

# Ein Schema zur phonetischen Annotation von Füllpartikeln und ihres Kontextes\*

Version 1.0

Malte Belz

*Institut für deutsche Sprache und Linguistik  
Humboldt-Universität zu Berlin*

16. Juli 2019

DOI:

<https://doi.org/10.18452/20163>

## Zusammenfassung

Diese Annotationsrichtlinie schlägt ein Schema zur phonetischen Annotation von Füllpartikeln und ihres Kontextes mithilfe einer Mehr-Ebenen-Annotation vor und demonstriert es anhand dreier Beispiele. Das Schema kann prinzipiell auf andere Sprachen übertragen werden. Es enthält ausführliche tabellarische Beschreibungen der Annotationswerte und ihrer Beziehung zu Werten auf anderen Ebenen.

## English abstract

This annotation guideline proposes a scheme for the phonetic annotation of filler particles (filled pauses) and their context using multi-level annotation and demonstrates it by three examples. In principle, the scheme can be transferred to other languages. It contains detailed tabular descriptions of the annotation values and their relationship to values on other levels.

---

\* Dieser Artikel ist Teil meiner in der Entstehung befindlichen Dissertation.

## 1 Füllpartikeln

Als Füllpartikeln bezeichne ich phonetisch-deskriptiv und funktional unvoreingenommen solche linguistischen Entitäten, die genuin in gesprochener Sprache vorkommen, deren orthographischen Repräsentationen im Deutschen mit *äh*, *ähm*, und *hm* wiedergegeben werden und die keine sekundäre Verwendung lexikalischer Wörter darstellen.<sup>1</sup> Dies schließt jedoch nicht aus, dass erstens phonetische Varianten realisiert werden können und zweitens phonetische Varianten möglich sind, für die keine einheitlichen graphematischen Realisierungen existieren, wie bei stark glottalisierte Formen oder Clicks. Keine Füllpartikeln sind extralinguistische Geräusche wie Husten, Räuspern, Lachen, etc. Diese Definition bildet die Grundlage für die Identifizierung von Füllpartikeln mithilfe des Schemas in Abschnitt 3.

## 2 Motivation

Die in der Literatur verwendeten symbolphonetischen Repräsentationen für Füllpartikeln im Deutschen sind divers. Neben vokalischen, vokalisch-nasalen und nasalen Repräsentationen mit kurzen und langen Vokalen unterschiedlicher Qualität (vgl. Tabelle 1) finden sich auch Sequenzen glottaler Plosive ohne deutlich wahrnehmbaren Vokal (Belz 2017) und Klicklaute (Trouvain und Malisz 2016).

**Tabelle 1:** Symbolphonetische Beschreibungen von deutschen Füllpartikelrepräsentationen in der Literatur.

Form	Referenz
[ə əm ɐ ɐ: ɐ:m] [ø: ø:] [œ œm] [hm]	Batliner u. a. (1995)
[e e: em e:m] [æ æm]	Rasoloson (1994)
[ɛ ɛ: ɛ:m]	Batliner u. a. (1995), Rasoloson (1994) und Schwitalla (2002)
[ɛm]	Batliner u. a. (1995) und Rasoloson (1994)
[ʔɛ ʔɛ: ʔɛm ʔɛ:m] [ʔœ ʔœ: ʔœ:m]	Willkop (1988)
[m]	Leeuw (2007)

Nicht nur die Vokalqualität wird unterschiedlich beschrieben, auch die Beschreibung der segmentalen Bestandteile variiert, beispielsweise durch die Einfügung eines glottalen Plosivs vor dem Vokal. Zudem werden Sequenzen glottaler Plosive ohne deutlich wahrnehmbaren Vokal beobachtet (Belz 2017). Auch Klicklaute (oder Clicks) können als Füllpartikel realisiert werden (Trouvain und Malisz 2016).

<sup>1</sup>Andere Bezeichnungen für Füllpartikeln sind *gefüllte Pausen*, *Füller*, *Häsitationen*, *Häsitationsmarker*, *Verzögerungsmarker*, *Diskursmarker*, und viele weitere.

