungen wie Abbruch oder schließende spitze Klammer bei z.B. Atmen).
z.B.
Kartoffel1@
Kart=1@
<A>1@>

Jedem aktiv überlagernden Element wird die Überlagerungsmarkierung ohne Leerstelle vorangestellt (unabhängig von eventuell weiteren Symbolen wie Wortklassenmarkierungen oder öffnende spitze Klammer bei z.B. Häsitation)
z.B.
@1Kartoffel
@1#drei
<@1<"ah>

Um die Zusammengehörigkeit einer passiven mit einer aktiven Überlagerung auszudrücken, steht in den Überlagerungsmarkierungen eine Zahl, die für das Überlagerungspaar gleich bleibt (auch bei mehreren betroffenen Elementen).
Diese Zahl wird während des gesamten Dialogs hochgezählt.
(erstes Überlagerungspaar, zweites Überlagerungspaar, drittes ..)
z.B.
A: hallo , ich1@ bin1@ der Herr ~Huber . wie2@ geht2@ ´s ?
B: @1hallo . <P> @2gr"u"s´ @2Sie .

Anmerkungen:

Für sprecherüberlagerte lexikalische Einheiten gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten,
für sprecherüberlagerte sonstige Ereignisse gelten die entsprechenden Konventionen zur Transliteration dieser Ereignisse.

Agrammatische Phrasenklammern und Geräuschüberlagerungsklammern stehen nach der passiven bzw. vor der aktiven Überlagerungsmarkierung.

Bei unterbrochenen lexikalischen Einheiten wird jeder Wortteil einzeln markiert. Die Markierungen für Sprecherüberlagerung stehen dann nach dem Unterstrich der Unterbrechungsstelle bzw. vor dem Unterstrich der Wiederaufnahme.
z.B.
A: Hutschachtel . wiederholen1@ Sie1@ das1@ .
B: @1*Hund_ <@1<"ah> @1_schachtel oder Hut2@_ +/sche=2@/+ _schachtel .
A: @2genau .

Siehe auch 3.3.1. Reihenfolge der Markierungssymbole .

Allgemeines:

Überlagert ein Sprecher A immer wieder den Beitrag des anderen (B) mit kleinen Einwürfen, kann zwischen den überlagerten Elementen auch eine Sprechpause <P> beim Sprecher B stehen. Falls während der Pause Geräusche zu hören sind, werden diese als Geräuschüberlagerungen im **Turn des Sprechers** B transliteriert.

1. Fall

Sprecher A spricht seinen Turn.

Sprecher B wirft ab und zu etwas ein (d.h. Sprecher B überlagert das Gesagte von Sprecher A) . Falls zwischen den Einwürfen von Sprecher B Sprechpausen <P> auftreten, in denen Sprecher A zu hören ist, wird dieses als (artikulatorische) Geräuschüberlagerung im Turn von Sprecher B transliteriert.

2. Fall

Sprecher A spricht seinen Turn und macht zwischendrin eine Sprechpause <P>, in der Sprecher B etwas sagt (d.h. Sprecher B überlagert die Sprechpause von Sprecher A). In diesem Fall wird die Pause von Sprecher A wie alle anderen sprecherüberlagerten Elemente seines Turns behandelt. <P>x@>

Um die Konventionen parsbar zu halten, müssen sich die Symbole für überlagernde/überlagerte lexikalische Einheiten (..n@ / @n..) und denen der sonstigen überlagernden Elemente wie Atmen, Pausen oder Häsitationen (..n@> / <@n..) etwas unterscheiden. Das zu beschreibende Phänomen ist davon jedoch nicht betroffen.
Die Zählung der überlagerten Stellen im Dialog erfolgt **unabhängig von den verschiedenen Markierungssymbolen**, also bei Sprecherüberlagerung lexikalischer

Einheiten mit lexikalischen Einheiten, anderer Elemente mit anderern Elementen oder einer Mischung der beiden Markierungsmöglichkeiten (..1@ / @1.. , ..2@> / <2@.. , ..3@ / <@3.. , ..4@> / @4.. , ...).

3.2.7.2 Geräuschüberlagerung

Symbol:

<:<..> ..:>

<..> steht für alle möglichen artikulatorischen und technischen Geräuschkategorien (z.B. <:<#> ..:>).

