

Konventionen zur orthographischen Verschriftung SpeechDat(II)

Christoph Draxler
Institut für Phonetik, Universität München
draxler@phonetik.uni-muenchen.de

Stand: 9.7.97

Einleitung

Die Verschriftung ist orthographisch, d.h. sie erfolgt gemäß der deutschen Rechtschreibung, mit einigen wenigen Markern für spezielle Phänomene gesprochener Sprache (z.B. Häitationen, unverständliche Passagen), artikulatorische und nichtartikulatorische Geräusche (z.B. Lachen, Atmen beziehungsweise Hintergrundgeräusche), und Fehler im Signal (z.B. Signalabbruch).

Die Verschriftung erfolgt in zwei Durchgängen:

1. Abhören des Signals und rein orthographische Verschriftung
2. Erneutes Abhören des Signals und Setzen der Marker

Die Sprachsignale müssen mit Kopfhörern in ruhiger Umgebung abgehört werden.

Groß- und Kleinschreibung

Wörter werden *unabhängig* von ihrer Position in einer Äußerung je nach *grammatischer Klasse* groß bzw. klein geschrieben. Das bedeutet, daß nur Substantive (auch Verbalsubstantive wie z.B. das Gehen), Eigennamen und einige Abkürzungen (z.B. AG, GmbH) groß geschrieben werden - alles andere (also auch die Höflichkeitsform der Personalpronomina Du, Sie, Ihr) klein, auch wenn es am Satzanfang steht (siehe auch den Abschnitt Zeichensetzung).

Rechtschreibung

Die Rechtschreibung erfolgt gemäß einem anerkannten Wörterbuch, z.B. Duden. Sonderfälle sind nur Ziffern, Zahlen und Buchstabierfolgen. Ziffern und Zahlen werden in orthographischer Form verschriftet, wobei Zahlen größer als 100 als mehrere Wörter verschriftet werden, um das Lexikon klein zu halten.

- "101" als ein hundert eins (statt einhunderteins)

In Buchstabierfolgen werden die einzelnen Buchstaben als Folge von Großbuchstaben geschrieben und der Rest orthographisch verschriftet. Wird ein Buchstabe als Laut gesprochen, so wird er in Anlehnung an Phoneme in Schrägstriche eingeschlossen.

- "A B B ß" als A doppel B scharfes S
- "D" als /D/

Zeichensetzung

Es werden keine Satzzeichen (.,!?:-) verwendet, d.h. es gibt keine Sätze im herkömmlichen Sinn. Damit entfällt auch die Großschreibung der Wörter am Satzanfang.

Anführungszeichen, Klammern und Apostrophe werden ebenfalls nicht verwendet.

Aussprache

Bei fehlerhaft ausgesprochenen Wörtern wird das intendierte Wort (soweit es rekonstruierbar ist) verschriftet und mit * direkt am Wortanfang, d.h. ohne Leerzeichen, markiert. Ist ein Wort nicht rekonstruierbar, wird das gehörte Wortfragment ebenfalls mit * markiert.

- “Verschriftigung“ als *Verschriftung
- “achtund nein neunundneunzig“ als *achtund nein neunundneunzig

Unverständliche Sprache

Passagen unverständlicher Sprache werden mit ** markiert, wobei jeweils ein Leerzeichen vor und nach dem ** steht.

- du bist ein solches ** wie ich noch nie eines getroffen habe

Geräusche

Geräusche werden mit folgenden Markern in eckigen Klammern [] verschriftet

1. [fil] für Häitationen (z.B. ähm, äh, öh, hm, usw. zwischen Wörtern)
2. [spk] für Sprechergeräusche (z.B. Lachen, Schmatzen, Husten, usw.)
3. [stat] für gleichbleibende Hintergrundgeräusche (z.B. im Auto, Straßenlärm, usw.)
4. [int] für Einzelgeräusche (z.B. Telefon, Türklingeln, Musik, Gespräch im Hintergrund)

Der Marker muß vor das erste betroffene Wort gesetzt und durch Leerzeichen abgetrennt werden:

- ich weiß nicht [fil] da muß [int] seid doch mal leise ich erst nachdenken

Prosodische Marker (optional)

Die außergewöhnliche Längung einzelner Silben oder Laute im Wort kann durch : an entsprechender Stelle im Wort gekennzeichnet werden.

- du hast ja keine A:hnung

Signalabbruch

Signalabbrüche werden mit einer ~ zu Beginn oder am Ende einer Verschriftung markiert, wobei die ~ direkt vor das erste bzw. hinter letzte Wort der Verschriftung gesetzt wird.

- ~undneunzig
- ~haben wir leider feststellen müssen daß sie die vorgeschriebene Mindestgeschw~
- drei hundert achtundsiebz~

Abgebrochene Wörter werden nicht restituiert.

Beurteilung der Signalqualität (optional)

Eine Beurteilung des Sprachsignals erfolgt in einer separaten Log-Datei. Darin werden Signale wie folgt beurteilt:

1. GARBAGE wenn das Signal nur Hintergrundgeräusche enthält oder völlig unpassende Sprache

2. NOISE wenn neben dem Sprachsignal deutliche Hintergrundgeräusche zu hören sind
3. OTHER wenn das Signal eine oder mehrere Häitationen, ungewöhnliche Aussprache oder teilweise unverständliche Sprache enthält
4. OK sonst