Die große Variabilität in der auditiven, symbolphonetischen Beschreibung von Füllpartikeln spricht dafür, dass die tatsächlich realisierten akustischen Formen nochmals diverser sind. Für Forschungsfragen zur allgemeinen phonetischen Variabilität von Füllpartikeln und zu ihrer Variabilität in bestimmten Kontexten besteht also das Desiderat eines phonetischen Annotationsschemas. Aufgrund der hohen Subjektivität von symbolphonetischen Lautkategorisierungen wird auf eine genaue Erkennung des Vokals verzichtet. Stattdessen arbeitet das Schema mit einem Platzhalter für den Vokal (*E*), dessen Formanten dann akustisch aus dem Signal heraus ermittelt werden können.

### 3 Phonetische Annotation

Das hier vorgeschlagene Annotationsschema enthält mehrere Ebenen. Es kann prinzipiell in allen Tools angewendet werden, die eine Mehr-Ebenen-Annotation unterstützen, beispielsweise in Praat (Boersma 2001) oder EXMARaLDA (Schmidt und Wörner 2009). Das Schema ist sowohl auf Monologe als auch auf Dialoge oder Gespräche mit mehreren Teilnehmenden anwendbar. Tabelle 2 visualisiert das Schema mithilfe des fiktiven Beispiels *und ⟨P⟩ ähm ja dann*, wobei ⟨P⟩ für eine stille Pause<sup>2</sup> steht.

**Tabelle 2:** Mehr-Ebenen-Annotation von Füllpartikeln. Der senkrechte Trennstrich ‚|‘ stellt eine Intervallgrenze dar. `dipl` enthält die Transliteration, ⟨P⟩ steht für eine stille Pause. Auf `fp` steht *fv* für eine Füllpartikel, *ap* für eine stille Pause im linken und *ps* für ein Segment im rechten Kontext. `segm` enthält die Segmente als SAMPA-Symbole (mit Ausnahmen: *E* steht für eine beliebige Vokalqualität, ‚\_‘ für eine stille Pause). Auf `phon` steht *mod* für modale Phonation.

Akustisches Signal						Ebene	Bezug	Alignierung mit	
und	⟨P⟩		<i>ähm</i>		ja	dann	<code>dipl</code>	Signal	Signal
	<i>ap</i>		<i>fv</i>		<i>ps</i>		<code>fp</code>	<code>dipl</code>	<code>dipl</code> und Signal
	_	?	<i>E</i>	<i>m</i>	<i>j</i>		<code>segm</code>	<code>fp</code>	<code>fp</code> und Signal
			<i>mod</i>				<code>phon</code>	<code>segm</code>	<code>segm</code>

Die Ebene `dipl` enthält hier die diplomatische Transliteration als Annotation auf dem akustischen Signal; sie kann jedoch auch andere schriftliche Repräsentationen gesprochener Daten enthalten. `dipl` ermöglicht die orthographische Annotation der Füllpartikelexponenten. Sollte keine Transliteration vorhanden sein, können Füllpartikeln direkt mit Bezug zum akustischen Signal annotiert werden.

<sup>2</sup>Für stille Pausen existieren verschiedene untere Schwellen. Das hier beschriebene Schema gibt keine kategorische Schwelle vor, orientiert sich aber an dem Wert von ca. 50 ms.

Die Ebene **fp** dient der Annotation von Füllpartikeln und ihres direkten segmentalen Kontextes (vgl. Tabelle 3).<sup>3</sup> Die Annotationswerte sind hierarchisch aufgebaut und zweistellig. Grenzen werden am Nulldurchgang aufsteigender Flanken im Oszillogramm gesetzt.

