

Malte Belz

Die Phonetik von *äh* und *ähm*

Akustische Variation von Füllpartikeln
im Deutschen

MOREMEDIA



OPEN ACCESS



J.B. METZLER

Die Phonetik von *äh* und *ähm*

Malte Belz

Die Phonetik von *äh* und *ähm*

Akustische Variation von Füllpartikeln
im Deutschen



J.B. METZLER

Malte Belz
Institut für deutsche Sprache und Linguistik
Humboldt-Universität zu Berlin
Berlin, Deutschland

Die Veröffentlichung wurde gefördert aus dem Open-Access-Publikationsfonds der Humboldt-Universität zu Berlin.

Die Arbeit wurde von der Humboldt-Universität zu Berlin als Dissertation angenommen und 2020 verteidigt.



ISBN 978-3-662-62811-9 ISBN 978-3-662-62812-6 (eBook)
<https://doi.org/10.1007/978-3-662-62812-6>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en) 2021 Dieses Buch ist eine Open-Access-Publikation. **Open Access** Dieses Buch wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Buch enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Anna Pietras

J.B. Metzler ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Inhaltsverzeichnis

Danksagung	vi
Abkürzungsverzeichnis	vii
1 Einführung	1
1.1 Forschungsfrage	1
1.2 Definition	3
1.3 Onomastikon	6
2 Phonetische Variabilität	11
2.1 Kategoriale Beschreibung	11
2.1.1 Orthographische Repräsentation	12
2.1.2 Symbolphonetische Repräsentation	16
2.2 Akustische Beschreibung	21
2.2.1 Segmente	21
2.2.2 Dauer	22
2.2.3 Grundfrequenz	24
2.2.4 Formanten	27
2.2.5 Phonationsart	30
2.2.6 Intensität	33
2.3 Fazit und Forschungsdesiderat	33
3 Funktionen und Kontexte	37
3.1 Multifunktionalität von Füllpartikeln	37
3.1.1 Symptom versus Signal	37
3.1.2 Funktionen	40
3.2 Kontext und phonetische Realisierung	48
3.2.1 Die Form-Kontext-Hypothese	48

3.2.2	Sequenzielle Kontexte	53
3.2.2.1	Segmente	53
3.2.2.2	Wörter	56
3.2.3	Hierarchische Kontexte	57
3.2.3.1	Syntaktische Phrasen und Äußerungen	57
3.2.3.2	Intonationsphrasen	59
3.2.3.3	Dialogstruktur	62
3.3	Paralinguistische Variation	67
3.3.1	Sprecherspezifität	67
3.3.2	Situationspezifität	69
3.4	Fazit und Forschungsdesiderat	70
4	Methode	73
4.1	Korpora	74
4.1.1	Studie I: GECO v.1.1	74
4.1.2	Studie II: BeDiaCo v.1	74
4.1.3	Diplomatische Transliteration und Segmentierung	76
4.2	Annotation	77
4.2.1	Füllpartikelannotation	77
4.2.2	Intonationsphrasenannotation	81
4.2.3	Dialogstrukturannotation	90
4.2.3.1	Segmentierung	90
4.2.3.2	Annotationswerte	92
4.3	Qualität der Annotationsrichtlinien	99
4.3.1	Füllpartikeln	99
4.3.2	Intonationsphrasen	100
4.3.3	Dialogzüge	101
4.4	Abfrage und phonetische Analyse	101
4.4.1	Dauer und Sprechgeschwindigkeit	102
4.4.2	Grundfrequenz	102
4.4.3	Formanten	102
4.4.3.1	Referenzvokale in GECO-FP	104
4.4.3.2	Referenzvokale in BeDiaCo	104
4.4.4	Phonationsart und Intensität	104

4.5	Statistik	108
4.6	Zugang und Wiederverwendung	110
4.6.1	GECO v.1.1	111
4.6.2	BeDiaCo v.1	111
5	Ergebnisse der explorativen Studie	113
5.1	Füllpartikeln	113
5.1.1	Allgemeine und sprecherspezifische Häufigkeit	113
5.1.2	Segmentale Formen	115
5.1.2.1	Nicht-glottale Formen	115
5.1.2.2	Glottale Formen	117
5.1.3	Dauer	118
5.1.3.1	Nicht-glottale Formen	118
5.1.3.2	Glottale Formen	121
5.1.4	Grundfrequenz	122
5.1.5	Formanten	127
5.1.6	Phonationsart	134
5.2	Form und Kontext	137
5.2.1	Sequenzielle Kontexte	137
5.2.1.1	Dauer	140
5.2.1.2	Grundfrequenz	143
5.2.1.3	Formanten	147
5.2.1.4	Phonationsart	150
5.2.2	Dialogzüge	151
5.2.2.1	Dauer	155
5.2.2.2	Grundfrequenz	156
5.2.2.3	Formanten	158
5.2.2.4	Phonationsart	163
5.3	Zusammenfassung	164
6	Ergebnisse der konfirmatorischen Studie	168
6.1	Füllpartikeln	168
6.1.1	Allgemeine und sprecherspezifische Häufigkeit	168

6.1.2	Segmentale Formen	170
6.1.2.1	Nicht-glottale Formen	170
6.1.2.2	Glottale Formen	173
6.1.3	Dauer	174
6.1.3.1	Nicht-glottale Formen	174
6.1.3.2	Glottale Formen	178
6.1.4	Grundfrequenz	179
6.1.5	Formanten	181
6.1.6	Phonationsart	190
6.2	Form und Kontext	192
6.2.1	Sequenzielle Kontexte	192
6.2.1.1	Dauer	195
6.2.1.2	Grundfrequenz	199
6.2.1.3	Formanten	203
6.2.1.4	Phonationsart	208
6.2.2	Dialogzüge	209
6.2.2.1	Dauer	213
6.2.2.2	Grundfrequenz	215
6.2.2.3	Formanten	216
6.2.2.4	Phonationsart	221
6.2.3	Intonationsphrasen	222
6.2.3.1	Dauer	225
6.2.3.2	Grundfrequenz	228
6.2.3.3	Formanten	230
6.2.3.4	Phonationsart	234
7	Zusammenfassung und Vergleich	236
7.1	Hypothese der Formvariabilität	236
7.2	Form-Kontext-Hypothese	238
8	Diskussion	243
8.1	Kontextfreie Formvariation	243
8.2	Kontextbedingte Formvariation	247
8.2.1	Sequenzieller Kontext und Dialogstruktur	247

8.2.2 Einfluss prosodischer Phrasierung	249
8.3 Fazit und Ausblick	252
Literatur	254

Danksagung

Die vorliegende Dissertation entstand an der Sprach- und literaturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin und beschäftigt sich aus einer phonetisch-korpuslinguistischen Perspektive mit *äh* und *ähm*. Diese Perspektive wurde maßgeblich von meinen beiden Betreuerinnen ermöglicht. In diesem Sinne bedanke ich mich bei meiner Erstbetreuerin Prof. Dr. Christine Mooshammer herzlich für die freundschaftliche Betreuung, die phonetischen Beratungen und die kritischen Gespräche über Füllpartikeln, aus denen ich immer neue Überlegungen für diese Arbeit mitnehmen konnte. Meiner Zweitbetreuerin Prof. Dr. Anke Lüdeling danke ich für die korpuslinguistische Betreuung und Expertise sowie für den Impetus, in meiner Masterarbeit Disfluencies zu untersuchen – ein Thema, das sich als nachhaltig interessant herausstellte.

Dem Institut für deutsche Sprache und Linguistik danke ich für die Übernahme der Versuchspersonenhonorare für die Aufnahmen des *Berlin Dialogue Corpus*. Meiner Promotionskommission danke ich für die Bereitschaft zur analogen Durchführung der Disputation im Mai 2020 während pandemischer Zeiten. Bei Dr. Jürgen Trouvain, Dr. Jelena Krivokapić und den Kolleg*innen am Institut für deutsche Sprache und Linguistik bedanke ich mich für die fachlichen Diskussionen über Phonetik und Linguistik; bei Dr. Felix Golcher für die Beratung über statistische Feinheiten; bei Dr. Raphael Winkelmann für die schnellen EMU-Hilfestellungen; bei Alina Zöllner und Megumi Terada für die Unterstützung bei der Erstellung von BeDiaCo; und bei Dr. Thomas Landgraf und Edith Scheifele für die bereichernden Gespräche über Wissen und Wissenschaft.

Meinen Eltern danke ich herzlich für ihre Unterstützung. Mein größter Dank gilt Dr. Carolin Odebrecht, die mir während der allmählichen Fertigstellung dieser Arbeit und über alle Höhen und Tiefen hinweg mit wissenschaftlicher Kritik, fachlicher Beratung und emotionaler Unterstützung zur Seite stand. Ihr widme ich dieses Buch.

Abkürzungsverzeichnis

⟨P⟩	Stille Pause
A ₃	Amplitude der dritten Harmonischen
BeDiaCo v. 1	Berlin Dialogue Corpus v. 1
BeDiaCo _r v. 1	Berlin Dialogue Corpus restricted v. 1
BeDiaCo _c	Berlin Dialogue Corpus compiled v. 1
f ₀	Grundfrequenz
F ₁ /F ₂ /F ₃	Erster/zweiter/dritter Formant
FG	Glottale Füllpartikel(n)
FP	Füllpartikel(n)
FV	Nicht-glottale Füllpartikel(n)
GECO v. 1.1	GERman COversations v. 1.1
GECO-FP	GERman COversations-Füllpartikeln v. 1
H ₁ /H ₂ /H ₃	Erste/zweite/dritte Harmonische
IP	Intonationsphrase
SAMPA	Speech Assessment Methods Phonetic Alphabet
V	Vokalische Füllpartikelform
VN	Vokalisch-nasale Füllpartikelform
VP	Versuchsperson



1 Einführung

1.1 Forschungsfrage

Diese Arbeit zeigt, dass Füllpartikeln wie *äh* und *ähm* verschiedene akustische Realisierungen annehmen, die nicht immer arbiträr sind (kontextfreie Variation), sondern in bestimmten Fällen durch ihren Auftretenskontext (kontextbedingte Variation) erklärt werden können. Zuvor werden die im Deutschen vorhandenen Füllpartikelformen beobachtet und quantifiziert. Damit füllt die Arbeit eine Lücke in der Beschreibung und Erklärung von Füllpartikeln im Deutschen – bis heute gibt es keine genaue Vorstellung, welches phonetische Material als Füllpartikel zu bewerten ist und in welchem Ausmaß diese Formen durch ihren Kontext bedingt werden. Stattdessen wird häufig untersucht, welche linguistischen Funktionen Füllpartikeln erfüllen. In Beispiel (1) kann die Füllpartikel *äh* zum Beispiel in mindestens zwei Lesarten so verstanden werden, dass sie das intensive Nachdenken der Sprecherin B anzeigt, oder aber der Hörerin A signalisiert, dass sie weitersprechen möchte – oder beides.

- (1) a. A: okay
 b. B: also ne kleinere Stadt is das ⟨P, 59 ms⟩ **äh** ⟨Einatmung, 280 ms⟩ das kennt auch niemand (BeDiaCoc frei_flf2_ch2, 263,5–267,3 s)

Für diese *funktionale* Problematik wird hier allerdings keine Lösung präsentiert, da dies den zweiten Schritt vor dem ersten setzte, nämlich die *formale* Problematik zu verstehen. Dieses Vorgehen (die funktionale Seite von Füllpartikeln unterspezifiziert zu lassen) ist insofern vorteilhaft, als es überhaupt erst ermöglicht, ihre *formale* Seite – hier die akustische Form – unvoreingenommen zu untersuchen. Demgemäß wird die Arbeit motiviert durch die Frage, ob denn wenigstens für die Formseite von Füllpartikeln Systematiken zu beobachten sind. Die Ergebnisse der Arbeit bestärken diese Annahme: Die akustische Form von Füllpartikeln ist in bestimmten

Kontexten systematisch vorhersagbar; es lassen sich Form-Kontext-Muster identifizieren. Im obigen Beispiel (1) wirkt sich die Position von *äh* zwischen zwei Pausen und am Ende einer Intonationsphrase (*also ne kleinere Stadt is das äh*) auf ihre akustische Form aus. *äh* hat in dieser Position tendenziell eine längere Dauer, als wenn es direkt ohne Pausen an den Kontext angeschlossen würde oder am Anfang der Intonationsphrase stünde, und seine Tonhöhe ist tendenziell tiefer als zu Beginn einer Intonationsphrase.

Exakter formuliert lautet die Forschungsfrage dieser Arbeit also: Gibt es eine Form-Kontext-Verbindung dergestalt, dass bestimmte phonetische Füllpartikelformen systematisch in einem spezifischen linguistischen Auftretenskontext (betrachtet werden Pausen und Segmente, Dialogstruktur und Intonationsphrasen) verwendet werden? Ziel dieser Arbeit sind (i) die Beschreibung phonetischer Formen nicht-lexikalischer Füllpartikeln in deutscher Spontansprache und (ii) die Betrachtung ihres direkten und indirekten linguistischen Auftretenskontextes als mögliche Erklärung ihrer Form. Diese Hypothesen werden aus der Literatur heraus motiviert und explorativ sowie konfirmatorisch getestet. In spontaner Sprache kann der Auftretenskontext einer Füllpartikel allerdings nicht streng kontrolliert werden. Methodisch arbeite ich daher korpusbasiert und quantitativ-qualitativ. Zusätzlich erfolgt die Untersuchung anhand von *aufgabenfreien* spontansprachlichen Dialogen. Somit wird vermieden, dass sich bestimmte Häufungen von Frage-Antwort-Abfolgen, wie sie in aufgabenbasierten Szenarien (z. B. dem Lösen eines Bilderrätsels) erwartbar sind, verzerrend auf die Häufigkeiten der gefundenen Form-Kontext-Muster auswirkt. Die Frequenzen von Füllpartikeln oder einiger bestimmter Formen sind alleine betrachtet nur bedingt aussagekräftig. Sie werden daher immer zusammen mit ihrer Distribution diskutiert. Dank der akustischen Analyse können dann Muster ihrer Verwendung in bestimmten Kontexten erkannt und erklärt werden.