Beispiele:

```
... de<Z>m <:<#> #neunundzwanzigsten:> August ...
... bei <:<Lachen> mir:> terminlich sehr ung"unstig
...
... <:<#> <Lachen> k"amen:> <:<#> <Lachen> mir:>
<:<#> <Lachen> die:> <:<#> <Lachen> Monate:> April ...

... <:<Lachen> ich:> <:<#> <Lachen> habe:> <:<Lachen>
heute:> keine Zeit ...
```

Definition:

Lexikalische Ereignisse können von **einem oder mehreren** Geräuschen überlagert sein,
– die entweder von einem Sprecher produziert (z.B. Husten) oder verursacht werden (z.B. in das Mikrofon blasen) oder
– die im Hintergrund zu hören sind wie Rascheln, Klopfen oder Hintergrundgemurmel.

Transliterationskonvention:

Eine geräuschüberlagerte lexikalische Einheit wird zusammen mit der entsprechenden Geräuschkategorie−Bezeichnung mit **<:** und **:>** geklammert.
Nach der öffnenden Klammer (<:) folgt das überlagernde Geräusch, eine Leerstelle und eventuell weitere überlagernde Geräusche plus jeweils folgender Leerstelle,
danach wird die überlagerte lexikalische Einheit transliteriert (gegebenenfalls mit entsprechenden Anfangssymbolen wie bei Namen oder Sprecherüberlagerung).
Direkt im Anschluß an die lexikalische Einheit folgt ohne Leerstelle die schließende Klammer (:>), unabhängig mit welchem Symbol die lexikalische Einheit abschließt (Abbruch, schwerverständlich, Wortunterbrechung oder Sprecherüberlagerung).

Anmerkungen:

Nur bei lexikalischen Einheiten wird Geräuschüberlagerung transliteriert.
Für jede überlagerte lexikalische Einheit wird die Geräuschüberlagerung **extra** markiert, **eine Klammerung mehrerer lexikalischer Einheiten**, die von einem durchgehenden Geräusch überlagert werden, **ist nicht möglich!**
Auch bei Wortunterbrechung wird jeder Wortteil **extra** mit Geräuschüberlagerung markiert.
Falls ein ganzer Turn mit einem durchgehenden Geräusch überlagert ist, kann auf die

Einzelmarkierung verzichtet werden und stattdessen in einem globalen Kommentar darauf hingewiesen werden.

Für geräuschüberlagerte lexikalische Einheiten gelten ansonsten die allgemeinen Konventionen für lexikalische Einheiten.

Die öffnende Geräuschüberlagerungsklammer einschließlich der überlagernden Geräusche steht
– nach der öffnenden Klammer der agrammatischen Phrase ohne Leerstelle
– vor allen anderen Symbolen wie Sprecherüberlagerung, technisch vorne abgeschnittenem Wort oder Wortklassifizierungssymbolen oder dem Unterstrich einer Wortfortführung bei Wortunterbrechung.

Die schließende Überlagerungsklammer steht
– nach eventuellen Symbolen, die artikulatorische Besonderheiten der lexikalischen Einheit bezeichnen,
– nach dem Symbol für passive Sprecherüberlagerung,
– nach dem Unterstrich einer Wortunterbrechung,
– vor der Endmarkierung bei agrammatischen Phrasen.

Siehe auch 3.3.1. Reihenfolge der Markierungssymbole.

3.2.8 Lokale Kommentare

Symbol:

<;..>

Beispiele:

```
... zum Beispiel <;"ubersteuert> ...
... am #eins% , #zwei% , #drei% <;Zahlen gefl"ustert>
#dritten M"arz ...
```

Definition:

Hinter jedem Turnelement kann im Prinzip ein lokaler Kommentar stehen. In einem lokalen Kommentar werden Bemerkungen zu Besonderheiten einzelner Turnelemente direkt nach den betreffenden Turnelementen vermerkt (im Gegensatz zum globalen Kommentar nach dem gesamten Turn). Das können bestimmte Sprechstile wie "emphatisch" oder "geflüstert" sein oder Bemerkungen bezüglich falsch verwendeter Grammatik.

Transliteration:

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

Innerhalb der Kommentare gelten die normalen orthographischen Regeln, bei Umlauten wird TeX−Schreibweise verwendet.

Sonderzeichen wie Abkürzungspunkte, Anführungszeichen (abgesehen von Umlauten) oder Spiegelstriche sollten in den Kommentaren nicht verwendet werden.