**Tabelle 3:** Annotation von Füllpartikeln mit direktem Kontext.

Annotationsebene	Annotationsart	Beschreibung
	<b>fp</b> ; bezieht sich auf <b>dipl.</b>	Spannenannotation.
		Füllpartikeln und ihr Kontext. Die Annotationswerte sind immer zweistellig.
Annotationswert	Wertbeschreibung	
f	Identifizierbare Füllpartikel (keine Silbenprolongationen, keine stillen Pausen, keine Atmungspausen).	
a	Antezedens. Linkes Intervall von <i>f</i> . Stille Pausen oder das letzte phonetische Segment des vorangehenden Signals.	
p	Postzedens. Rechtes Intervall von <i>f</i> . Stille Pausen oder das erste phonetische Segment des nachfolgenden Signals.	
f {	v	Ein nicht-glottales Phänomen ohne jene in <i>fc</i> und <i>fg</i> .
	c	Ein nicht näher spezifizierter Clicklaut.
	g	Ein glottales Phänomen.
	x	Nicht in <i>fv</i> , <i>fc</i> oder <i>fg</i> kategorisierbar.
a {	s	Im Antezedens ist ein <b>Lautsegment</b> enthalten.
	c	Im Antezedens ist ein <b>Click</b> enthalten.
	p	Im Antezedens ist eine <b>stille Pause</b> enthalten.
	r	Im Antezedens ist eine <b>überwiegend stille Pause</b> enthalten, die unbestimmte artikulatorische oder vegetative Reflexe enthält.
	h	Im Antezedens ist eine <b>Atempause</b> enthalten.
	e	Im Antezedens ist <b>Extralinguistisches</b> enthalten (Lachen, Räuspern, Schlucken, Husten, etc.).
	t	Im Antezedens ist ein <b>Turn</b> des Dialogpartners enthalten.
p {	s	Im Postzedens ist ein <b>Lautsegment</b> enthalten.
	c	Im Postzedens ist ein <b>Click</b> enthalten.
	p	Im Postzedens ist eine <b>stille Pause</b> enthalten.
	r	Im Postzedens ist eine <b>überwiegend stille Pause</b> enthalten, die unbestimmte artikulatorische oder vegetative Reflexe enthält.
	h	Im Postzedens ist eine <b>Atempause</b> enthalten.
	e	Im Postzedens ist <b>Extralinguistisches</b> enthalten (Lachen, Räuspern, Schlucken, Husten, etc.).
	t	Im Postzedens ist ein <b>Turn</b> des Dialogpartners enthalten.

Die Ebene **segm** enthält die segmentale Annotation der auf der Ebene **fp** identifizierten Füllpartikeln und ihres direkten Kontextes (vgl. Tabelle 4).<sup>4</sup> Die Segmen-

<sup>3</sup>Im Falle von Pausen entspricht der direkte Kontext auch nicht-segmentalem Kontext.

<sup>4</sup>Lautsegmente werden der Einfachheit halber in SAMPA dargestellt (<http://www.phon.ucl>).

tierung orientiert sich daran, ob ein Laut gerade schon oder gerade nicht mehr als Laut einer bestimmten Klasse erkennbar ist. Grenzen werden am Nulldurchgang aufsteigender Flanken im Oszillogramm gesetzt.

Tabelle 4: Segmentale Annotation.

Annotationsart	Beschreibung
Annotationsebene	segm; bezieht sich auf fp
Annotationsart	Spannenannotation
Beschreibung	Annotation der Segmente, aus denen die Füllpartikeln bestehen sowie der Segmente, die in den Ante- und Postzedentia enthalten sind.
Annotationswert	Wertbeschreibung
SAMPA	Jeder auf der fp-Ebene mit <i>bs</i> , <i>as</i> , <i>fv</i> , <i>ps</i> , <i>qs</i> annotierte Exponent wird mit seinem tatsächlichen symbolphonetischen Wert oder Werten in weiter Transkription mithilfe von SAMPA annotiert (es gelten die Ergänzungen A–F, s. u.). Jeder Lautwert wird separat annotiert (beispielsweise [ʔɛ:] als <i>fv</i> in fp mit zwei Token ? und E). Die Segmentierung orientiert sich an der Ausdehnung der als unterscheidbar identifizierbaren Laute. Grenzen werden in Praat dort annotiert, wo der Laut gerade schon oder gerade nicht mehr als Laut einer bestimmten Klasse erkennbar ist. Grenzen werden am Nulldurchgang aufsteigender Flanken im Oszillogramm gesetzt.
?   G   GP   GT   GG	<b>Ergänzung A:</b> Ein bis drei glottale Plosive werden mit ? annotiert. Sequenzen von mehr als drei glottalen Plosiven, die enger als 50 ms beieinander stehen, werden mit G annotiert. Sequenzen von mehr als zwei glottalen Plosiven, die weiter als 50 ms entfernt auftreten (entspricht 20 Hz), werden mit GP annotiert. Glottale Transitionen antezedenter Segmente in den fv-Vokal werden mit GT annotiert. Die Abgrenzung zum Vokal in fv wird dort getroffen, wo die regelmäßige modale Phonation nach links hin endet. Glottalisierte Phasen mit geschlossenem Mund werden mit GG annotiert.
Q	<b>Ergänzung B:</b> Hohe gepresste Stimmlage vor dem Vokal in fv wird auf segm mit Q annotiert.
E	<b>Ergänzung C:</b> Vokale in fv werden immer als E annotiert. Ihre Segmentierung endet nach der letzten erkennbaren Vokalamplitude.
0	<b>Ergänzung D:</b> Mikropausen innerhalb von fv unter 100 ms werden mit 0 annotiert. Diese können auch das erste Segment unter fv auf segm sein, wenn davor keine Pause vorhanden ist.
#	<b>Ergänzung E:</b> Unvollständige Stimmlippenschwingung mit geringer Amplitude, die an den linken und rechten Enden von Vokalen auftreten kann, wird mit # annotiert.
x	<b>Ergänzung F:</b> Unentscheidbare Intervalle werden mit x annotiert.
—	Werte auf Ebene fp mit <i>ac</i> , <i>pc</i> , <i>ap</i> , <i>pp</i> , <i>ar</i> , <i>pr</i> , <i>ae</i> , <i>pe</i> , <i>at</i> , <i>pt</i> werden mit einem Unterstrich annotiert.
in	Atempausen ( <i>ah</i> , <i>ph</i> ), in denen eingeatmet wird.
ex	Atempausen ( <i>ah</i> , <i>ph</i> ), in denen ausgeatmet wird.