(i) Die Beschreibung phonetischer Formen nicht-lexikalischer Füllpartikeln im Deutschen ist ein Forschungsdesiderat, welches trotz oder gerade aufgrund der intensiven Forschung zu Füllpartikeln in den letzten zwei Dekaden entstanden ist. Eine Auswahl der in der Literatur beschriebenen Formenvielfalt sind *ah*, *äh*, *ähm*, *öh*, *öhm*, *hm*, *mh* für schriftbasierte Formen (vgl. Abschnitt 2.1.1) und [ə əm v: e: ʔɛ: ʔɛ:m ʔœ: ʔœ:m m] für symbolphonetische, auditiv perzipierte Formen (vgl. Abschnitt 2.1.2). Neben diesen qualitativen und subjektiv beeinflussbaren Beschreibungen

gen der Form ermöglicht die messphonetische Untersuchung des akustischen Signals eine objektive Sicht auf die Formvariationen. Hierüber liegen für das Deutsche bislang noch kaum Erkenntnisse vor. Zur Quantifizierung der akustischen Formen werden Parameter der akustischen Phonetik wie Grundfrequenz, Formantwerte, Dauer und Phonationsart herangezogen (vgl. Abschnitt 2.2).

(ii) Der direkte und indirekte linguistische Auftretenskontext von Füllpartikeln wird zwar in einigen Studien – besonders für das Englische – thematisiert, ist aber meist nicht der Kern der Untersuchung. Dennoch deuten Indizien darauf hin, dass bestimmte Formen in bestimmten Kontexten überzufällig vorkommende Auftretensmuster zeigen (vgl. Kapitel 3). Die Beschreibung der Formen beschränkt sich jedoch überwiegend auf orthographisch repräsentierte Formen in der Dichotomie *äh* vs. *ähm*. Der Kontext einer Füllpartikel kann auf unterschiedlichen hierarchischen Ebenen betrachtet werden, von der segmentalen Ebene über die Wortebene und Ebene der syntaktischen Phrasengrenzen bis hin zu prosodischen Phrasen und Äußerungseinheiten. Eine quantitative akustische Analyse der phonetischen Form in Kombination mit einer solchen Kontextanalyse stellt einen vielversprechenden Schritt zur Erklärung ihrer Formvariabilität dar. In dieser Arbeit untersuche ich, ob der Auftretenskontext einer Füllpartikel ihre phonetische Realisierungsform erklären kann. Drei verschiedene Kontextstrukturen unterschiedlicher Granularität werden dabei betrachtet: der adjazente sequenzielle Kontext der Füllpartikel (Laute und Pausen), der Kontext der dyadischen Interaktion im Gespräch (Dialogzüge) und schließlich als mittlere Einheitengröße auf prosodischer Ebene der Intonationsphrasenkontext.

1.2 Definition

Mit dem Begriff der *Füllpartikel* beschreibe ich nicht-lexikalische Entitäten, die orthographisch typischerweise mit *äh*, *ähm*, *hm* repräsentiert, jedoch phonetisch nicht ausschließlich in dieser Form produziert werden, sondern variieren können. Als Füllpartikeln kategorisiert werden auch Entitäten, deren Vokalqualität variiert (z. B. [œ:]) oder die nur aus Konsonanten [pf] oder glottalen Plosivketten bestehen [ʔʔʔ].¹ Keine Füllpartikeln in diesem Sinne sind stille Pausen, Atmungspausen, sowie para-

¹Ob auch Formen wie Clicks (Gil 2013; Trouvain 2015; Trouvain & Malisz 2016) darunter subsumiert werden, ist ein offene Frage (vgl. Günthner 2017; Smith & Clark 1993). Für diese Arbeit schließe ich Click-Formen vorerst aus der Definition aus.

oder extralinguistische Geräusche wie Lachen, Husten, Räuspern, oder Pfeifen. Abbildung 1.1 stellt die in dieser Arbeit entwickelte und verwendete Formenhierarchie dar.

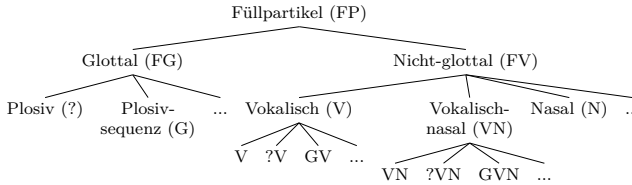


Abbildung 1.1: Füllpartikelformenhierarchie in dieser Arbeit.

Füllpartikeln (FP) gliedern sich in glottale (FG) und nicht-glottale (FV) Formen. Glottale Füllpartikeln bestehen in Abgrenzung zu nicht-glottalen Füllpartikeln *nur* aus glottalen Lautsegmenten und beinhalten keine perzeptiv wahrnehmbare vokalische Struktur. Nicht-glottale Füllpartikeln hingegen beinhalten hauptsächlich Lautsegmente, die im Vokaltrakt artikuliert werden. Stimmhafte Konsonanten und Vokale werden demnach trotz ihrer Glottisaktivität zu FV gezählt. Glottale Füllpartikeln zerfallen in mindestens zwei Gruppen, nämlich in einen singulären glottalen Plosiv [ʔ] (dargestellt mit dem Symbol ,?‘) und Sequenzen glottaler Plosive (G). Nicht-glottale Füllpartikeln zerfallen in drei große Gruppen, nämlich vokalische (V), vokalischnasale (VN) und nasale (N) Füllpartikeln. Die Darstellung dieser Gruppierungen schließt nicht aus, dass noch andere Formen vorkommen können, wie beispielsweise *ähmä*, *ähf*, oder [f:], was durch die Auslassungspunkte (...) symbolisiert wird. Jede der nicht-glottalen Gruppierungen ist eine quasi-phonologische Abstraktion konkreterer Realisierungen, wie V, ?V, oder GV, in denen vor Vokalen in der V-Formgruppe zusätzlich Variation durch die Voranstellung eines glottalen Plosivs oder von Sequenzen glottaler Plosive möglich ist. In dieser Arbeit ist mit den Abkürzungen V und VN also (sofern nicht anders spezifiziert) die vokalische und vokalischnasale Formen ohne genauere Betrachtung ihrer phonetischen Realisierung gemeint.

Die Verwendung des Begriffs der *Füllpartikel*, der ein terminologisches Novum außerhalb der im Deutschen und Englischen geläufigen Termini *gefüllte Pause*, *Füller* und *Häsitationen* sowie *filled pause*, *filler* und *hesitation* bildet (vgl. auch Abschnitt 1.3), wird wie folgt motiviert.

Erstens entsteht während der Produktion von Füllpartikeln keine Pause – weder artikulatorisch noch akustisch (vgl. auch den Abschnitt zur *gefüllten Pause* auf Seite 6).² In Abbildung 1.2 ist die deutliche Anwesenheit eines Signals für die Füllpartikel *ähm* zu beobachten, während die stille Pause (P) rechts von ihr tatsächlich kein akustisches Korrelat zeigt.³ Das akustische Signal von *ähm* birgt die gleichen Schwierigkeiten der Identifizierung, Abgrenzung und Segmentierung lautlicher Kontinua, die für jedes andere akustische Signal auch gelten.

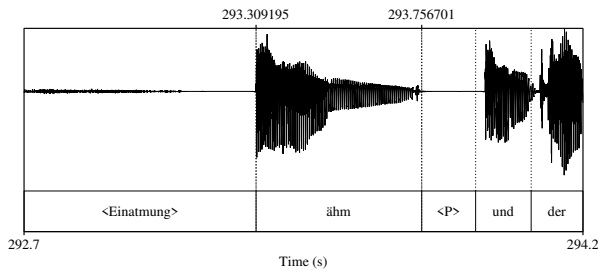


Abbildung 1.2: Akustisches Signal (Oszillogramm) für eine Atmungspause, gefolgt von einer Füllpartikel *ähm* und einer stillen Pause (BeDiaCo_c: frei_f3f4_ch1).

Zweitens können Füllpartikeln verschiedene sprachliche Funktionen erfüllen (vgl. Abschnitt 3.1.2).⁴ Die bisherigen Termini, die dies abzubilden suchen, tradieren aus der linguistischen Pragmatik: *Verzögerungsmarker*, *Interjektion*, *Gliederungspartikel*, *Diskursmarker*, *Diskurspartikel*. Die Köpfe *-marker* und *-partikel* ordnen das Phänomen einer grammatischen Beschreibungsebene zu. Allerdings impliziert *-marker*, dass etwas markiert wird, ohne dass aufgrund ihrer Unterspezifizierung hervorgeht, was markiert wird. In diesem Sinne ist *-partikel* neutraler. Jedoch suggerieren die Nichtköpfe *Diskurs-* und *Gliederung-* eine bestimmte Situationspezifität oder Funktion, welche Füllpartikeln in die große Gruppe von (auch lexikalischen) *Diskurs-* und *Gliederungspartikeln* stellt.

²Was nebenbei die Oxymorondebatte auflöst (vgl. Abschnitt 1.3).

³Im artikulatorischen Signal können jedoch durchaus zungenartikulatorische Bewegungen während der akustischen Pause vorhanden sein (vgl. Ramanarayanan et al. 2009; Rasskazova et al. 2018). Eine akustische Pause bedeutet also nicht automatisch auch eine artikulatorische Pause.

⁴Die Funktionen von Füllpartikeln sind divers; sie können Verzögerungen, Turns, Reparaturen, Informationsstatus, Salienz, Aufmerksamkeit oder Dialogstruktur markieren, sie sind jedoch keine Feedbackseinheiten wie *mhm*.

Hier entsteht meines Erachtens die Notwendigkeit für einen phänomenalen Begriff, der sowohl die grammatische Dimension⁵ einerseits als auch die Variabilität der funktionalen und formalen Realisierungen andererseits abbilden kann, ohne dabei funktional vorzugreifen.

Der morphologische Kopf des in dieser Arbeit verwendeten Begriffs, *-partikel*, ordnet Füllpartikeln den Partikeln zu, die als nicht-flektierbar und lexikalisch oft unterspezifiziert definiert sind. Syntaktisch sind Füllpartikeln sehr wahrscheinlich wesentlich freier in ihrer Stellung als andere Partikeln, wenngleich sie Präferenzen für bestimmte Positionen zeigen können.⁶ Mit dem Nichtkopf *Füll-* wird die vorhandene Aktivität im Sprachsignal neutral beschrieben, ohne sie *a priori* funktional zu spezifizieren. Nebenbei hat der Begriff der *Füllpartikel* den Vorteil, dass er Teile der bisherigen Terminologien in sich aufnimmt und gleichzeitig die Integration von Füllpartikeln in grammatische Beschreibungen ermöglicht. Im Übrigen lässt sich *Füllpartikel* (und *filler particle* als englische Übersetzung) mit der schon bekannten Abbeviation FP abkürzen.

1.3 Onomastikon

Dieses Kapitel dient der weiteren Motivation des Begriffs der *Füllpartikel*, indem es in gebotener Kürze einen Einblick in die bisher verwendeten Termini gibt und zeigt, dass diese nicht deutlich abgegrenzt sind. Zur Vermeidung von Denotationsüberschneidungen vermeiden manche Autor*innen die Verwendung von Hyperonymen übrigens gänzlich; so spricht Fox Tree (2001) immer nur von „*um* and *uh*“. Dieses eher deskriptive Vorgehen hat den Vorteil, nicht *a priori* eine phonetische, phonologische oder graphematische Gestalt der untersuchten Formen und deren funktionalen und theoretischen Status festlegen zu müssen.⁷

Der Terminus *gefüllte Pause* (engl. ‚filled pause‘) wird häufig im Zusammenhang mit Sprachplanungsschwierigkeiten verwendet. Darunter fallen vokalische Formen

⁵Spätestens seit ihrer formalen semantischen Beschreibung durch Ginzburg et al. (2014) können Disfluencies (insbesondere Selbstreparaturen, aber auch Füllpartikeln) als von ihrer postulierten Position an der ‚grammatischen Peripherie‘ rehabilitiert gelten.

⁶Beispielsweise vor zu-Infinitiven (Bada & Genç 2008; vgl. auch Abschnitt 3.2.1). Die Kategorisierung als Partikeln eröffnet somit auch die Möglichkeit einer syntaktischen Beschreibung (auf Syntax wird in dieser Arbeit jedoch nicht eingegangen).

⁷Für einen ausführlichen Überblick über die Formseite vgl. Kapitel 2.

(z. B. Boomer & Dittmann 1962; Christenfeld 1995; Lallgee & Cook 1969; Leeuw 2007; Maclay & Osgood 1959), vokalisch-nasale Formen (z. B. Christenfeld 1995; Lallgee & Cook 1969; Leeuw 2007; Schegloff et al. 1977), nasale Formen (z. B. Leeuw 2007; Maclay & Osgood 1959) und manchmal auch Clicks (Smith & Clark 1993: 27). Manche Autoren fassen diese Definition etwas weiter und nehmen „non-language sounds, as well as lengthened syllables [...], non-identifiable sound combinations [...], filler words [...] and phrases [...]“ (Faure 1980: 287) hinzu. Für das Deutsche definiert Jäger (1979) gefüllte Pausen als „Laute, durch die sich der Sprecher selbst unterbricht und die keinen Wortsinn bzw. ein beginnendes Wort erkennen lassen“ (ebd.: 13). Jedoch subsumiert er zudem noch „inkohärente Geräusche wie Husten, Räuspern“ hinzu (ebd.). Gefüllte Pause wird häufig als Schlagwort für Arbeiten über Füllpartikeln verwendet. Insbesondere zusammen mit seinem Gegenstück, der *ungefüllten Pause*, beschreiben die beiden Begriffspaare *prima facie* einen schlüssigen Zusammenhang. Beide Termini propagieren eine Pause, nämlich mit und ohne sprachlichem Material.⁸ Bisweilen wird jedoch kritisiert, dass der Begriff gefüllte Pause ein Oxymoron sei (Ehlich 1986: 219; Smith & Clark 1993: 27), da eine Pause, welche ja *per definitionem* die Abwesenheit einer Tätigkeit kennzeichnet, mit einer Tätigkeit, nämlich der Produktion akustischer Signale, gefüllt wird. Somit sei die ‚Pause‘ aber keine Pause mehr. Was hierbei verkannt wird, sind unterschiedliche Ebenen, auf denen etwas passieren oder abwesend sein kann. Als Pause wird hier eben nicht die Abwesenheit eines artikulatorischen oder akustischen Signals, sondern eines lexikalisch interpretierbaren Signals verstanden. Die Kritik von Ehlich und Smith & Clark bezieht sich auf letztere Interpretation.⁹ Nähme man jedoch einen lexikalischen Gehalt von Füllpartikeln an (wie bspw. Clark & Fox Tree 2002), so ließe sich auch auf der lexikalisch interpretierbaren Ebene nicht mehr von einer Pause sprechen – der Begriff der gefüllten Pause wäre Makulatur und es entstünde eine terminologische Lücke. Möglicherweise besteht ein Zusammenhang zwischen dieser Lücke und der Etablierung des Begriffs *Füller*. Ungeachtet dieser theoretischen Erwägungen wird manchmal sogar explizit erwähnt, dass der Begriff nicht

⁸Die terminologische Nähe zwischen stiller und gefüllter Pause führt dazu, dass in manchen Studien stille und gefüllte Pausen nicht getrennt ausgewertet oder als funktional äquivalent interpretiert werden (z. B. Boomer & Dittmann 1962; Hawkins 1971).