3.2.9 Off–Talk

Symbol:

<ROT> –> Read Off–Talk

<OOT>–> Other Off–Talk

Beispiel:

```
... ich<OOT> mu"s<OOT> [NA] den<OOT> !KEYAladdin<OOT>
 [NA] ber"uhren<OOT> [PA] [B3 fall] , aha<OOT> [B3
cont] .
```

```
...<A> chinesisches<ROT> <!1 schinesische>
Fastfood<ROT> [NA] [B2] , asiatische<ROT>
Spezialit"aten<ROT> [NA] [B2] , ge"offnet<ROT> [B2] ,
Freitag<ROT> [B3 cont] .
```

Definition:

Als Off–Talk werden sämtliche Äußerungen des Benutzers bezeichnet, die nicht direkt als Anweisung, Rückmeldung oder Frage an das System gerichtet sind.

Es werden 2 Formen von Off–Talk unterschieden:
– Selbstgespräch / Lautes Denken / Fluchen
=> Diese Form bekommt den Marker <OOT>, was für **O**ther **O**ff–**T**alk) steht

– Ablesen von präsentiertem Text
=> Diese Form bekommt den Marker <ROT>, was für **R**ead **O**ff–**T**alk steht

Transliterationskonvention:

Der Marker für Off–Talk <OOT>/<ROT> wird als letzter anschließender Marker bei jedem Wort angehängt.

Nur das Symbol von Sprecherüberlagerung darf nach dem Off–Talk–Marker stehen, andere Markierungen am Wortende stehen davor. (Bsp. Ah<OOT>1@, das<Z><OOT> ist%<OOT>...)

Mit Off–Talk–Markern können nur lexikalische Einheiten versehen werden. Sonderereignisse wie Häsitationen oder unverständliche Passagen, können nicht Off–Talk sein, da sie ansich schon eine metasprachliche Funktion erfüllen oder nicht beurteilt werden können.

Anmerkung:

In Zweifelsfällen wird Metatalk **nicht** markiert!

3.2.10 Prosodisches Labeling

3.2.10.1 Phrasengrenzen

Symbole:

[B2]	schwache Phrasengrenze
[B3 rise]	starke Phrasengrenze mit leicht steigender Intonation
[B3 cont]	starke Phrasengrenze, wobei die Intonation nicht signifikant fallend oder steigend ist.
[B3 fall]	starke Phrasengrenze mit leicht fallender Intonation
[B9]	irreguläre Phrasengrenze

Definition:

Gesprochene Äußerungen werden durch intonatorischce Muster untergliedert. Teile einer Äußerung werden beim Sprechen zu sog. Intonationsphrasen (prosodische Einheiten) zusammengefaßt. Dabei werden Einschnitte im Redefluß auditiv wahrgenommen. Diese Einschnitte sollen durch **Phrasengrenzen** markiert werden.

Man unterscheidet dabei verschieden starke Einschnitte:

intermediäre Phrasengrenzen B2: Diese Phrasengrenze tritt innerhalb einer prosodischen Phrase auf (meist nach einer Häsitation), wobei das globale Intonationsmuster bzw. das Sprechtempo erhalten bleibt. Sie ist schwächer als eine B3−Grenze.

Beispiel:

w000_pk1_020_SMA: auf dem Display [PA] [B2] sehen Sie verschiedene Regionen [PA] [B2] f"ur eine Wettervorhersage [PA] [B3 fall] . w"ahlen [NA] Sie einen Ort [PA] aus [B3 fall] .

Intonationsphrasengrenzen B3: Ein starker Einschnitt im Redefluß wird markiert. B3 ist die »normale« Phrasengrenze (und daher meist mit fallender Intonation [B3 fall]). Normalerweise ist die B3−Grenze durch eine Pause markiert, bei schneller Sprechweise kann der Einschnitt meist nur durch starke Schwankungen in der Grundfrequenz oder durch einen Wechsel im Sprechtempo erkannt werden. Normalerweise tritt eine B3−Grenze mit einem PA auf. In bestimmten Fällen kann auch ein NA stehen. (Beispiel: <"ahm> [B2] Moment [NA] [B3 fall])

Beispiele:

w000_pk1_003_AAA: <"ahm> [NA] [B2] f"ur ~M"unchen [PA] [B3 fall] . <P> ~M"unchen [PA] [B3 rise] ?

w001_pk1_008_SMA: ... verwenden Sie bitte [NA] f"ur eine Eingabe [NA] das Kommandowort [NA] !KEYSmartKom [PA] [B3 cont] oder !KEYComputer [PA] [B3 fall] . ...