[ac.uk/home/sampa/index.html](http://ac.uk/home/sampa/index.html)), besucht am 15.07.2019.

Auf der Ebene **phon** (vgl. Tabelle 5) kann die Phonationsart der in Füllpartikeln enthaltenen Vokale annotiert werden.

**Tabelle 5:** Phonationsartannotation.

Annotationsart	Wertbeschreibung
<b>Annotationsebene</b>	<b>phon</b> ; bezieht sich auf <b>segm</b>
<b>Annotationsart</b>	Tokenannotation
<b>Beschreibung</b>	Annotation der Phonationsart.
Annotationswert	Wertbeschreibung
glo	Bezogen auf <b>segm</b> für den mit SAMPA annotierten Vokal innerhalb des Intervalls <i>fv</i> von Ebene <b>fp</b> , wenn dieser Vokal <b>vollständig</b> glottalisiert ist.
mod	Bezogen auf <b>segm</b> für den mit SAMPA annotierten Vokal innerhalb des Intervalls <i>fv</i> von Ebene <b>fp</b> , wenn dieser Vokal <b>vollständig</b> mit modaler Phonation artikuliert ist.
asp	Bezogen auf <b>segm</b> für den mit SAMPA annotierten Vokal innerhalb des Intervalls <i>fv</i> von Ebene <b>fp</b> , wenn dieser Vokal <b>vollständig</b> mit behauchter Phonation artikuliert ist.

## 4 Beispiele

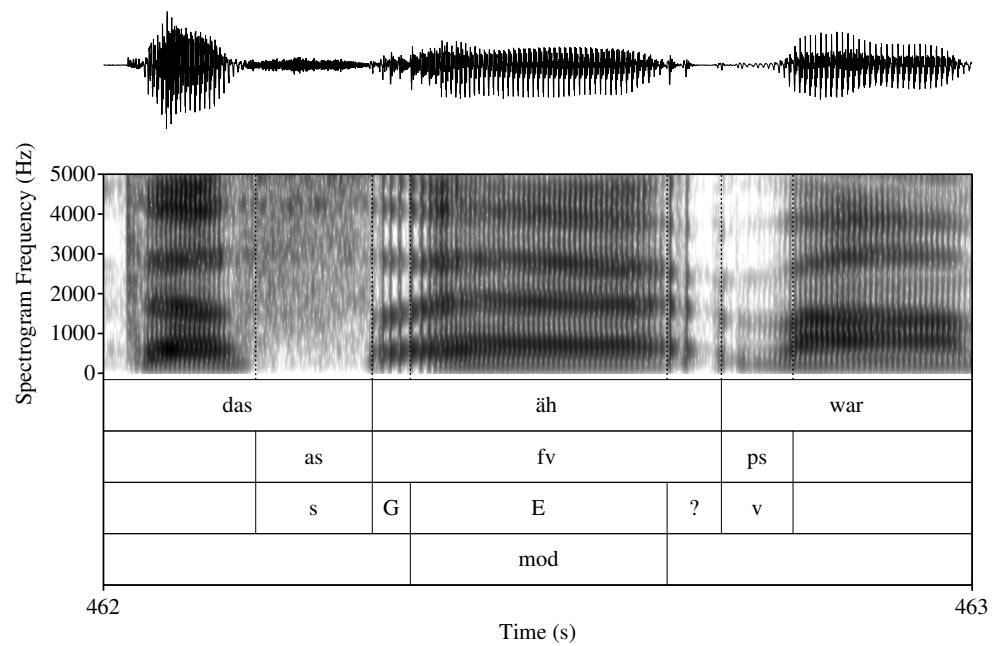
Drei Beispiele werden vorgestellt.<sup>5</sup> Abbildung 1 enthält die Annotation einer nicht-glottalen Füllpartikel mit vorangehender Einatmung. Sie enthält einen glottalen Plosiv und ist direkt an das nächste Segment (*j* in *ja*) angeschlossen.