⁹Die vermutete Abwesenheit eines lexikalisch interpretierbaren Signals lässt sich gut mit dem Verständnis und der Bezeichnung von Füllpartikeln als Häsitationsmarker beziehungsweise *Häsitation* (vgl. z. B. Faure 1980: 287; Rose 2015) vereinbaren (s. Seite 9).

beschreibungsadäquat ist: „The present results contradict the common idea that filled pauses are rest times.“ (Duez 2001: 44).

Der Terminus *Füller* (engl. ‚filler‘) etabliert sich seit den 1990er-Jahren in der einschlägigen Forschung und wird synonym zu *gefüllte Pause* verwendet, macht jedoch andere Annahmen über den lexikalischen Status von Füllpartikeln: „If *uh* and *um* are words, [...] it is misleading to call them filled pauses. [...] we will call them *fillers*“ (Clark & Fox Tree 2002: 75). Damit ändert sich auch die Perspektive auf das Phänomen – es wird diskutiert, ob Füllpartikeln nicht doch linguistische Signale (oder gar Wörter) sind, und nicht nur Symptome der Sprachplanung. Allerdings wird *filler* auch weiterhin als Sprachplanungsbegriff in der Bedeutung ‚mark a hesitation on the part of the speaker‘ (Corley & Stewart 2008: 589f.) gebraucht. *Füller* und *gefüllte Pause* werden mittlerweile austauschbar verwendet (vgl. Brennan & Williams 1995; Fraundorf & Watson 2011; Rendle-Short 2004; Smith & Clark 1993), meinen aber nicht zwangsläufig dasselbe.

Der Terminus *Disfluenz*¹⁰ (Engl. ‚disfluency‘) erfasst eine große Menge an Phänomenen gesprochener Sprache, denen zugeschrieben wird, aus Planungsproblemen in einer beliebigen Stufe des Sprachproduktionsprozesses zu emergieren. B. A. Fox & Jaspersen (1995: 709) verstehen unter Disfluenzen, dass sie (i) den Sprachfluss unterbrechen und (ii) keinen propositionalen Gehalt zur Äußerung hinzufügen. Disfluenz wird daher, wenn sich der Begriff auf nicht-pathologische Phänomene bezieht, als Gegenstück einer ‚perfekten‘, ‚flüssigen‘ Sprachproduktion verstanden – aus generativer Perspektive ist eine Disfluenz daher auch ein Epiphänomen der Sprachverwendung (Chomsky 1965). Besonders häufig werden hierunter die englischen Äquivalente¹¹ zu Füllpartikeln sowie (sprachübergreifend) stille Pausen, Prolongationen, Wiederholungen und Reparaturen verstanden (Broen & Siegel 1972: 221; Corley & Stewart 2008: 589; Eklund 2004; Fox Tree 1995; MacGregor 2008: 7; Mahl 1956; Lickley & Bard 1998: 203; Shriberg 1994). *Disfluenz* steht also in hyperonymischer Relation zu *Füllpartikel*. Jedoch wird der Begriff *Disfluenz* seit einiger Zeit in Frage gestellt, da bestimmte Merkmale und Vorkommen von Disfluenzen systematisch auftreten und somit Informationen tragen können, die der Hörer interpretiert – „the presence, type

¹⁰Zur Diskussion der Begriffe *Disfluency* vs. *Dysfluency* mit Fokus auf pathologische Sprachstörungen vgl. Wingate (1984).

¹¹Die sich natürlich sowohl in ihrer graphematischen als auch in ihrer phonetischen Form unterscheiden, vgl. Kapitel 2.

and location of a disfluency carries potentially useful information“ (Arnold & Tannenhaus 2011: 3) Damit würde die historisch primäre Funktion des Disfluenzbegriffs (eine Disfluenz stört die flüssig ablaufende Sprachproduktion) invertiert. Da *Disfluenz* dies *eo ipso* nicht leisten kann, etablieren sich andere Termini. So verwendet Götz (2013) die Begriffe *Fluenceme* und *fluency enhancement*, um zu verdeutlichen, dass die hierunter verstandenen Phänomene (die immer noch die oben beschriebenen sind) womöglich gerade dazu beitragen, die Sprachproduktion flüssiger¹² zu gestalten.

Der Terminus *Häsitation* (Engl. ‚hesitation‘) oder *Häsitationsmarker* (Engl. ‚hesitation marker‘, Kjellmer 2003) subsumiert als funktionsimplizierender Begriff Füllpartikeln, aber auch Prolongationen (Betz et al. 2017) sowie stille Pausen und Wiederholungen (Lickley 2015; Lounsbury 1954). *Häsitation* ist konnotiert mit der Bedeutung des Zögerns.¹³ Gemeint ist die Verzögerung der *Message*, also des „Output[s] der Mikro- und Makroplanung“ (Dietrich 2007: 141), „where speakers just delay continuation of the message“ (Lickley 2015: 455). *Häsitationen* werden häufig auch hyponymisch zu den Disfluenzen geschlagen. Tatsächlich wird aber auch für *Häsitation* im Kontext der *Häsitationsforschung* selbst explizit festgestellt, dass der Begriff nicht immer zutreffend ist (Kjellmer 2003: 171). Vielmehr können Formen wie Füllpartikeln, und hierunter verstehe ich auch stark glottalisierte Formen wie [ɛ:], auch andere Funktionen erfüllen. Diese Funktionen sind sehr vielfältig (s. Kapitel 3); sie reichen von Aufmerksamkeitsherstellung bis hin zur emphatischen Verwendung. Kjellmer votiert daher für den neutraleren Begriff der ‚gefüllten Pause‘. Mit Blick auf die oben skizzierte Begriffsproblematik der gefüllten Pause (s. Seite 7) muss man diese Neutralität jedoch infrage stellen.

Weitere Begriffe ohne Anspruch auf Exhaustivität sind *Verzögerungsmarker* (Engl. ‚delay markers‘, Hudson Kam & Edwards 2008), *Gefüllte-Pause-Wörter* (Engl. ‚filled-pause words‘, Lake et al. 2011), *Konversationsgeräusche* (Engl. ‚conversational grunts‘, Ward 2006), und *Planer* (Engl. ‚planner‘ (Tottie 2013: 53; Jucker 2015a,b).

Zusammenfassend gliedert sich die terminologische Beschreibung von Füllpartikeln einerseits in Begriffe, die Füllpartikeln als Epiphänomen der Sprachverwendung

¹²Zur Unterscheidung kognitiver Flüssigkeit, Flüssigkeit der Äußerung und perzipierter Flüssigkeit vgl. Segalowitz (2010).

¹³Lat. *haesitare* – allg. *festhängen*, ling. *stottern* (Hau 2006). *Stottern* und *Häsitation* bezeichnen mittlerweile unterschiedliche Phänomene. Während mit *Stottern* ein pathologisches, angeborenes oder erworbenes Phänomen beschrieben wird, wird *Häsitation* nicht mehr pathologisch verwendet.

betrachten (*Disfluenz, gefüllte Pause, Verzögerungsmarker, Hässitation*). Sie suggerieren, dass das Auftreten der Phänomene hauptsächlich durch Sprachplanungsschwierigkeiten entsteht. Andererseits werden Begriffe verwendet, die Füllpartikeln wortähnlichen Status und pragmatischen Gehalt zuschreiben – mithin also positiv auch für den Verstehensprozess des Hörers zu werten ist (*Füller, Fluenceme*). Erschwerend kommt hinzu, dass manche Termini für beide Bedeutungen verwendet werden (*gefüllte Pause*). Schließlich existieren parallel zu den hier diskutierten Termini weitere Begriffe, welche die grammatische Perspektive von Füllpartikeln zu beschreiben versuchen, wie *Gesprächswort* (Schwitalla 2002), *Interjektion, Gliederungspartikeln* (Ehlich 1986; Willkop 1988), *Diskurspartikeln* (Fischer 2000), sowie *Hässitationsmarker* und *Diskursmarker*. Die Auswahl eines Begriffes determiniert in gewisser Weise die Sicht auf das Phänomen – daher wird in dieser Arbeit nur von *Füllpartikeln* gesprochen.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





2 Phonetische Variabilität

In diesem Kapitel beschreibe ich den Forschungsstand zur auditiven und akustischen Variabilität von Füllpartikeln. In Abschnitt 2.1 fasse ich die Variabilität in der auditiven Wahrnehmung und ihre qualitative, orthographische und symbolphonetische Repräsentationsbreite zusammen. Hieraus motiviere ich die eingehendere Betrachtung der phonetischen Realisierung. In Abschnitt 2.2 stelle ich messphonetische Parameter vor, mithilfe derer Füllpartikeln quantitativ beschrieben werden können, wie die signalgestützte Einteilung in Segmentsequenzen, Dauer, Grundfrequenz, Formanten, Phonationsart und Intensität. Zu jedem Parameter diskutiere ich vorhandene Studien zum Deutschen, in begründeten Fällen und zum besseren Verständnis auch Studien über Füllpartikeln in anderen Sprachen. Das Kapitel bereitet so die weitere Untersuchung von Füllpartikeln im Deutschen vor und motiviert ihre Analyse unter Berücksichtigung des linguistischen Kontextes (s. Kapitel 3).

2.1 Kategoriale Beschreibung

Mit dem Begriff der auditiven Repräsentation bezeichne ich die Beschreibung von Füllpartikeln mittels einer orthographischen Repräsentation, wie beispielsweise *äh* und *ähm*, oder einer symbolphonetischen (ohrenphonetischen) Repräsentation wie beispielsweise [ɛ:] und [ɛ:m]. Das jahrzehntelang vorherrschende Primat der Schriftlichkeit in der linguistischen Forschung (Bloomfield 1927; Klein 1985; Linell 1982, 2005) hat dazu geführt, dass selbst so variable Sprachlaute wie Füllpartikeln in vielen Studien graphemisch wiedergegeben werden (O'Connell & Kowal 2004). Aufgrund ihrer (teilweisen) Aufnahme in Wörterbücher und damit ihrer (teilweisen) schriftlichen Normierung spreche ich von orthographischen Repräsentationen. Die folgenden Abschnitte zeigen zunächst die Variabilität orthographischer Repräsentationen, bevor diese mit der Variabilität symbolphonetischer Repräsentationen verglichen wird. In Tabelle 2.1 ist der Unterschied zwischen symbolphonetischer und orthographischer

Repräsentation schematisiert dargestellt.

Tabelle 2.1: Schema zweier auditiver Repräsentationsebenen anhand des fiktiven Beispiels *und ähm*, gefolgt von einem hörbaren Einatmungsgeräusch. Die symbolphonetische und orthographische auditive Transliteration beschreibt die Einzelsegmente kategorial. Das postzedente Einatmungsgeräusch kann beispielsweise nicht dargestellt werden.

		Antezedentia		Füllpartikel	Postzedentia
Kategorial	Orthographisch	und	ähm
	Symbolphonetisch				?ɛm

2.1.1 Orthographische Repräsentation

Der Duden beschreibt noch 2001 nur die Lemmata *äh* und *hm*, jedoch nicht *ähm* (Dudenredaktion 2001).¹⁴ In der Forschung finden sich allerdings Abweichungen von diesen präskriptiven Einträgen. Die Abstraktion von der lautlichen Form und die damit einhergehende Kategorisierung in eine vokalische (V), eine vokalisch-nasale (VN) und eine nasale (N) orthographische (oder besser: „schriftsprachlich geläufige“, Willkop 1988: 246, Fn. 164) Repräsentation findet sich häufig in Studien zur Frequenz oder zur Distribution von Füllpartikeln. Zwar werden hierbei auch Schreibvarianten gezählt – so extrahieren Wieling et al. (2016) aus dem *Forschungs- und Lehrkorpus Gesprochenes Deutsch*¹⁵ Tokens mit den Werten *ähm*, *öhm*, *äh*, *öh*, und für das Englische aus verschiedenen Korpora Tokens mit den Werten *uh*, *um*, *erm*, *er*, *ehm* – diese werden aber schließlich fast immer unter zwei Hyperonymen zusammengefasst, „the UM form, which consists of a neutral vowel followed by a final labial nasal, and the UH form, which consists of a neutral vowel in an open syllable“ (ebd.: 199). Es findet somit eine *a-priori*-Kategorisierung bezüglich der Anzahl ihrer enthaltenen Segmente und ihrer Vokalqualität statt, die meist nicht weiter untersucht werden.

Diese Art der Darstellung führt jedoch oftmals zu Situationen, in denen nicht klar ist, ob von einer bestimmten Variante der Variablen Füllpartikeln oder von der Variablen selbst die Rede ist – was zu der grundsätzlicheren Frage führt, ob die beobachteten Repräsentationen überhaupt auf eine Variable zurückgeführt werden

¹⁴Mittlerweile hat auch *ähm* einen Eintrag: <https://www.duden.de/suchen/dudenonline/%C3%A4hm>, besucht am 24.05.2020.

