BMBF

Verbmobil
Verbundvorhaben

Aussprachevarianten in der VERBMOBIL-Transliteration - Regeln zur konsistenteren Verschriftung

Susanne Burger
Elke Kachelrieß

Ludwig Maximilians Universität München

Vm

Memo 111
August 1996

August 1996

Susanne Burger
Elke Kachelrieß

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

Tel.: (089) 2180 - 2806
e-mail: burger@sun1.phonetik.uni-muenchen.de

Gehört zum Antragsabschnitt: 14 VERBMOBIL/PHONDAT

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

Abstract:

Dieses Memo soll Regeln zur konsistenteren Verschriftung der VERBMOBIL-Aussprachekommentare auf orthographischer Ebene vorschlagen.
This Memo proposes regularities for a more consistent transliteration of the VERBMOBIL-pronunciation-comments at the orthographic level.

Siehe auch: http://www.phonetik.uni-muenchen.de/VMTechDoks.html

Error reports to burger@phonetik.uni-muenchen.de

Inhaltsverzeichnis

1 Einleitung

Trotz des ausführlichen Handbuchs zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL -3.0 (Kohler at al, Kiel 1994, [1]) und des Konventionslexikons zur Transliteration von Spontansprache (Burger, München 1995, [2]) stößt man bei der Verschriftung von Spontansprache immer noch auf ungelöste Probleme. Die Autoren beschäftigen sich seit einiger Zeit mit den Aussprachekommentaren in der VERBMOBIL-Transliteration, um zu sehen, in wie weit schon anhand dieser Kommentare Aussagen über Aussprachevarianten im Deutschen gemacht werden können. Die Transliteration der Aussprachevarianten erfolgte bisher relativ ungeregelt, was einerseits statistische Untersuchungen erschwerte und andererseits schon beim Transliterieren der VERBMOBIL-Dialoge häufig zu Nachfragen seitens der Transliterierer führte.
Dieses Memo soll konsistente Regeln zur Verschriftung dieser Aussprachekommentare vorschlagen. Die Regeln wurden bei der Münchner und Karlsruher Verschriftung der letzten Ausgaben der VERBMOBIL-CD-ROMs bereits angewandt.

2 Was sind Aussprachekommentare?

2.1 Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL-3.0 (Kohler at al, Kiel 1994)

Das Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL -3.0 [1] sagt dazu (unter III.3.7 (4), S. 17) :

"...
- Aussprachevarianten: Die standardisierte Transliteration eines gesprochenen Wortes kann durch Angaben hinsichtlich Dialektaussprachen, anderer Stilformen (als in der Rechtschreibung impliziert) oder sonstiger Abweichungen (wie z.B. "Deutschslands" mit nichtkanonischem Fugen-s) ergänzt werden. Diese Angaben werden mit "!" eingeleitet, in orthographischer Repräsentation zwischen "< >" gesetzt und stehen hinter der standardisierten Transliteration des Gesprochenen. Nach "!" wird durch eine Ziffer angegeben, auf wieviele vorangehende lexikalische Einheiten sich der Aussprachekommentar bezieht. Die orthographische Repräsentation des Aussprachekommentars wird mit einer Leerstelle von "!Ziffer" getrennt.
..."

2.2 Konventionslexikon zur Transliteration von Spontansprache (Burger, München 1995)

Im Konventionslexikon [2] kann man die Konventionen zur orthographischen Transliteration spontansprachlicher Phänomene und zu Hintergrundgeräuschen unter bestimmten Stichwörtern nachschlagen. Nach einigen Beispielen wird die Symbolik gezeigt, es folgt eine kurze "Gebrauchsanweisung" und eine Liste der möglichen Zeichenkombinationen. Zum Thema Aussprachevarianten findet sich hier folgendes:

"...
Aussprachevariante:

Beispiel:

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

Symbol: <!n ..>

Dialektaussprachen, andere Stilformen oder sonstige Abweichungen werden in korrekter Form nach Duden verschriftet. Die eigentlich geäußerte Version wird hinter der korrekten Form
nach einer Leerstelle,
in spitzen Klammern,
nach einem !,
einer Ziffer, die sich auf die betroffenen lexikalischen Einheiten bezieht,
und einer Leerstelle transliteriert.

vor <!n ..> *darf stehen:*
Leerstelle
..."

2.3 Was leistet ein Aussprachekommentar?

- zeigt an, daß eine Äußerung von der kanonischen Form stark abweicht
- arbeitet nur mit Alphabet und Apostrophen
- gibt im Rahmen der Möglichkeiten Informationen über die Form der Abweichung

- keine enge Transkription, sondern Vorklassifikation von Aussprachevarianten in großen Datenmengen

2.4 Wann wird nicht kommentiert?

- standardisierte Aussprache,
  z.B. warten = /wa:tn/
- schwerverständlich:
  z.B. %Montag
- unverständlich:
  z.B. <%>
- Neologismus/schwerer Versprecher (Versprecher, der zu Wortneuschöpfungen führt):
  z.B. *Maistag
- Wortfragment:
  z.B: Sam=
- End-e Reduzierung bei Nominal-, Adverbial- und Verbalflexion:
  z.B. hab'
- Tilgung des Initialvokals bei "es" und unbestimmtem Artikel:
  z.B. 's, 'ne
- Verschmelzung von Präposition und bestimmtem/unbestimmtem Artikel:
  z.B. ans, f"ur'n
- bereits in den Duden als umgangssprachlich eingegangene Varianten:
  z.B. nix, zwo.