Abbildung 2 zeigt eine etwas komplexere Annotation einer nicht-glottalen Füllpartikel, die zu beiden Seiten ohne Pause in Lautsegmente eingebettet ist. Sie enthält zu Beginn Sequenzen glottaler Plosive und an ihrem Ende ein bis drei glottale Plosive.

Abbildung 3 beschreibt eine glottale Füllpartikel. Da hierfür keine orthographische Repräsentation existiert, wird auf der Transliterationsebene **dipl** eine Beschreibung eingefügt ((klicken)). Ihr linker Kontext ist zugleich der rechte Kontext einer vorangehenden Füllpartikel und erhält daher den Kombinationswert *pr\_ar*.

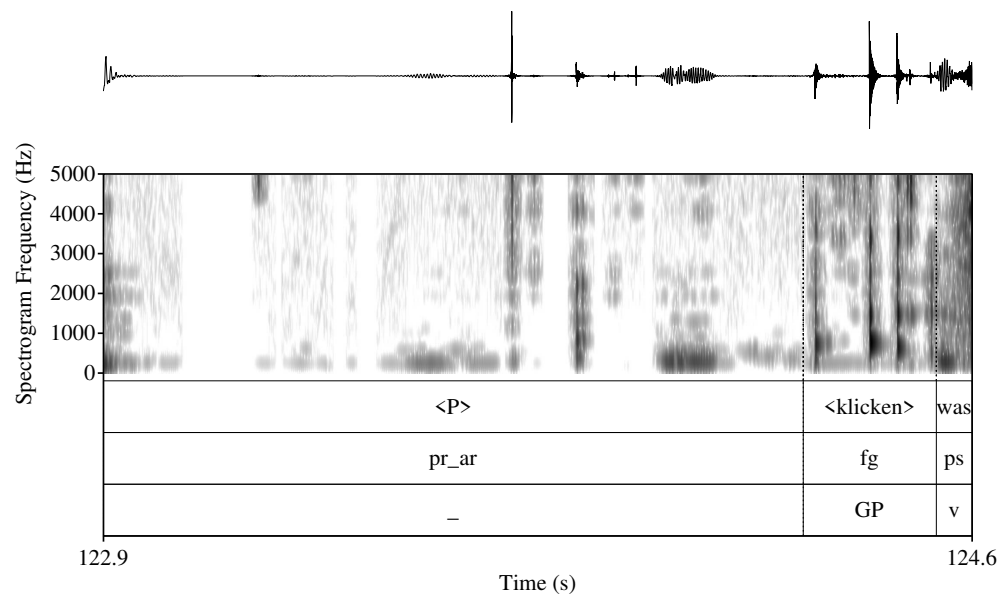
<sup>5</sup>Die gezeigten Daten sind am Lehrstuhl für Sprachwissenschaft des Deutschen: Phonetik/Phonologie unter der Leitung von Prof. Dr. Christine Mooshammer und mir aufgenommen worden. Sie enthalten aufgabenfreie spontane Dialoge und werden perspektivisch als Korpus gesprochener Sprache veröffentlicht.