¹⁵<http://agd.ids-mannheim.de/folk.shtml>, besucht am 20.12.2018.

können. Beispielsweise trägt ein Artikel von Bosker et al. (2014) *um* im Titel¹⁶, die Autoren untersuchen jedoch nur die Variante *uh* – im Titel selbst wird *um* demnach als Variablenamen verwendet.¹⁷ Andere Artikel wiederum beschreiben mit *um* im Titel durchaus die tatsächliche Variante *um* als Ausprägung der Variablen *Füllpartikel*, wie in Corley et al. (2011).¹⁸

Tabelle 2.2 gibt einen Überblick (ohne Anspruch auf Vollständigkeit) über die Variabilität der in der Literatur verwendeten orthographischen Formen für deutsche und englische Füllpartikeln.¹⁹ Es kristallisieren sich vier größere Gruppen heraus: die ⟨u⟩- und ⟨e⟩-Gruppen vor allem für englische Formen²⁰, die ⟨a/ä⟩-Gruppe vor allem für deutsche Formen, und die ⟨m⟩- beziehungsweise konsonantische Gruppe. Nach der Betrachtung der Distribution (und dies nur für das Deutsche und Englische) liegt es nahe, folgender Generalisierung zuzustimmen: „[I]t seems fairly common for filled pauses to have at least two forms in a given language, including both a prolonged vowel sound and a vowel (usually) followed by a nasal“ (Lickley 2015: 458).

Die in Tabelle 2.2 aufgeführten Schreibvarianten lassen vermuten, dass Füllpartikeln einer Einzelsprache entweder von Forschenden unterschiedlich perzipiert werden oder innerhalb einer Einzelsprache verschiedene Versionen produziert werden. Trotz einer im Deutschen und Englischen vorhandenen orthographischen Normierung werden immer wieder Schreibvarianten propagiert, von denen ich annehme, dass die Forschenden sie als relevante Abweichung von der orthographischen Normierung gehört haben und mitteilen möchten.²¹

Abweichende Formen können jedoch sehr unterschiedliche Gründe haben. Die häufigsten Beschreibungen korrespondieren mit den jeweils im Deutschen (*äh*) und Englischen (*uh*, *er*, *um*) normierten Schreibungen für Füllpartikeln (vgl. Dudenredakti-

¹⁶ „Native ‘um’s elicit prediction of low-frequency referents, but non-native ‘um’s do not“.

¹⁷ Eine solche *Pars-pro-toto*-Verwendung findet sich auch für die Varianten *um*, *uh* und *er* in Christenfeld (1995) mit dem Titel „Does it hurt to say um?“ und für die Varianten *um*, *uh*, *er* und *ah* in Christenfeld (1996) mit dem Titel „Effects of a Metronome on the Filled Pauses of Fluent Speakers“.

¹⁸ „Why *Um* Helps Auditory Word Recognition: The Temporal Delay Hypothesis“.

¹⁹ Englisch wurde als Vergleichswert ausgewählt, da ungleich mehr Forschung zu englischen Füllpartikeln existiert, und beide Sprachen als Sprachen germanischer Provenienz vergleichbar sind.

²⁰ Die Präferenz für die Wahl einer orthographischen Repräsentation scheint auch mit dem britischen Englisch (Präferenz für *er*, *erm*) und dem amerikanischen Englisch (Präferenz für *uh*, *uhm*) zusammenzuhängen (MacGregor 2008: 95; Hughes et al. 2016: 127).

²¹ Begründungen für eine bestimmte graphemische Repräsentation sind m. W. in keiner der von mir rezipierten Studien zu finden.

Tabelle 2.2: Orthographische Repräsentationen (OR) von Füllpartikeln in der deutsch- und englischsprachigen Literatur aus psychologischer, psycholinguistischer, kommunikativer, pragmatischer, prosodischer, kontrastiver, lernersprachlicher, korpuslinguistischer, physiologischer und grammatischer Perspektive (jedes Häkchen symbolisiert mindestens eine Erwähnung der OR in einer der referenzierten Studien).

OR	Englisch	Deutsch	Referenzen
<i>uh</i>	✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓		Acton 2011; Bortfeld et al. 2001; Bosker et al. 2014; Brennan & Williams 1995; Christenfeld 1995, 1996; Clark & Fox Tree 2002; Fox Tree 2001; Fraundorf & Watson 2013; Hudson Kam & Edwards 2008; Hughes et al. 2016; Kasl & Mahl 1965; Laserna et al. 2014; Lickley 2001; Schachter et al. 1991; Schegloff 2010; Shriberg & Lickley 1993; Siegel et al. 1969; Smith & Clark 1993
<i>uhh</i>	✓		O'Shaughnessy 1992
<i>um</i>	✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓		Acton 2011; Bortfeld et al. 2001; Bosker et al. 2014; Brennan & Williams 1995; Christenfeld 1995, 1996; Clark & Fox Tree 2002; Cook 1971; Fox Tree 2001; Fraundorf & Watson 2013; Hawkins 1971; Hudson Kam & Edwards 2008; Lallgee & Cook 1969; Laserna et al. 2014; Lickley 2001; Schachter et al. 1991; Shriberg & Lickley 1993; Siegel et al. 1969; Smith & Clark 1993
<i>umm</i>	✓		O'Shaughnessy 1992
<i>uhm</i>	✓✓✓		Hughes et al. 2016; Kasl & Mahl 1965; Schegloff 2010
<i>er</i>	✓✓✓✓✓ ✓✓✓✓✓ ✓✓		Brand & Götz 2013; Christenfeld 1995, 1996; Collard et al. 2008; Cook 1971; Corley et al. 2007; Fraundorf & Watson 2013; Hawkins 1971; Kjellmer 2003; Lallgee & Cook 1969; Schachter et al. 1991; Siegel et al. 1969
<i>erm</i>	✓✓		Brand & Götz 2013; Kjellmer 2003
<i>eh</i>	✓✓	✓✓	Brand & Götz 2013; Kasl & Mahl 1965; Deutsch: Künzel 1987; Rasoloson 1994
<i>em</i>	✓✓		Brand & Götz 2013; Lickley 2001
<i>hem</i>	✓		Lounsbury 1954
<i>hm</i>	✓	✓✓	Smith & Clark 1993; Deutsch: Batliner et al. 1995; Künzel 1987
<i>mm</i>	✓		Smith & Clark 1993
<i>mh</i>		✓✓	Jäger 1979; Schönle & B. Conrad 1985
<i>haw</i>	✓		Lounsbury 1954
<i>ah</i>	✓✓✓✓✓ ✓✓✓	✓	Bortfeld et al. 2001; Christenfeld 1996; Fraundorf & Watson 2013; Hawkins 1971; Kasl & Mahl 1965; Lallgee & Cook 1969; Schachter et al. 1991; Tannenbaum et al. 1965; Deutsch: Schönle & B. Conrad 1985
Ä		✓	Todt 1981
<i>äh</i>		✓✓✓✓✓ ✓✓✓	Batliner et al. 1995; Belz 2013; Belz & Klapi 2013; Belz & Reichel 2015; Belz et al. 2017; Jäger 1979; Keseling 1989; Künzel 1987; Willkop 1988
<i>ähm</i>		✓✓✓✓✓ ✓✓	Batliner et al. 1995; Belz 2013; Belz & Klapi 2013; Belz et al. 2017; Keseling 1989; Künzel 1987; Willkop 1988
<i>öh</i>		✓✓	Jäger 1979; Rasoloson 1994
<i>öhm</i>		✓	Rasoloson 1994

on (2001) und Oxford Dictionaries²²). Weder bei den dieser Normierung folgenden noch bei den abweichenden Schreibungen ist eindeutig, welche Motivation zugrunde liegt. Die normierte Schreibung und damit eine gewissen Abstraktion von der Variabilität in der phonetisch-phonologischen, graphematischen oder orthographischen Repräsentation ist eine tradierte methodische Vorgehensweise, welche es ermöglicht, generalisierte oder prototypische Aussagen über ein Phänomen oder eine Wortform zu treffen. Wenn also phonologische, graphematische oder orthographische Wörter einer Sprache mit einer abstrakten Lemmaform repräsentiert werden, unabhängig von ihrer möglicherweise reduzierten, klitisierten, epenthetisierten oder elidierten Aussprache, liegt es nahe, diese Variabilitätsreduktion auch für die orthographische Repräsentation von Füllpartikeln anzunehmen, und dabei nicht mehr jede Einzelinstanz zu betrachten. Eine solche orthographische Repräsentation hat den Vorteil, dass linguistische Aussagen über die abstrahierte Form für eine Sprachvarietät postuliert werden können, ohne die konkreten phonetischen Einzelfälle zu betrachten. Dies gilt auch und gerade für die vielfältigen akustisch möglichen Formen von Füllpartikeln. Als Beispiel sei eine Untersuchung ungarischer Füllpartikeln genannt, da diese die phonetischen Realisierungen genauer wiedergibt, als dies bisher in deutschen und englischen Studien der Fall ist. Die Verteilung der phonetischen Realisierungen ähnelt dabei stark einer Zipf-Verteilung (Zipf 1949), d. h. es gibt einige häufige und viele seltene Formen: 78,5% der ungarischen „gefüllten Pausen“ bestehen aus [ə], 10,9% aus [m], 7,6% aus [əm], 0,3% aus [əmh], und je 0,2% aus [həh], [ɛh] und [ɛə] (Horváth 2010: 294). Aus statistischer Sicht ist es häufig sinnvoll, sich auf die wenigen häufigen Formen zu stützen, da die vielen seltenen Formen zu wenige Instanzen für generalisierbare Aussagen bilden.

Andererseits lässt sich die graphemische Vielfalt in Tabelle 2.2 auch mit der allgemeinen Tendenz erklären, dass Wörter in der grammatischen Peripherie einer Sprache (damit sind häufig genuin gesprochensprachliche Phänomene gemeint) eine weniger fixierte Orthographie zeigen können. Dies ist besonders im Bereich der Partikeln und Interjektionen der Fall, in denen kleine Änderungen der phonetischen Realisierung zu einer Änderung der kontextuellen Bedeutung führen können, die aber orthographisch wenig normiert ist (vgl. *hm* vs. *mhm* vs. *hmhm*).

²²<https://en.oxforddictionaries.com/definition/uh>, <https://en.oxforddictionaries.com/definition/um>, <https://en.oxforddictionaries.com/definition/er>, besucht am 29.03.2019.

Die bisher gezeigte Variabilität in den orthographischen Repräsentationen lässt die Schlussfolgerung zu, dass Füllpartikeln eine hohe phonetische Variabilität ausstellen. Diese hohe phonetische Variabilität führt dazu, dass trotz einer existierenden orthographischen Normierung Abweichungen im Höreindruck *ad hoc* und wiederum orthographisch dokumentiert werden. Inwiefern sich die Variabilität abermals erhöht, wenn man die Granularität der Betrachtungsweise verfeinert, werde ich im nächsten Abschnitt anhand der in der Literatur erwähnten symbolphonetischen Repräsentationen von Füllpartikeln betrachten und diskutieren.

2.1.2 Symbolphonetische Repräsentation

In vielen Studien zu Füllpartikeln wird eine symbolphonetische Kategorisierung angegeben, d. h. die gehörten Segmente werden mithilfe des internationalen phonetischen Alphabets (IPA) transkribiert. Bei der meines Wissens ersten solchen symbolphonetischen Erwähnung der Vokalqualität von Füllpartikeln handelt es sich um den im Englischen vorkommenden untermittelhohen vorderen ungerundeten Vokal [ɛ] (Bloomfield 1933: 186).²³ Möglicherweise hat diese Ersterwähnung die weitere symbolphonetische Beschreibung von Füllpartikeln beeinflusst.

Tabelle 2.3 enthält eine arbiträre, nicht-exhaustive Zusammenstellung symbolphonetischer Kategorisierungen, wie sie für Füllpartikeln in deutschen und englischen Studien auftreten. Die Suche nach einer beschreibungsadäquaten symbolphonetischen Repräsentation von Füllpartikeln ist nicht der Hauptzweck der Studien in Tabelle 2.3. Vielmehr geben die Repräsentationen den subjektiv-perzeptiven Eindruck der Autorinnen und Autoren wieder. Ziel der Tabelle ist ein Überblick über die verwendeten Variantenvielfalt symbolphonetischer Repräsentationen von Füllpartikeln.

Es lassen sich vier Unterschiede in der symbolphonetischen Beschreibung erkennen: (i) es existieren vokalische, konsonantische, und vokalisch-konsonantische Formen; (ii) die Qualität des Vokals schwankt; (iii) die Epenthese von glottalen Plosiven ist unregelmäßig; (iv) über die Länge des Vokals herrscht kein Konsens. Die Repräsentationen in Tabelle 2.3 bestärken die Feststellung, dass Füllpartikeln „sehr

²³Bloomfield erwähnt eine weitere phonetische Form, deren Symbolwert jedoch aufgrund der nachträglich editierten Typographie nicht exakt zu identifizieren ist, und die entweder das in der englischen Phonologie vorhandene halbtiefe vordere ungerundete [æ] oder ein rhotisiertes Schwa [ə] beschreiben mag.

Tabelle 2.3: Symbolphonetische Repräsentationen (SR) von Füllpartikeln in der deutsch- und englischsprachigen Literatur aus psycholinguistischer, kommunikativer, pragmatischer, prosodischer, kontrastiver, lernersprachlicher und grammatischer Perspektive (jedes Häkchen symbolisiert mindestens eine Erwähnung der SR in einer der referenzierten Studien).