2.5 Beispiel für einen Aussprachekommentar

```
machen wir mal das n"achste <!5 mach´ ma moi des n"akschte>
```

Erklärung:
mach' : die Endsilbe fehlt
ma : ungangssprachlich für wir (eher süddeutsch)
moi : süddeutsch für mal
des : sehr häufige Variante von das
n"akschte : dialektale Variante von nächste, eher alemannisch

3 Warum sollte man über konsistente Regeln nachdenken?

3.1 Aussprachekommentare sind durchaus aussagekräftig

- Dialektausprägung
  Bei der Suche nach Dialogen mit starkem Dialekt bietet es sich an, nach Transliterationen mit vielen Aussprachekommentaren zu suchen.

- welcher Dialekt
  dialekt-typische Aussprachevarianten,
  z.B. “August <!1 Auguscht>” (alemannisch), “Beispiel <!1 Beischbui>” (bairisch)

- soziolektaler Sprachgebrauch
  z.B. umgangssprachliche Aussprachevarianten, wie “Entschuldigung <!1 ’tschuldigung>”, “das <!1 des>”

- Grad der “Wohlartikuliertheit”
  verschliffenes versus akurates Sprechen,
  z.B. “Name ist <!2 Nam’ is’>”, “f”unft”agiges <!1 f”unft”ag’es>”

- Sprachfehler
  ein Beispiel: “drei <!1 darei>”:
  Sprecher könnte Schwierigkeiten mit der kombinierten Artikulation von Plosiv und Trill haben

- Häufigkeit bestimmter Varianten
  (Siehe unter: 5. Erste Daten!)

- statistische Aussagen über große Datenmengen
  Solange wenig segmentierte und gelabelte Daten vorliegen, können Statistiken über die Aussprachekommentare aller Transkriptionen schon über Phänomene im voraus informieren, die mit segmentiertem Material dann genauer erforscht werden könnten.

3.2 Ohne Regeln werden ähnliche Phänomene auf verschiedene Weise beschrieben

Zum Beispiel fanden sich mindestens fünf verschiedene Kommentare zu einer Variante von “jetzt”, die alle das Fehlen von initialem und finalem Laut beschreiben:

jetzt<!1 ’etz’>
<!1 etz’>
<!1 ’etz>
<!1 etz>
<!1 ez> ...

- die Transliterationsebene sollte eher kategorisch arbeiten
- statistische Auswertung wird sonst erschwert
- leichteres Transliterieren auch für Nicht-Phonetiker
- Vorteile für die technische Weiterverarbeitung

4 Wie könnten Regeln aussehen?

4.1 Grundfaktoren

4.1.1 Orthographische Einschränkungen

Wodurch beschränkt die Orthographie mit ihrer begrenzten Symbolik (Alphabet, Apostroph) das Transliterieren von Aussprachevarianten?

- kann viele Lautvarianten nicht oder nur grob repräsentieren (z.B. Schwa-Laut)
- Lautqualitäten wie Längung oder Kürzung sind nur schwer darzustellen
- Fälle, bei denen die Standardaussprache sich von der orthographischen Verschriftung unterscheidet, aber dialektale Varianten sich eher auf die Aussprache einzelner Buchstaben beziehen, können nur schwer beschrieben werden.
  Zum Beispiel bei “st”, das als Standardaussprache /St/ gesprochen wird, muß bei der dialektalen Variante /st/ (z.B. nord-niedersächsisch: /staIn/ im Gegensatz zur Standardaussprache /StaIn/) entweder ganz auf einen Kommentar verzichtet werden oder im Kommentar ein Ersatzbuchstabe wie “`"s`” verwendet werden.
- weder kann die Anzahl elidierter Laute immer eindeutig bestimmt werden, noch bietet die Orthographie hier Beschreibungsmöglichkeiten, weil z.B.

Aneinanderreihungen von Apostrophen in den Transliterationskonventionen nicht vorgesehen sind

- prosodische Besonderheiten sind nicht darstellbar

Wo kann man die Orthographie in den Aussprachekommentaren sinnvoll einsetzen?

- Kategorienbildung
- Markierung von Phänomen-Klassen
- Markierung von Elisionspositionen

4.1.2 Phänomen-Klassen

Folgende Klassen wurden in VERBMOBIL-CD-ROM 1-5, 7,12 bisher gefunden:

- Reduktion/Elision
- Modifikation/Substitution
- Lautaddition
- Enklise

Die hier vorgeschlagenen Regeln für die Verschriftung der Aussprachekommentare basieren vor allem auf den bisher gefundenen Phänomen-Klassen.