**Abbildung 2:** Nicht-glottale Füllpartikel, annotiert auf *dipl* mit *ähm*, auf *fp* mit *as* (linker Kontext, Segment), *fv* und *ps* (rechter Kontext, Segment); auf *segm* mit *s* (SAMPA), *G* (mehr als drei glottale Plosive), *E* (Vokal), *?* (ein bis drei glottale Plosive), *v* (Laut nach SAMPA); auf *phon* mit *mod* (modale Phonation).





**Abbildung 3:** Glottale Füllpartikel, annotiert auf `dip1` mit `<klicken>`, auf `fp` mit `pr_ar` (zugleich rechter Kontext einer vorangehenden Füllpartikel und linker Kontext der glottalen Füllpartikel, Pause enthält artikulatorische Reflexe) und `ps` (rechter Kontext, Segment); auf `segm` mit `,_'`; `GP` (mehr als zwei glottale Plosive im Abstand von 50 ms) und `v` (Laut nach SAMPA).

## 5 Auswertung

Das Schema kann mit verschiedenen Tools ausgewertet werden, die Mehr-Ebenen-Annotationen unterstützen. Eine Auswertung der Häufigkeiten von und der Beziehungen zwischen den annotierten Werten kann beispielsweise mit ANNIS erfolgen (Krause und Zeldes 2016). Für die Auswertung der phonetischen Parameter wie der Formanten in dem mit *E* markierten Vokal auf der Ebene *segm* bietet sich eine Konvertierung der annotierten Daten in eine EMU-Datenbank (Winkelmann u. a. 2017) und ihre Analyse in R (R Core Team 2018) mit dem Paket *emuR* (Winkelmann u. a. 2018) an.

## Literatur

- Batliner, Anton, Andreas Kießling, Susanne Burger und Elmar Nöth (1995). *Filled pauses in spontaneous speech*. URL: [https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25232/1/report\\_88\\_95.pdf](https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25232/1/report_88_95.pdf) (besucht am 15.07.2019).
- Belz, Malte (2017). „Glottal filled pauses in German“. In: *Proceedings of DiSS 2017. The 8th Workshop on Disfluency in Spontaneous Speech* (Stockholm, Sweden). Hrsg. von Robert Eklund und Ralph L. Rose. TMH-QPSR 58(1), S. 5–8.
- Boersma, Paul (2001). „Praat, a system for doing phonetics by computer“. In: *Glott International* 5 (9), S. 341–345.
- Krause, Thomas und Amir Zeldes (2016). „ANNIS3: A new architecture for generic corpus query and visualization“. In: *Digital Scholarship in the Humanities* 31 (1), S. 118–139. DOI: 10.1093/llc/fqu057.
- Leeuw, Esther de (2007). „Hesitation Markers in English, German, and Dutch“. In: *Journal of Germanic Linguistics* 19 (2), S. 85–114. DOI: 10.1017/S147054270700049.
- R Core Team (2018). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Rasoloson, Janie Noëlle (1994). *Interjektionen im Kontrast. Am Beispiel der deutschen, madagassischen, englischen und französischen Sprache*. Bd. 22. Arbeiten zur Sprachanalyse. Frankfurt am Main: Peter Lang.
- Schmidt, Thomas und Kai Wörner (2009). „EXMARaLDA - Creating, analysing and sharing spoken language corpora for pragmatic research“. In: *Pragmatics* 19 (4), S. 565–582.
- Schwitalla, Johannes (2002). „Kleine Wörter. Partikeln im Gespräch“. In: *Über Wörter. Grundkurs Linguistik*. Hrsg. von Jürgen Dittmann und Claudia Schmidt. Freiburg im Breisgau: Rombach, S. 259–283.

- Trouvain, Jürgen und Zofia Malisz (2016). „Inter-Speech Clicks in an Interspeech Keynote“. In: *Interspeech* (San Francisco), S. 1397–1401. DOI: 10.21437/Interspeech.2016-1064.
- Willkop, Eva-Maria (1988). *Gliederungspartikeln im Dialog*. Bd. 5. Studien Deutsch. München: Iudicium.
- Winkelmann, Raphael, Jonathan Harrington und Klaus Jänsch (2017). „EMU-SDMS. Advanced speech database management and analysis in R“. In: *Computer Speech & Language*. DOI: 10.1016/j.cs1.2017.01.002.
- Winkelmann, Raphael, Klaus Jaensch, Steve Cassidy und Jonathan Harrington (2018). *emuR. Main Package of the EMU Speech Database Management System*. R package version 1.1.1.