SR	Englisch	Deutsch	Referenzen
[ə]	✓✓✓✓✓ ✓	✓✓	Boomer & Dittmann 1962; Goldman-Eisler 1961; Maclay & Osgood 1959; Martin & Strange 1968; Shriberg & Lickley 1993; Stenström 1990; Deutsch: Batliner et al. 1995 ^b & Rasoloson 1994
[e]		✓	Batliner et al. (1995) ^b
[æ:] ^a	✓✓	✓	Hughes et al. (2016) & Stenström (1990) ^a ; Deutsch: Batliner et al. (1995) ^b
[e:]		✓	Batliner et al. (1995) ^b
[əm]	✓✓	✓✓	Shriberg & Lickley 1993; Stenström 1990; Deutsch: Batliner et al. (1995) ^b & Rasoloson (1994)
[em]		✓	Batliner et al. (1995) ^b
[ə:m] ^a	✓✓✓		Batliner et al. (1995) ^b , Hughes et al. (2016) & Stenström (1990) ^a
[e:m]		✓	Batliner et al. (1995) ^b
[e]		✓	Rasoloson (1994)
[e:] ^a		✓	Rasoloson (1994) ^a
[em]		✓	Rasoloson (1994)
[e:m] ^a		✓	Rasoloson (1994) ^a
[ɛ]	✓✓✓	✓✓✓	Englisch: Goldman-Eisler 1961; Maclay & Osgood 1959; Martin & Strange 1968; Deutsch: Batliner et al. (1995) ^b ; Rasoloson (1994) & Schwitalla (2002)
[ɛ:]		✓✓✓✓	Belz & Reichel 2015; Rasoloson 1994; Schwitalla 2002; Batliner et al. (1995) ^b
[ʔɛ]		✓	Willkop 1988
[ʔɛ:] ^a		✓	Willkop 1988 ^a
[ɛm]		✓✓	Batliner et al. (1995) ^b & Rasoloson (1994)
[ɛ:m]		✓✓	Batliner et al. (1995) ^b , Rasoloson (1994) & Schwitalla (2002)
[ʔɛ:m] ^a		✓	Willkop 1988 ^a
[œ]	✓	✓✓	Englisch: Martin & Strange 1968; Deutsch: Batliner et al. 1995; Schwitalla 2002
[ʔœ]		✓	Willkop 1988
[ʔœ:] ^a		✓	Willkop 1988 ^a
[œm]		✓	Batliner et al. (1995) ^b
[ʔœ:m] ^a		✓	Willkop 1988 ^a
[ø]		✓	Batliner et al. 1995 ^b
[ø:]		✓	Schwitalla 2002
[øm]		✓	Batliner et al. (1995) ^b
[ɑ]	✓		Goldman-Eisler 1961
[æ]	✓✓	✓	Goldman-Eisler 1961; Maclay & Osgood 1959; Deutsch: Rasoloson 1994
[æm]		✓	Rasoloson 1994
[m]	✓✓✓	✓	Goldman-Eisler (1961), Maclay & Osgood (1959) & Martin & Strange (1968); Deutsch: Batliner et al. (1995), Leeuw 2007
[hm]		✓	Batliner et al. (1995)
[r]	✓✓✓		Goldman-Eisler 1961; Maclay & Osgood 1959; Martin & Strange 1968

^a Formen mit Doppelpunkt (,:') werden als intendierte Längungsdiakritika (,:') kategorisiert.

^b Transkription im Original in SAMPA. Batliner et al. (1995) führt auch Formen mit verlängertem Nasal auf, die hier nicht übernommen werden.

unterschiedlich realisiert werden können, z. B. mit Vokalen verschiedener Klangfarbe und Dauer, mit oder ohne folgenden oder vorausgehenden Nasalkonsonanten [m], usw.“ (Künzel 1987: 37). In manchen Fällen werden die in Tabelle 2.3 beschriebenen Formen mit einem initialen glottalen Plosiv dargestellt. Phonetisch ist im Deutschen beides möglich, was mit der phonologische Regel der Glottalplosivepentheses ([ʔ]-Epenthese) vor betonten Vokalen oder vor Vokalen am Wortanfang beschrieben wird.

$$\text{[ʔ]-Epenthese: } \emptyset \longrightarrow [\text{ʔ}] / \left\{ \begin{array}{l} \text{V} - \text{'V} \\ \# - \text{V} \end{array} \right\} \quad (\text{Hall 2011: 66})$$

Diese phonologische Regel besagt, dass vor wortinitialen Vokalen (Wortanfang ist markiert mit #) im Deutschen ein glottaler Plosiv [ʔ] eingefügt wird. Die Regel wird als ‚quasi-obligatorisch‘ bezeichnet, da sie nicht immer eintritt. So stellen Pompino-Marschall & Žygis (2010) fest, dass die Realisierung des glottalen Plosivs von zusätzlichen Faktoren abhängen kann. Inhaltswörter werden stärker glottal markiert als Funktionswörter, betonte Erstsilben stärker als unbetonte, tiefe Vokale stärker glottal als nicht-tiefe und eine schnelle Sprechgeschwindigkeit führt weniger häufig zur Knacklautepentheses als eine langsame (ebd.: 12 f.). Für die phonetische Realisierung von Füllpartikeln bedeutet dies, dass keine eindeutige Vorhersage möglich ist. Zwar bestehen sie aus einer einzigen betonten Silbe, aber nicht immer aus tiefen oder halbtiefen Vokalen, sodass sowohl Formen mit als auch ohne Knacklaut erwartet werden.

Während die symbolphonetisch beschriebenen Füllpartikelvokale für das Englische zwischen Schwa [ə] und dem vorderen halbtiefen ungerundeten Vokal [æ] variieren, (vgl. auch Shriberg 2001), scheint es im Deutschen mehr Varianten zu geben. Neben dem neutralen mittleren ungerundeten Vokal Schwa [ə] wird der vordere obermittelhohe gespannte ungerundete Vokal [e] und sein gerundetes Pendant [ø] sowie der vordere untermittelhohe ungespannte ungerundete Vokal [ɛ], der vordere untermittelhohe gespannte ungerundete Vokal [ɛ:] und der vordere untermittelhohe ungespannte gerundete Vokal [œ] genannt.²⁴ Die konsonantischen Vorkommen fürs Deutsche bestehen aus dem bilabialen Nasal [m], der von einem glottalen stimmlosen Frikativ präzediert werden kann [hm].

²⁴Ob andere Ausprägungen nicht-verbalen Vokalisierungungen die gleiche Funktion wie Füllpartikeln übernehmen, bleibt offen. Denkbar sind Kategorien wie Einatmen, Ausatmen, Schnalzen, Prolongationen und das Produzieren anderer, nichtphonologischer Laute. Zu Funktionen vgl. Kapitel 3.

Abbildung 2.1 zeigt die Positionen der standardmäßig für das Deutsche beschriebenen symbolphonetischen Darstellungen im Vokaltrapez der Monophthonge nach Kohler (1999a). Der aufgespannte Raum der bisher erwähnten Werte ist mit einer gestrichelten Linie markiert. Es wird deutlich, dass die Vokalqualität von Füllpartikeln auf der vertikalen Dimension zwischen halbtiefen und obermittelhohen, sowie auf der horizontalen Dimension zwischen vorderen und zentralen Positionen verortet wird.

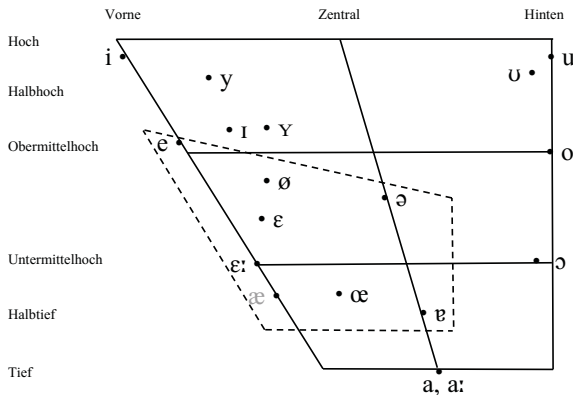


Abbildung 2.1: Vokaltrapez deutscher Monophthonge nach Kohler (1999b: 87) mit den zusätzlichen Phonen [ə], [ɐ] und dem phonologisch nicht-muttersprachlichen [æ] (in grau). Die gestrichelte Linie umrahmt die in der Literatur genannten perzipierten Vokalqualitäten in Füllpartikeln.

Zwei Vokale verdienen eine nähere Betrachtung: Schwa [ə] und halbtiefes Schwa [ɐ]. Interessanterweise wird Schwa als Füllpartikelvokal im Deutschen erst von Rasoloson (1994) erwähnt, nicht jedoch in früheren Beschreibungen (vgl. Willkop 1988). Theoretisch tritt Schwa im deutschen phonologischen System nur in unbetonten Silben auf (Hall 2011: 70), womit die Einheiten [ə] [ə:] [əm] und [ə:m] als Wörter phonologisch restringiert sind. Auch das halbtiefe Schwa ist phonologisch nicht betonbar²⁵ und nach gängiger Auffassung kein Phonem des Deutschen, sondern ein Allophon von [ʀ] (ebd.: 71). Das halbtiefe Schwa als Vokalrealisierung in Füllpar-

²⁵ Auch phonetisch mit kontrastivem Fokus nicht, da hier wahrscheinlich ein sekundärer Diphthong realisiert wird, beispielsweise in *Hast du Opa* [ʔo:pa:] *gesagt?* – *Nein, Oper* [o:peɐ].

tikeln erfährt nur eine Erwähnung in der Literatur (Batliner et al. 1995).²⁶ Die Vokalrepräsentation der Füllpartikeln wird demnach (wie auch die Glottalplosivepenthese) eher phonetisch als phonologisch dargestellt, was darauf hindeutet, dass für ihre Variabilität noch keine regelhaften Muster identifiziert worden sind. Dieses in dieser Arbeit zu untersuchende Forschungsdesiderat wird davon gestützt, dass keine der in Tabelle 2.3 erwähnten Studien die Auswirkung des linguistischen Kontext auf die phonetische Repräsentation untersucht.

Dass die Vokalqualität von Füllpartikeln im Englischen eher als Schwa repräsentiert wird, im Deutschen jedoch eher breiter gestreut ist, könnte mit sprachspezifischen artikulatorischen Einstellungen („articulatory settings“) zusammenhängen (Honikman 1964). In der artikulatorischen Ruheposition einer Sprache würde eine Phonation dann eine Füllpartikel in dieser Sprache hervorbringen – was wohl auch Schmidt (2001: 17) mit der Beschreibung als Neutralvokal und „Intonationsträger mit geöffnetem Mund“ gemeint hat – jedoch gibt es hierfür bis dato weder artikulatorische (Gick et al. 2004) noch akustische Evidenz (Foulkes et al. 2004).²⁷ Vielmehr scheint es so, also ob „filled pauses [...] have targets of their own“ (Gick et al. 2004: 231). Die Ergebnisse in Abschnitt 5.1.5 und Abschnitt 6.1.5 deuten ebenfalls in diese Richtung.

Schließlich birgt die kategoriale ohrenphonetische Beschreibung von Füllpartikeln die Gefahr, regelhafte, kontextuell bedingte Abweichungen in der phonetischen Realisierung unentdeckt zu lassen, die aber möglicherweise systematisch variieren (beispielsweise die Dauer einer Füllpartikel, ihre Vokalqualität oder ihre Grundfrequenz). Aus den bisher gesehenen impressionistisch beschriebenen Formen von Füllpartikeln ergibt sich das Desiderat einer signalphonetischen Analyse, um ein genaueres Bild der akustischen Realisierungsvariabilität zu erhalten. Im nächsten Kapitel stelle ich daher die Parameter akustischer Realisierung von Füllpartikeln und die bisherigen Erkenntnisse hierüber vor.

²⁶In jüngster Zeit wird außerdem untersucht, ob [e] im Deutschen (bei norddeutschen Versuchspersonen) noch als distinktiv zu [a] wahrgenommen wird (Bokelmann et al. 2016; Ratheke & Mooshammer to appear).

²⁷So finden Foulkes et al. (2004) eine Vokalqualität im obermittelhohen vorderen Vokalraum in der Nähe von [e] für Füllpartikeln im Dialekt der Region von Newcastle upon Tyne in Nordengland. Dies ist doppelt auffällig, da sich der Vokal [e] in diesem Dialekt von jedwedem lexikalischem Vokal unterscheidet, und die Lage von [e] im Vokaltrapez des Englischen eine Position einnimmt, die eklatant von der bisher für englische Füllpartikelvokale festgestellten Schwa-Qualität [ə] abweicht.

2.2 Akustische Beschreibung

Mit dem Begriff der akustischen Repräsentation bezeichne ich die Beschreibung von Füllpartikeln mittels im akustischen Signal vorhandener, kontinuierlicher Parameter. Die in diesem Abschnitt betrachteten phonetischen Parameter sind Dauer, Grundfrequenz, Formanten, Phonationsart und Intensität, wobei Intensität später aus der Studie ausgenommen wird (vgl. Abschnitt 4.4.4). Sprachlaute treten im akustischen Signal nicht kategorial auf, sondern werden kontinuierlich produziert. Die Segmentierung von Füllpartikeln sollte daher signalgestützt erfolgen (vgl. Abschnitt 4.2.1). Tabelle 2.4 zeigt schematisch die Ausprägungen akustischer Parameter, die ein oder mehrere Segmente betreffen können.

Tabelle 2.4: Schema verschiedener Repräsentationsebenen anhand des fiktiven Beispiels *und ähm*, gefolgt von einem hörbaren Einatmungsgeräusch. Die symbolphonetische und orthographische Transliteration beschreibt die Einzelsegmente kategorial. Die hörbaren Einzellaute sind segmentiert und mit dem Signal aligniert, der Vokal bleibt unterspezifiziert. Die akustischen Maßeinheiten werden in Kapitel 4 beschrieben.