4.2 Bezug zwischen Aussprachekommentar und orthographischer Repräsentation

4.2.1 Lexikalische Einheiten

Im Aussprachekommentar sollen nach Möglichkeit betroffene lexikalische Einheiten noch erkennbar sein. Deshalb werden die Aussprachevarianten mit Leerstellen auseinander gehalten.

```
machen wir mal das n"achste <!5 mach´ ma moi des n"akschte>
hab´ ich noch einen <!4 hob i´ no´ an>
```

Aber: Eine Ausnahme bilden das Zusammentreffen von Anfangs- und Endlautelisionen (siehe 4.3.2) und Enklisen (siehe 4.4.2).

4.2.2 Orthographie

Generell gelten im Zweifelsfall die orthographischen Regeln.

Beginnt die orthographische Repräsentation mit einem Großbuchstaben, ist das auch bei der dazugehörigen Aussprachevariante der Fall.

```
Donnerstag <!1 Donnaschag>
```

Wenn die Aussprache nichts anderes nahelegt, bleibt scharfes s, (ß, verschriftet als "s) bzw. Doppel-s in der Aussprachevariante wie in der orthographischen Repräsentation bestehen, ebenso wie Längungs-h, Doppelvokale oder Buchstabenkombinationen, für die eine spezielle Standardaussprache gilt ("st" = /St/).

4.3 Reduktion/Elision

4.3.1 Apostroph

Unterscheidet sich eine Äußerung von ihrer orthographischen Repräsention lediglich durch die Abwesenheit eines oder mehrer Laute, so wird die Elisions*stelle* im Aussprachekommentar mit *einem* Apostroph gekennzeichnet.

```
meinen <!1 mein´>
irgendwie <!1 irgen´wie>
nat"urlich <!1 ´t"urlich>
```

Aber: kein Kommentar bei fehlendem End-e, Reduktion des unbestimmten Artikels und Präpostion/Artikelverschmelzung (siehe 2.4)

4.3.2 Apostroph zwischen Wörtern

Entfällt bei einem Wort der Endlaut/Endlaute, beim folgenden Wort der Anfangslaut/Anfangslaute, so steht zwischen beiden nur *ein* Apostroph, der wiederum nur die Elisionsposition bezeichnet. Die betroffenen lexikalischen Einheiten werden in diesem Fall *nicht* mit Leerstelle getrennt.

```
nach Haus <!2 na´aus>
und dann <!2 un´a´>
schon einen <!2 scho´n>
```

4.4 Modifikation/Substitution

4.4.1 Lautmodifikation

Werden in der Aussprachevariante Laute durch andere ersetzt oder Laute hinzugefügt, dann wird im Aussprachekommentar die so modifizierte Äußerung verschriftet. In diesem Fall steht kein Apostroph, selbst wenn zusätzlich Laute weggelassen wurden.

```
gr"u"s <!1 es>
siebenundzwanzig <!1 sienzwanzesch>
"uberhaupt <!1 "uberhaupts>
```

4.4.2 Enklise

Zwischen enklitisierten Varianten steht kein Apostroph; diese Wörter werden zusammengeschrieben.

```
k"onnen wir <!2 k"omma>
haben wir <!2 hamma>
kannst du <!2 kannste>
```

5 Erste Daten

Alle Statistiken wurden über VERBMOBIL-CD1 - CD5, CD7 und CD12 erstellt. Zum Bearbeitungszeitpunkt lag CD14 noch nicht vor. Es wurden nur in Bonn, Karlsruhe und München produzierte Dialoge verwendet; das deshalb, weil die Aussprachekommentare dieser Dialoge mit Erlaubnis der Produzenten nach den hier beschriebenen Regeln in München korrigiert wurden.

Insgesamt fanden sich 257673 Wörter, wovon 9893 (3.85%) mit Aussprachekommentaren versehen waren.
Innerhalb dieser Kommentare fanden sich 28735 Apostrophe, also 28735 mal waren Lautelisionspositionen markiert worden.

Wörter insgesamt	**257673**
Wörter mit Kommentaren	**9893**
Apostrophe in Kommentaren	**28735**

"Hitliste" der häufigsten Aussprachekommentare:

Platz	Orthographie	Variante	Häufigkeit
1.	ist	is'	692
2.	das	des	592
3.	einen	ein'	216
4.	Ihnen	Ihn'	182
5.	wir	ma	151

6 Weitere Vorhaben

- Vorklassifikationen aus der Transliteration
- Elisionspositionen initial-medial-final
- Modifikationen (dialektal, Versprecher)
- Häufigkeiten bestimmter Varianten
- Statistiken über große Datenmengen

7 Literatur

[1] Kohler,Lex,Pätzold,Scheffers,Simpson,Thon (1994): Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL – 3.0. VERBMOBIL Techdok-11-94, Universität Kiel.

[2] Burger (1995): Konventionslexikon zur Transliteration von Spontansprache. VERBMOBIL Techdok-36-95, Universität München
http://www.phonetik.uni-muenchen.de/VMTraLex.html