Sowohl B2− als auch B3−Grenzen können in Verbindung mit einer Pause auftreten.

irreguläre Phrasengrenzen B9: Wenn der Sprecher zögert oder eine Pause macht, um sich seine weitere Wortwahl zu überlegen, tritt eine irreguläre Phrasengrenze auf (meist in Kombination mit). Die Intonationsphrase wird dann meist unter− bzw. abgebrochen und somit nicht beendet. Dasselbe passiert auch bei Versprechern, die vom Sprecher dann korrigiert werden.B9−Grenzen, die nicht in Kombination mit Zögerungen autreten, lassen sich meist durch eine unnatürliche Pause innerhalb der Phrase erkennen.

Transliteration:

B9 ist die einzige Phrasengrenze, die zwischen unterbrochenen lexikalischen Einheiten auftreten kann!

Alle anderen Phrasengrenzen werden **nach** dem letzten Wort der Phrase gelabelt, abgetrennt durch Leerzeichen und immer **vor** Interpunktionszeichen. Phrasengrenze stehen somit auch immer **vor** einem Geräusch, Pause oder Atmen.

Grundsätzlich gilt: am Satzende steht immer eine Phrasengrenze. Absätze, die noch stärker voneinander abgehoben sind, werden grundsätzlich nicht zusätzlich markiert.

3.2.10.2 Akzente

Symbol:

[NA]	Nebenakzent
[PA]	Phrasenakzent
[EK]	Emphase/Kontrast

Definition:

Hervorgehobene Wörter in der gesprochenen Äußerung werden durch Akzente markiert. Bei den Akzenten werden unterschiedlich starke Betonungen unterschieden. Die Markierung erfolgt bei jeder Phrase einzeln. Die Akzente werden ebenfalls wortweise, also nach dem betroffenen Wort markiert. Sie stehen dabei innerhalb der Phrase.

Man unterscheidet folgende Akzente:

Phrasenakzent [PA]: Dieser Akzent liegt auf dem am stärksten hervorgehobenen Wort der Phrase auf der lexikalisch betonten Silbe. Das ist das Wort mit der wichtigsten Information für den Hörer. Normalerweise tritt dieser Akzent nur einmal pro Phrase auf. Es können aber auch mehrere gleichstarke oder auch kein PA auftreten. Leitlinie ist die Perzeption, nicht ein theoretisches Constraint! (Fälle ohne PA treten typischerweise beim Beiseitesprechen auf. Z.B.

```
Hilfe [NA] [B3 fall] .
```

Beispiel:

```
w001_pk1_000_SMA: hallo [PA] [B3 fall], ich bin
~SmartKom [PA] [B3 fall]. ich kann Ihnen Auskunft [NA]
[B2] "uber das Wetter [NA] [B2] und das Kinoprogramm
[PA] geben [B3 fall] . ich reagiere [NA] auf Ihre
m"undliche [NA] Anweisung [PA] [B3 cont] oder auf
Eingaben [NA] mit dem <*EN>touch<Z> [PA] [B9]
<*EN>screen [B3 fall] . um mich zu aktivieren [PA]
[B2] , verwenden Sie bitte das Kommandowort [NA]
```

```
!KEYComputer [PA] [B2] oder !KEYSmartKom [PA] [B3
fall] . <#>
```

Nebenakzent [NA]: Alle weiteren hervorgehobenen Wörter innerhalb einer Phrase werden durch den Nebenakzent markiert. Die Markierung erfolgt wieder nach dem betroffenen Wort, wobei berücksichtigt werden sollte, daß sich die Markierung immer auf die lexikalisch betonte Silbe bezieht.

Beispiel:

```
>w001_pk1_000_SMA: ... ich reagiere [NA] auf Ihre
m"undliche [NA] Anweisung [PA] [B3 cont] oder auf
Eingaben [NA] mit dem <*EN>touch<Z> [PA] [B9]
<*EN>screen [B3 fall] . ... <#>
```

Emphase/ Kontrast [EK]: Bei besonders starken Betonungen kann anstelle eines Phrasenakzents eine Emphase bzw. ein Kontrast treten. Dieses Etikett wird immer dann verwendet, wenn ein Unterschied zu Vorangegangenem hervorgehoben werden soll.