		Antezedentia	Füllpartikel	Postzedentia
Auditiv	Orthographisch und	ähm	
	Symbolphonetisch		?ɛm	
Akustisch	Segmentierung		? Vokal m	
	Dauer		ms ms ms	
	Grundfrequenz		Hz	
	Formanten		Hz	
	Phonation		modal	
	Intensität		H ₁ -A ₃ H ₁ -A ₃	

2.2.1 Segmente

Mit Segmenten bezeichne ich hier die signalphonetisch unterscheidbaren zeitlich sequenziellen Laute, die in Füllpartikeln enthalten sind. Bisher werden im Deutschen symbolphonetisch Instanzen von Vokal (V) oder Glottalplosiv-Vokal (GV), Vokal-Nasal (VN) oder Glottalplosiv-Vokal-Nasal (GVN) und Nasal (N) beobachtet (vgl. Tabelle 2.3). Während der Nasal in der Literatur nur bilabial auftritt, kann die Vokalqualität durchaus divers realisiert werden (vgl. Abschnitt 2.2.4). Dass der glottale

Plosiv vor Vokalen tatsächlich signalphonetisch vorhanden ist, wird fürs Deutsche von Rodgers (1999) bestätigt. Zusätzlich zu diesen abstrahierten Segmentklassen nehme ich an, dass die segmentale Ausprägung nicht auf die obigen Entitäten begrenzt, sondern potenziell offen für andere Realisierungen ist, die Füllpartikeln zugeordnet werden können. Auch ist weder für den bilabialen Nasal noch für die GV-/GVN-/V-/VN-/Varianten klar, ob ein glottaler Frikativ [h] involviert ist.

2.2.2 Dauer

Die Dauer (oder Länge) einer Füllpartikel bezeichnet die zeitliche Ausdehnung von ihrem akustischen Beginn bis zu ihrem Ende. Je nach Kontext können diese Grenzen unterschiedlich gut bestimmt werden. Steht die Füllpartikel zwischen zwei stillen Pausen, kann ihr Beginn und Ende relativ eindeutig bestimmt werden.²⁸ Steht die Füllpartikel zwischen zwei Segmenten, beeinflusst meist die Art des Segments die Genauigkeit der Abgrenzung und damit der gemessenen Länge der Füllpartikel. So sind beispielsweise V-Füllpartikeln nach Plosiven und Frikativen gut abgrenzbar, nach Vokalen jedoch nicht.²⁹

Gemessen wird entweder die Gesamtlänge einer Füllpartikel oder ihre einzelnen Segmente. Vokalisch-nasale (VN) Instanzen werden als länger als vokalische (V) Instanzen beschrieben, da sie zusätzlich einen Nasal besitzen (Clark & Fox Tree 2002).³⁰ In einer der frühesten quantitativen Studien im Deutschen erzählen 56 Biologiestudenten einen fünf Seiten langen Text nach und mussten anschließend 35 Minuten darüber sprechen (Todt 1981). Die mittlere Dauer von „Ä“ wird mit 180 ms angegeben, die kürzeste Dauer mit 60 ms und die längste Dauer mit 500 ms.³¹ In Bildbeschreibungen ausschließlich männlicher Versuchspersonen beträgt die mittlere Füllpartikeldauer (unnormalisiert) 377 ms für V- und 491 ms für VN-Formen (Jessen 2012). Tatsächlich gibt es kaum Studien zu Füllpartikeln, die auch einen detaillierten Überblick über die (sprechernormalisierte) Dauer angeben. Leeuw (2007:

²⁸ Auch hier gibt es jedoch je nach Heuristik unterschiedliche Möglichkeiten, eine Grenze zwischen stiller Pause und Füllpartikel zu setzen, da die Amplitude (sichtbar im Oszillogramm) graduell zu- und abnimmt.

²⁹ Im besten Fall orientiert sich die Segmentierung auch am Sonagramm, um die Abgrenzung beispielsweise anhand der in den Formanttrajektorien sichtbaren Vokaltransitionen einzugrenzen.

³⁰ Einschränkung muss festgehalten werden, dass Clark & Fox Tree (2002) die Dauern von Füllpartikeln nicht gemessen, sondern aufgrund ihrer Transkriptionen geschätzt haben.

³¹ Unklar bleibt, ob tatsächlich nur die V-Form und nicht auch die VN-Form untersucht wurde.

96) vergleicht in Interviews mit 21 deutschsprachigen und 16 englischsprachigen Versuchspersonen deutsche mit englischen V-, VN- und N-Füllpartikeln. Da ihr Ziel eine kontrastive Analyse ist, stellt sie nur den Vergleich zwischen Einzelsprachen an (und findet keinen signifikanten Unterschied), analysiert jedoch nicht die Dauern der einzelnen V-, VN-, und N-Repräsentationen innerhalb einer Einzelsprache.

Tabelle 2.5: Übersicht über Dauern von vokalischen (V), vokalisch-nasalen (VN) und nasalen (N) Füllpartikeln in verschiedenen Studien, sortiert nach V-Dauer (geringste zuerst).

Aufnahme	Sprache	Situation	VP	\bar{x}_V^a (s) in ms	\bar{x}_{VN} (s) in ms	\bar{x}_N (s) in ms	Studie
Labor	DEU	Interviews	21 (m/w)	317 (113)	457 (161)	470 (234)	Leeuw (2007)
Labor	DEU	Bildbe- schreibung	72 (m)	377	491,3	–	Jessen (2012) ^b
Labor	ENG	Interviews	16 (m/w)	379 (207)	493 (199)	330 (178)	Leeuw (2007)

^a \bar{x} steht für das arithmetische Mittel, *s* für die Standardabweichung einer Stichprobe.

^b Mittelwerte von mir ermittelt.

Neben der Gesamtdauer von Füllpartikeln sind Interaktionen der Dauer mit dem Kontext erwähnenswert, die in Studien zu anderen Sprachen festgestellt wurden. In niederländischen Bildbeschreibungen findet Swerts (1998: 491) längere Füllpartikeln in initialer als in nicht-initialer Position. In spontansprachlichen ungarischen Monologen und aufgabenbasierten Dialogen sind diejenigen Füllpartikeln, die zwischen zwei stillen Pausen stehen, länger als der Durchschnitt aller Füllpartikeln (Gósy et al. 2017: 161). Und in V- und VN-Füllpartikeln von je drei Äußerungen 42 Sprecher in englischen aufgabenbasierten Monologen (O’Shaughnessy 1992) liegen Füllpartikeln an größeren syntaktischen Grenzen im Bereich von 200–500 ms, während sie innerhalb kleiner syntaktischer Einheiten (wie Nominalgruppen, ebd.: I-521) im Schnitt kürzer (170–320 ms) sind. Es liegt nahe, einen prosodischen Effekt der Füllpartikelposition auf die Dauer anzunehmen, was in Abschnitt 6.2.3 schließlich gezeigt wird.

Manchmal wird auch das Verhältnis zwischen dem Vokal und dem Nasal in VN-Repräsentationen betrachtet. Shriberg (1994) berichtet jedoch für englische VN-Repräsentationen keine signifikante Korrelation zwischen der Vokal- und der Nasaldauer derselben Entität in 238 Instanzen der 70 untersuchten Sprecherinnen und Sprecher: „In some tokens the vowel is sustained; in others the nasal is sustained;

in still others both phones are sustained“ (Shriberg 1994: 175 f.). Jessen (2012: 154) postuliert für das Deutsche, dass „ein *äh* vor *m* kürzer ist als ein *äh* ohne nachfolgendes *m*, da die Laute in Lautkombinationen meistens einen Kürzungseffekt aufweisen“. Dieser Effekt wird in Abschnitt 5.1.3 und 6.1.3 analysiert.

2.2.3 Grundfrequenz

Die Grundfrequenz (f_0) als messphonetische physikalische Eigenschaft wird durch die „Rate der Schwingung der Stimmlippen bestimmt“ (Pompino-Marschall 2009: 39). Sie kann interindividuell („kürze Stimmlippen schwingen schneller als längere“, ebd.) sowie intraindividuell variieren. Die intraindividuelle Variation kann linguistisch funktional eingesetzt werden (vgl. Ladd 1996). Das auditive Korrelat der Grundfrequenz ist die Tonhöhe der Stimme, welche die Basis der Intonation, also der wahrgenommenen Stimmtonmelodie konstituiert (vgl. Laver 1994: 194). Im Folgenden stelle ich zunächst zwei Studien zur Höhe und Kontur der Grundfrequenz englischsprachiger Füllpartikeln vor, bevor ich auf das Deutsche eingehe. Mithilfe dieses Parameters können Aussagen darüber getroffen werden, ob sich Füllpartikeln einheitlich oder unterschiedlich verhalten, und ob sie möglicherweise mit einer anderen f_0 geäußert werden als ihre Umgebung und somit prosodisch prominent werden.³²

In englischen aufgabenbasierten Monologen zeigen V- und VN-Füllpartikeln eine abfallende (5–20 Hz) oder flache Intonationskontur bei relativ niedriger Grundfrequenz, die in den unteren 15% des Wertebereichs der Sprecher*innen endet. An syntaktischen Grenzen³³ beginnen sie mit einer höheren f_0 und fallen dann ab, wobei unklar bleibt, ob an diesen Grenzen auch eine neue Intonationsphrase beginnt. Innerhalb syntaktischer Einheiten zeigen sie tiefere f_0 -Muster (O’Shaughnessy 1992: I-523).

Shriberg & Lickley (1993) untersuchen V- und VN-Füllpartikeln aus amerikanischen aufgabenbasierten Monologen und britischen spontanen Dialogen. Sie schränken die Analyse auf alle Füllpartikeln ein „that followed lexical material indicating that the utterance could not end before the filled pause“ (ebd.: 174). Sie messen den Füllpartikeln vorangehenden und folgenden f_0 -Gipfel sowie den ersten und letzten

³²Zu Tonakzent im Deutschen als Korrelat für prosodische Prominenz vgl. Baumann & Winter (2018).

³³Hier wird nicht spezifiziert, um welche syntaktische Grenzen es sich handelt.

f_0 -Wert von Füllpartikeln. Wenn Füllpartikeln ein hoher Gipfel vorangeht, so tendieren diese signifikant häufig auch zu höheren f_0 -Werten. Insgesamt verhält sich also die Grundfrequenz in Füllpartikeln ähnlich wie die Grundfrequenz ihres Kontextes. Shriberg & Lickley finden keinen Einfluss der Dauer auf die Grundfrequenz.

Rasoloson (1994) untersucht vokalische und vokalisch-nasale Füllpartikeln im Deutschen³⁴ in 132 sogenannten „Informationsdiskursen“ zwischen 1975 und 1986, das sind Alltagssituationen, in welchen eine Information eingeholt wird (ebd.: 28 f.). Sie findet

keine funktionale Typologisierung von ÄH [Zusammenfassung von V- und VN-Füllpartikeln, M.B.] nach tonalen Strukturen [...]. Aus ohrenphonetischen und instrumentellen Analysen ergibt sich bei allen ÄH seine ebene Tonstruktur, die sich jedoch nach der Tonhöhenstufe (hoch, mittel, niedrig) und nach der Form (kurz, einfach, lang oder iteriert) voneinander unterscheiden läßt. Der Ton hat also in der Regel für die Interjektion ÄH keine bedeutungsunterscheidende Funktion. Der steigende Ton (äh?) stellt einen Ausnahmefall dar und ist [...] mit der Illokution ‚Frage‘ in Verbindung zu setzen. (ebd.: 43 f.)

Batliner et al. (1995) untersuchen deutsche Spontansprache von 56 Frauen und 81 Männern aus München und Karlsruhe, die Termine vereinbaren. Füllpartikeln zeigen wie bei Shriberg & Lickley (1993) eine geringere f_0 als ihre Umgebung. Der f_0 -Verlauf ist fallend. Batliner et al. folgern, dass sich Füllpartikeln wie „parenthetische Chunks“ verhalten, die eine geringere Grundfrequenz als ihre Umgebung haben (Batliner et al. 1995: 4).

Klug (2013) untersucht deutsche Spontansprache von zehn männlichen Polizeibeamten mit dialektalem Hintergrund in Notrufmitschnitten.³⁵ Auch in diesen Daten zeigen Füllpartikeln eine tiefere f_0 -Realisierung als die restlichen Äußerungen.

Braun & Rosin (2015) analysieren die Grundfrequenz von acht deutschen Frauen zwischen 45 und 65 Jahren, die an drei Tagen jeweils ein paar Minuten aufgabenfreie Spontansprache in ihrem Zuhause produzierten.³⁶ Der f_0 -Mittelwert der analysierten Füllpartikeln ist signifikant geringer als die über die restliche Sprachaufnahme gemittelte Grundfrequenz. Dieser Vergleich ist allerdings weniger aussagekräftig als

³⁴Außerdem im Englischen, Französischen und Madagassischen.

³⁵Wobei unklar bleibt, welcher Dialekt untersucht wird und ob der gleiche Dialekt für alle Beamten zutrifft.

³⁶Ursprünglich hatten sie 10 Versuchspersonen aufgenommen, zwei allerdings auch deswegen ausgeschlossen, „because they [...] did not produce nearly as many hesitations as were expected“ (Braun & Rosin 2015).

der Vergleich mit dem adjazenten Auftretenskontext von Füllpartikeln, da er sehr wahrscheinlich auch für andere Wörter gefunden werden kann.

Die N-Füllpartikel wird in Studien zur Füllpartikelgrundfrequenz nicht thematisiert, obwohl oder möglicherweise gerade weil sie ein großes Spektrum an Grundfrequenzrealisierungen enthalten kann. Schmidt (2001: 25) fasst Form-Funktions-Zuordnungen von *hm* und *hmhm* mit einem idealisierten Tonhöhenverlauf zusammen. Am ehesten einer Füllpartikel zuzuordnen ist der Formprototyp 2, der einen langen progredienten Grundfrequenzverlauf zeigt.³⁷

Tabelle 2.6 fasst die Ergebnisse der genannten Studien nochmals kurz bezüglich ihrer Aufnahmesituation, ihrer Sprache, ihres Registers (inkl. Probandenanzahl und -eigenschaften) sowie ihrer f_0 -Höhe und -Kontur zusammen.

Tabelle 2.6: FP-Grundfrequenz im Englischen und Deutschen für V- und VN-Formen.

Aufnahme	Sprache	Situation	f_0	Studie
Feld	DEU	Zuhause, 8 weibl. Versuchspersonen zw. 45–65 J.	< als mittlere f_0	Braun & Rosin 2015
Feld	DEU	Notrufmitschnitte, 10 männl. Beamte zw. 32–54 J.	< als mittlere f_0	Klug 2013
Labor	DEU	339 Min. Mensch-Mensch-Terminvereinbarungen, 56 weibl. u. 81 männl. Versuchspersonen	< als Umgebungsäußerung; Richtung: fallend	Batliner et al. 1995
Feld	DEU	132 Informationsdiskurse verschiedener Versuchspersonen	hoch, mittel, niedrig; Richtung: eben, in Fragen steigend	Rasoloson 1994
Labor	ENG	6 aufgabenfreie spontane Dialoge und Mensch-Fakemaschine-Reisevereinbarungen (<i>Wizard-of-Oz</i>) von 14 männl. und 15 weibl. Sprechern	ähnlich wie Umgebungsäußerung; Richtung: fallend	Shriberg & Lickley 1993
Labor	ENG	Je die ersten drei Äußerungen aus Mensch-Maschine-Reisevereinbarungen von 42 verschiedene Versuchspersonen	niedrig; fallend	Richtung: O'Shaughnessy 1992

Es wird deutlich, dass sowohl im Deutschen als auch im Englischen und in unterschiedlichen Registern die Grundfrequenz von V- und VN-Füllpartikeln ähnliche Muster zeigt. Füllpartikeln sind in die von einigen Studien berücksichtigten linguistischen Kontexte eingebettet und zeigen dort eine etwas niedrigere Grundfrequenz als ihre Umgebung, mit fallender Richtung. V- und VN-Füllpartikeln mit steigender Grundfrequenz kommen nur in Fragen vor (vgl. auch „*um, excuse me!*“ Lickley

³⁷Schmidt bezeichnet diesen Typ als „Turnhalten, Turnbeanspruchen“, mit den hörerseitigen Paraphrasen „Ich muss nachdenken“ und „Moment mal...“.

2015: 463). Der Effekt, dass Füllpartikeln eine niedrigere f_0 als ihre direkte lautliche Umgebung zeigen, wird in dieser Arbeit repliziert (vgl. Abschnitt 5.2.1.2 und 6.2.1.2).

2.2.4 Formanten

Während in Abschnitt 2.1 die perzipierte, kategoriale Vokalqualität und ihre symbolphonetische Beschreibung erörtert wurde, widme ich mich hier der akustischen, im Signal gemessenen kontinuierlichen Vokalqualität von Füllpartikeln. Die Vokalqualität kann mithilfe der ersten beiden Formanten F_1 und F_2 quantifiziert werden.³⁸ F_1 beschreibt die Zungenhöhe in der vertikalen Dimension (superior-inferior), wobei niedrige F_1 -Werte auf eine hohe Zungenposition (beispielsweise [i ɪ ʊ u]) und hohe F_1 -Werte auf eine niedrige Zungenposition deuten (beispielsweise [a]). F_2 beschreibt die Zungenlage in der horizontalen Dimension (anterior-posterior), wobei niedrige F_2 -Werte auf eine hintere Zungenlage (beispielsweise [u o ɔ]) und hohe F_2 -Werte auf eine vordere Zungenlage deuten (beispielsweise [i e ɛ]).³⁹ Prinzipiell können wie bei der Grundfrequenz sowohl statische Messpunkte (z. B. in der Mitte eines Vokals) als auch die Formanttrajektorien über die gesamte Vokallänge hinweg untersucht werden. Die bisherige Literatur hat sich fast nur an statischen Messpunkten orientiert – aufgrund ihrer reduzierten Komplexität sind diese einfacher zu untersuchen als die Trajektorien. Zudem hat schon Todt (1981: 598) für Füllpartikeln in Nacherzählungen von Texten bei deutschen Muttersprachlern festgestellt, dass „der Mittel- und Endteil ihrer Formantenverläufe [...] sich als sehr variabel [erwies], und zwar auch dann, wenn nur die vom gleichen Versuchsindividuum stammenden Laute miteinander verglichen wurden“. Auch in der vorliegenden Arbeit werden keine konsistenten Ergebnisse zu Formanttrajektorien bei Füllpartikeln gefunden. Im Folgenden diskutiere ich zwei Studien zur Füllpartikelvokalqualität im Deutschen und ziehe anschließend zur weiteren Erläuterung sprecherspezifischer Variabilität eine Studie

³⁸Formanten sind die im Vokaltrakt entstehenden Resonanzfrequenzen der Stimmlippenschwingungen, die durch die Zungenposition modifiziert wird. Die Schwingungen werden dadurch an bestimmten Stellen in ihrem Spektrum verstärkt, und durch die Zungenposition beeinflusst. Diese verstärkten Frequenzbereiche erhöhter Energie lassen sich signalphonetisch ermitteln und geben einen Hinweis darauf, mit welcher Zungenposition ein Vokal artikuliert wird.

³⁹Die Beziehung zwischen artikulatorischer Stellung und akustischem Signal ist jedoch nicht uneindeutig. Ein bestimmtes akustisches Signal kann durch unterschiedliche Zungenpositionen zustande kommen (Johnson et al. 1993).

zu englischen Füllpartikeln und zur Motivierung der kontextuellen Betrachtung eine Studie zu ungarischen Füllpartikeln hinzu.⁴⁰

Klug (2013) untersucht deutsche Spontansprache von zehn männlichen Polizeibeamten mit dialektalem Hintergrund in Notrufmitschnitten.⁴¹ Sie analysiert fünf äquidistante Messpunkte aus den Formanttrajektorien und findet neben monophthongischen auch diphthongische und verzeinzelt triphthongische Realisierungen des vokalischen Anteils der Füllpartikeln. Die monophthongischen Vokale sind dabei innerhalb der Wertemenge für [ə ɐ ø: ʏ ʊ] und [œ] zu finden.⁴² Klug schließt aus ihrer Studie, dass der Gebrauch und die Realisierung von Füllpartikeln tendenziell sprecherspezifisch sein kann. Allerdings macht sie keine Aussage über den linguistischen Kontext – es ist somit nicht nachprüfbar, ob der Unterschied zwischen den Sprechern durch die Verwendung anderer linguistischer Kontexte zustandekommt.

Pätzold & Simpson (1995) untersuchen drei männliche Sprecher⁴³ aus dem norddeutschen Raum in spontansprachlichen dialogischen Mensch-Mensch-Terminvereinbarungen. Ihre Füllpartikelvokale wurden in der Mitte gemessen und können im Bereich [ɛ ə ɐ] verortet werden. Linguistische Kontexte der Füllpartikeln bleiben unberücksichtigt. Abbildung 2.2 zeigt die Mittelwerte und Standardabweichungen dieser Füllpartikelvokale in einem Vokaltrapez zusammen mit Referenzmonophthongwerten männlicher Sprecher aus demselben Korpus (Pätzold & Simpson 1997).⁴⁴ Die drei Sprecher aus Pätzold & Simpson (1995) zeigen auf den ersten Blick sehr unter-

⁴⁰Da ich mich in dieser Arbeit besonders auf die einzelsprachliche Betrachtung fokussiere, blende ich weitere Studien aus, obschon kontrastiv weiterführende Befunde existieren, beispielsweise im Französischen (Vasilescu et al. 2005), im kanadischen Dialekt Outaouais (Séguin 2008), im Ungarischen (Horváth 2010), im Polnischen (Karpíński 2013), und im Russischen (Stepanova 2007). Candea et al. (2005) vergleichen Arabisch, Chinesisch, Französisch, Deutsch, Italienisch, Portugiesisch, Englisch und Spanisch und zeigen, dass Füllpartikeln nicht in allen Sprachen eine zentrale Position im Vokaltrapez einnehmen müssen, sondern verschieden realisiert werden können.

⁴¹Wobei unklar bleibt, welcher Dialekt untersucht wird und ob der gleiche Dialekt für alle Beamten zutrifft.

⁴²Messwerte sind nicht in der Studie enthalten.

⁴³Pseudonyme und Metadaten der drei Sprecher: TIS, 25 Jahre, niedersächsischer Hintergrund, Student; OLV, 34 Jahre, schleswig-holsteinischer Hintergrund, Student; GEP, 58 Jahre, niedersächsischer Hintergrund, Professor (vgl. die Korpusdokumentation zum *Kiel Corpus* unter https://www.isfas.uni-kiel.de/de/linguistik/forschung/kiel-corpus/docs/Info_KielCorp_2017.pdf, besucht am 29.01.2019).

⁴⁴Da Formantwerte sprecherspezifische Daten sind und unter anderem von biologischen Eigenschaften wie der individuellen Vokaltraktlänge abhängen, sollten sie zur Vergleichbarkeit der Sprecher und Sprecherinnen untereinander normalisiert werden (vgl. Adank et al. 2004). Eine solche Normalisierung kann diese rekonstruierte Abbildung der veröffentlichten Daten nicht leisten.

schiedliche Vokalrealisierungen. Aus messphonetischer Perspektive hingegen liegen die drei Mittelwerte der Sprecher in Abbildung 2.2 ungefähr in dem von Klug (2013) eruierten Formantraum für Füllpartikeln ([ə ɐ ø; ʏ ʊ œ]). Weiterhin wird der Unterschied zwischen einer auditiven und einer akustischen Herangehensweise deutlich. Überträgt man den für Füllpartikelvokale symbolphonetisch auditiv beschriebenen (gestrichelten) Raum aus Abbildung 2.1 im Abschnitt 2.1.2 zur Orientierung in die Abbildung 2.2, so liegt der Vokal [ɻ] im auditiv bestimmten Raum aus Abbildung 2.1 außerhalb der auditiv beschriebenen Vorkommen, im akustisch bestimmten Raum in Abbildung 2.2 jedoch innerhalb der akustisch bestimmten Vorkommen. Dies rührt daher, dass die auditiv beschriebenen Vokalqualitäten sowohl phonologisch als auch perceptiv motiviert sind, während messphonetisch ermittelte Vokalmittelwerte oft näher beieinander liegen, als eine qualitative Beschreibung suggeriert. In diesem Fall folgt daraus, dass der Vokal [ɻ] auditiv nicht als Füllpartikelvokal beschrieben wird, messphonetisch jedoch (zumindest nach den Formantdaten aus Pätzold & Simpson 1997) nahe [ə] und [ø] im potenziellen Formantraum für die Realisierung von Füllpartikelvokalen liegt.

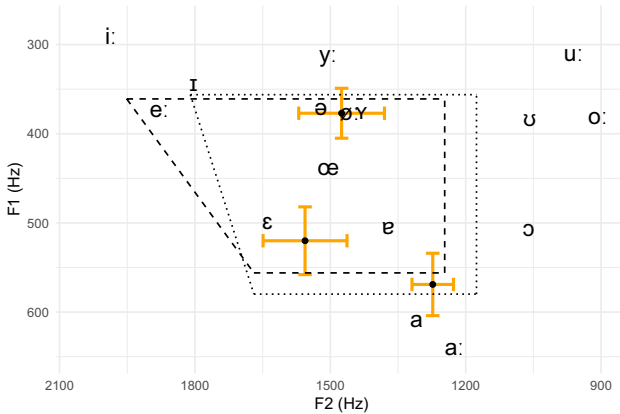


Abbildung 2.2: Mittelwerte (schwarze Punkte) und Standardabweichungen (orangene Balken) für V- und VN-Formen der drei Sprecher in Pätzold & Simpson (1995), abgebildet in dem aus Formantmittelwerten von Monophthongen männlicher deutscher Sprecher im *Kiel Corpus* rekonstruierten Vokaltrapez (Pätzold & Simpson 1997), mit auditiv beschriebenen (gestrichelt, vgl. Abbildung 2.1) und akustisch gemessenen FP-Qualitäten (gepunktet, vgl. Klug 2013; Pätzold & Simpson 1995).

Gósy et al. (2017) wertet die Füllpartikelvokale 30 ungarischer Lerner des Englischen aus. Sie finden einen Unterschied in F_1 für den Faktor Position, d. h. wenn der Vokal zwischen einem Wort und einer Pause steht (egal in welcher Reihenfolge) ist F_1 um ca. 18 Hz niedriger, als wenn der Vokal zwischen zwei stillen Pausen steht (ebd.: 162). Dies ist meines Wissens die erste Studie, die für die Vokalqualität von Füllpartikeln auch den Kontext berücksichtigt. Offen bleibt, ob dieser Unterschied perceptiv wahrnehmbar ist.

Die Ergebnisse der bisherigen Studie zeigen, dass Unterschiede in der Füllpartikelvokalrealisierung besonders auf sprecherspezifische und crosslinguistische Unterschiede zurückgehen. Beides ist erwartbar. Crosslinguistische Unterschiede sind erwartbar, da die Sprachen der Welt unterschiedlich große Vokalräume besitzen, und es daher unterschiedliche Präferenzen für die Vokale in Füllpartikeln geben kann. Sprecherspezifische Unterschiede sind erwartbar, da interindividuelle Unterschiede eine der wenigen Konstanten der phonetischen Forschung sind – sie spiegeln zufällig verteilte physiologische, psychische, somatische, soziale und regionale Variablen wider, die sich in den Äußerungen konkreter Sprecherinnen und Sprecher niederschlagen. Der linguistische Kontext als möglicher Einfluss auf die Füllpartikelvokalrealisierung wird in der Literatur nur am Rande thematisiert.⁴⁵

2.2.5 Phonationsart

Sprache kann mit verschiedenen Stimmlippeneinstellungen (Phonationsarten) produziert werden. Für diese Arbeit relevant sind modale und glottale Phonation.⁴⁶ Modale Phonation (*modal voice*) bezeichnet die regelmäßige Schwingung der adduzierten Stimmlippen. Glottale, glottalisierte oder laryngalisierte Phonation (*creaky voice*) bezeichnet eine knarrende oder schnarrende Stimme mit geringer Grundfrequenz, begleitet von einer nahezu vollständigen Dämpfung der Glottalimpulse (Redi & Shattuck-Hufnagel 2001: 414). In Abgrenzung zur Glottalisierung in phraseninitialer und -medialer Position bezeichne ich Glottalisierung am Ende einer Phrase, die mit einer fallenden Grundfrequenz einhergeht, als *Laryngalisierung* (vgl. Kohler

⁴⁵So analysieren Candea et al. (2005) nur Füllpartikeln, die von stillen Pausen umgeben sind (Pause-Wort-Pause); Gósy et al. (2017) erweitern diesen Kontext um die Variante Wort-Pause beziehungsweise Pause-Wort.