Beispiel:

```
w001_pk1_019_AAA: ach so [PA] [B3 fall] . <A> <"ahm>
<P> und wie [PA] [B3 rise]? !KEYComputer [PA] [B2] ,
wie [EK] gebe ich +/das Wetter/+ +/da=/+ [PA] [B9] den
Ort [EK] auf dem Display [PA] ein [B3 fall] ?
```

Transliteration:

Akzente werden wie Phrasengrenzen **nach** dem letzten Wort der Phrase gelabelt, abgetrennt durch Leerzeichen und immer **vor** der Phrasenmarkierung und somit **vor** der Interpunktion. Akzente werden vor den Phrasengrenzen verschrfitet, können aber auch ohne Phrasengrenze stehen.

Erhält ein Wort einen Aussprachekommentar, so steht dieser vor der prosodischen Markierung.

.

Anmerkungen:

Zur Transliteration der Prosodie gibt es eine gesonderte Anleitung mit Hörbeispielen, unter:

http://www.phonetik.uni–muenchen.de/smartkom/prosodieanleitung.html

3.3 Tabellen

3.3.1 Reihenfolge der Markierungssymbole bei lexikalischen Einheiten

Agram	Geräusch	Spr. Überlagerung	Anf. Symbol	Lang. Marker	Lex.	Defekt	Off–Talk	Spr. Überlagerung	Geräusch	Agram
+/ –/	<:<#> <Ger"ausch> <Lachen>	@1	# * ~ & !KEY $	<*XX>	Wort <Z>	% =	<ROT> <OOT>	1@	:>	/+ /–

Eine lexikalische Einheit kann aus folgenden Kombinationen bestehen:

Unterbrechung	Anf. Symbol	Lexem	Endsymbol
_(..)	**$**	**Wort**	**+[_]**
<T_>		Wo<Z>rt	–[_]
			––
			<_T>
			=

3.3.2 Liste aller verwendbarer Symbole

<*tEN>	fremdsprachiger Turn (JA,DE, ..)
<*tROT>	Read Off–Talk (Turn)
<*tOOT>	Other Off–Talk (Turn)
..´..	Apostroph (Wort–Reduktion)
..–.. (––)	Bindestrich (Komposita)
..+..	Pluszeichen(Namens–Komposita)
$..	Buchstabierung und Buchstabiersequenzen
~..	Namen
&..	Akronyme
#..	Zahlen
*..	Neologismus

<*XXX>..	fremdsprachiges Wort (FRA,ITA, ..)
!KEY..	Kommandowort
..<Z>..	Zögerung
..%	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
<!n ..>	Aussprachekommentar
. / ? / ,	Interpunktion
+/..	Beginn Wiederholung/Korrektur
../+	Ende Wiederholung/Korrektur
–/..	Beginn False Start
../–	Ende False Start
<A>	Atmen
<"ah>	Häsitation/gefüllte Pause
<"ahm>	Häsitation/gefüllte Pause
<hm>	Häsitation/gefüllte Pause
<h"as>	Häsitation/gefüllte Pause
<%>	unverständliche Sprachproduktion
<Lachen>	nonverbale artikulatorische Geräusche
<Ger"ausch>	nonverbale artikulatorische Geräusche
<#>	technische Geräusche
<P>	Sprechpause
<PP>	szenario–bedingte Pause
@n..	aktive Sprecherüberlagerung

..n@	passive Sprecherüberlagerung
<@n..	aktive Sprecherüberlagerung von Ereignissen
..n@>	passiver Sprecherüberlagerung von Ereignissen
<:<..> ..	Beginn Geräuschüberlagerung
..:>	Ende Geräuschüberlagerung
<;..>	lokaler Kommentar
<ROT>	Read Off−Talk (Wort)
<OOT>	Other Off−Talk (Wort)
[B2]	schwache Phrasengrenze
[B3 rise]	starke Phrasengrenze mit leicht steigender Intonation
[B3 fall]	starke Phrasengrenze mit leicht fallender Intonation
[B9]	irr. Phrasengrenze
[NA]	Nebenakzent
[PA]	Phrasenakzent
[EK]	Emphase/Kontrast

3.4 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 spontansprachlicher 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.