⁴⁶Weitere Phonationsarten sind Flüstern, behauchte Phonation, Falsetto und verschiedene Mischformen (vgl. Laver 1994: 199).

et al. 2005). Glottalisierung entsteht durch die fest geschlossene adduzierte Stellung der Aryknorpel (der posterioren Aufhängung der Stimmlippen an den *Cartilagine arytenoideae*), so dass die Stimmlippen nur an ihrem anterioren Ende in Richtung des Kehlkopfknorpels (*Cartilago thyroidea*) aneinanderschlagen können (Ladefoged & Johnson 2011: 149 f.). Hierbei entsteht eine tiefe Grundfrequenz (zwischen 25 und 50 Hz) und die aufeinanderfolgenden Pulse der Stimmlippen sind unregelmäßig (Laver 1994: 194 f.), bei niedrigerem subglottalem Luftdruck (Blomgren et al. 1998: 2653). Der perzeptive Eindruck von Glottalisierung kann auch durch andere glottale Einstellungen entstehen, beispielsweise durch aperiodische Stimmlippenschwingungen (Aperiodizität), durch reguläre Alternation der Form, Dauer oder Amplitude glottaler Perioden (Diphonie), oder durch *glottal squeak*, einem plötzlichen Wechsel zu einer hohen Grundfrequenz bei geringer Amplitude (Redi & Shattuck-Hufnagel 2001: 414; Pompino-Marschall & Žygis 2010: 4 f.).

Wieso die Beschreibung der Phonation für die Füllpartikelforschung interessant sein kann, zeigen exemplarisch Daten aus den von mir untersuchten Korpora (vgl. Kapitel 4) in Abbildung 2.3. Im linken Bild ist die modale Phonation eines *äh* mit regelmäßigen Perioden, gleichmäßiger Amplitude und vorangehendem glottalen Plosiv, im rechten Bild die glottale Phonation eines *äh* mit aperiodischen Stimmlippenschwingungen, und im unteren Bild eine Aneinanderreihung glottaler Plosive zu sehen. Aus dem linken Bild lässt sich entnehmen, dass es bei der [ʔ]-Epenthese phonetisch nicht immer zu genau einem glottalen Plosiv kommt, sondern wie in Abbildung 2.3 links oben auch zwei (oder mehr) aufeinanderfolgende Ereignisse produziert werden können.

Nach Blomgren et al. (1998: 2652) sprechen Männer aufgrund ihrer biologischen Voraussetzungen (durchschnittlich größerer Kehlkopf und längere Stimmlippen) mit einer etwas tieferen Stimme als Frauen. Während dieser Unterschied bei modaler Phonation bemerkbar ist, kann der Unterschied bei glottaler Phonation verschwinden: (US-amerikanische) Männer produzieren glottale Stimme mit 49 Hz (Standardabweichung 3,7 Halbtöne), (US-amerikanische) Frauen mit 48 Hz (Standardabweichung 3,5 Halbtöne). Aufgrund dieser Interaktion beim perzeptiven Eindruck ist die Phonationsart ein Parameter, der bei der Auswertung der Grundfrequenz berücksichtigt werden sollte.

Im Kiel Corpus spontaner gesprochener Sprache findet Rodgers ein ähnliches Mus-

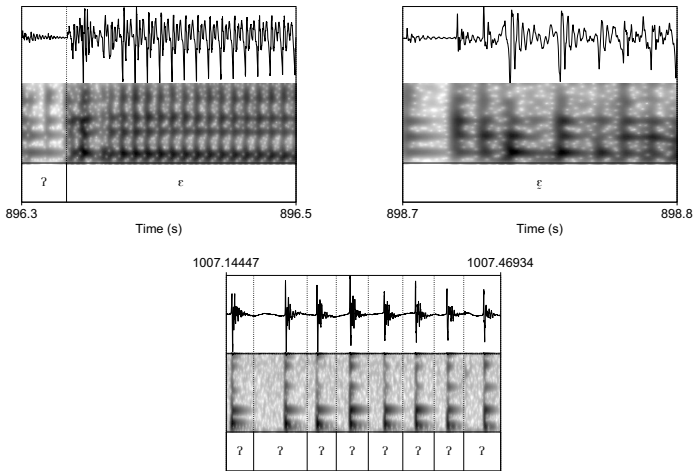


Abbildung 2.3: Glottaler Plosiv [ʔ] mit nachfolgendem modal phoniertem Vokal [ɛ] (linkes Bild) und glottalisierter Vokal [ɛ̥] (rechtes Bild) und Sequenz glottaler Plosive (unteres Bild). Quelle rechts und links: Berlin Dialogue Corpus compiled v. 1 (BeDiaCo_c) frei_flf2_ch1, Quelle unten: GERman COversations-Füllpartikeln v. 1 (GECO-FP) multi_F-K_right.

ter für Füllpartikeln wie für andere Wörter in Spontansprache.⁴⁷ Von 134 V- und VN-Füllpartikeln in akzentuierter Position werden 11 mit einem einzelnen glottalen Plosiv vor dem Vokal [ʔɛ], 59 mit einzeitigem glottalem Plosiv und glottalisiertem Vokal [ʔɛ̥], 53 mit glottalisiertem Vokal ohne vorangehenden glottalen Plosiv [ɛ̥] und 11 mit modal phoniertem Vokal ohne vorangehenden glottalen Plosiv [ɛ] produziert (Rodgers 1999: 268).⁴⁸

Eine verfeinerte Analyse nimmt Shriberg vor, indem sie in englischen Füllpartikeln den Vokal nicht als ganzes analysiert, sondern modal und glottal realisierte Abschnitte identifiziert. Ihr Fazit ist, dass Füllpartikeln tendenziell mit Knarrstimme („creaky voice“) enden (Shriberg 2001: 162). Die glottalisierte Realisierung von Füll-

⁴⁷Offen bleibt, wie viele Sprecherinnen und Sprecher zu diesem Zeitpunkt in die Analyse einbezogen wurden.

⁴⁸Rodgers (1999) verwendet zur Darstellung des Vokals die Majuskel V in eckigen Klammern mit dem Tilde-Diakritikum als Subskript. Aus typographischen Gründen verwende ich hier das Zeichen ε – dies ist hier jedoch nicht als Aussage über die Qualität des Vokals zu verstehen.

partikeln wird auch in anderen Sprachen beobachtet (für Schwedisch Horne 2006, 2009; für Französisch Duez 2001; für Polnisch Karpiński 2013: 69).

2.2.6 Intensität

Über die Intensität von Füllpartikeln ist mir nur eine Studie zum Japanischen bekannt. Dort zeigen die in Monologen produzierten Füllpartikeln „/aH/, /eH/, and /oH/“ [sic] (Maekawa & Mori 2015: 61) eine signifikant niedrigere Intensität als dieselben Vokale in lexikalischen Wörtern (ebd.). Unklar ist allerdings, was dieses Ergebnis bedeutet. Auch wird in dieser Studie der direkte Kontext der Füllpartikeln nicht berücksichtigt. Die Intensität von Füllpartikeln bleibt ein interessantes Forschungsdesiderat, da sie möglicherweise auch eine Erklärung über ihre wahrgenommene Prominenz bietet.

2.3 Fazit und Forschungsdesiderat

Während schon die kategorialen Beschreibungen von Füllpartikeln eine gewisse Variabilität zeigen, so erhöht sich die Variabilität weiter, wenn man ihre akustischen Realisierungsmöglichkeiten einbezieht. Füllpartikeln variieren in Segmentanzahl, Dauer, Grundfrequenz, Formanten, Phonationsart und Intensität. Dieses Variabilitätspotenzial wirft folgende Fragen auf: (i) Welche phonetischen Füllpartikelvarianten werden im Deutschen produziert? (ii) Handelt es sich um freie Variabilität oder ist sie von bestimmten Faktoren abhängig?

Ad (i): Dass es Variabilität in der Sprachproduktion gibt, ist bekannt. Die tatsächlich realisierte phonetische Form eines Lemmas hängt immer auch von der Variabilität spontaner gesprochener Sprache ab, welche je nach Äußerungssituation, sozialem Stand, Sprecherabsicht, Gesprächsthema oder idiosynkratischen Gewohnheiten andere Formen präferiert (einen Überblick geben bspw. Fiehler et al. 2004). Es gibt Hinweise, dass sich solche Faktoren mittel- oder unmittelbar auch auf die Füllpartikelrepräsentationen auswirken – vgl. Studien über Füllpartikeln zu Äußerungssituationen (Moniz et al. 2014; Rendle-Short 2004; Schachter et al. 1991), sozialem Stand (Fruehwald 2016; Laserna et al. 2014; Tottie 2011), Sprecherabsicht (Arciuli et al. 2010), Gesprächsthema (Bortfeld et al. 2001) oder idiosynkratischen Gewohn-

heiten (Braun & Rosin 2015; McDougall & Duckworth 2017).⁴⁹ Trotz der intensiven Beforschung von Füllpartikeln in den letzten Jahrzehnten wurde ihre phonetische Realisierung im Deutschen und auch in anderen Sprachen bislang nur ansatzweise betrachtet. Die Frage nach der realisierten Variabilität kann unter Berücksichtigung der oben genannten potenziell variabilitätsinduzierenden Faktoren nur für bestimmte Äußerungssituationen beantwortet werden, indem beispielsweise aufgabenfreie spontane Dialoge einer gewissen sozialen Schicht eines gewissen Alters mit einem bestimmten Startthema untersucht wird. Diese deskriptive Analyse der Füllpartikelrepräsentationen im Deutschen für spontansprachliche aufgabenfreie Dialoge ist ein Teilziel dieser Arbeit. Davon ausgehend können dann Generalisierungen abgeleitet werden.

Ad (ii): Wo die phonetische Variabilität bisher untersucht wird, beschränkt sich ihre Analyse oftmals auf über alle gefundenen Instanzen hinweg errechnete Mittelwerte. Der linguistische Kontext⁵⁰ und dessen Auswirkung auf die lautliche Gestalt von Füllpartikeln wird nur selten betrachtet – wenn doch, dann ist die Betrachtung oft nicht mit der in anderen Studien vergleichbar. Auf den ersten Blick frei verteilte linguistische Varianten können jedoch auf verschiedenen linguistischen Ebenen systematisch verteilt sein.⁵¹ Der sprachliche Ko- und Kontext eines linguistischen Phänomens beeinflusst seine sprachliche Gestalt – sowohl phonetisch also auch auf höheren linguistischen Ebenen. Phonetisch adjazente Laute werden unterschiedlich stark voneinander beeinflusst, da die Artikulation nicht kategorial-segmental, sondern kontinuierlich erfolgt.⁵² Diesen lautlichen Kontext nenne ich Mikrokontext oder sequenziellen Kontext. Den Kontext einer höheren linguistischen Ebene (Wort, Satz, Dialog) nenne ich Makrokontext oder hierarchischen Kontext. Diese Arbeit setzt sich daher zum Ziel, Füllpartikeln in ihren Mikro- und Makrokontexten zu untersuchen, da diese möglicherweise einen Teil ihrer auftretenden Varianten erklären können.

⁴⁹Der letzte Punkt macht Füllpartikeln interessant für die forensische Sprechererkennung (z. B. Hughes et al. 2016; Künzel 1987; Tschäpe et al. 2005).

⁵⁰*Kontext* verwende ich zur Beschreibung des unmittelbaren als auch des mittelbaren Kontextes. Der unmittelbare Kontext (auch *Kotext*) meint das unmittelbare vorangehende oder nachfolgende Wort beziehungsweise aus phonetischer Perspektive das unmittelbar vorangehende oder nachfolgende Segment. Der mittelbare Kontext besteht aus linguistischem Material höherer Ebenen (Morphologie, Syntax, Semantik, Pragmatik, Prosodie, Diskursstruktur).

⁵¹Vgl. die soziale Stratifizierung von /r/ in New Yorker Kaufhäusern (Labov 1972) oder Faktoren zur Erklärung der Genitivalternation im Englischen (Szmrecsanyi 2010).

⁵²Zur Koartikulation im Deutschen s. Kohler (1974), für einen Überblick über linguale Koartikulation s. Recasens (1999).

Tabelle 2.7 fasst die bisher vorgestellten Ergebnisse für deutsche Füllpartikeln aus Abschnitt 2.2 zusammen.

Tabelle 2.7: Zusammenfassung bisher bekannter akustischer Parameter für Füllpartikeln im Deutschen.

Parameter	Ausprägung	Kontextabhängigkeit
Segmentabfolge	V GV GVN VN N	–
Dauer	V ca. 270 ms, VN ca. 390 ms	–
Grundfrequenz	f_0 hoch, mittel, niedrig, Ton nicht bedeutungsunterscheidend	Tiefere f_0 als Umgebung, fallend
Vokalqualität/ Formanten	[[ə ɐ ø: ɪ ʊ œ]], F_1 / F_2 ♂ 500/1450 Hz, F_1 / F_2 ♀ 580/1550 Hz	–
Phonationsart	modal u. glottalisiert	–
Intensität	–	–

Berücksichtigt werden muss, dass die in Tabelle 2.7 dargestellten Varianten aus verschiedenen Äußerungssituationen stammen (beispielsweise die Lösung gemeinsamer Aufgaben vs. Interviews vs. Notrufe).⁵³ Weitere potenziell konfundierenden Faktoren für die in Tabelle 2.7 beschriebenen Merkmale von Füllpartikeln sind diatopische und diastratische Variationsmöglichkeiten. Während Äußerungssituationen experimentell kontrolliert werden können, ist der diatopische und diastratische Einfluss ein sprecherspezifisches Merkmal, welches im Idealfall konstant gehalten wird, wovon für die Studien, aus denen sich Tabelle 2.7 zusammensetzt, nicht ausgegangen werden kann. Im Folgenden werden die Variationsmöglichkeiten der akustischen Parameter und die daraus resultierenden Forschungsdesiderate nochmals kurz beschrieben.

1. **Segmentabfolge:** Welche Varianten existieren?

Einigkeit herrscht über die Existenz der Varianten V und VN. Offen ist indes, ob für die Epenthese des glottalen Plosivs zu Beginn von Füllpartikeln im Deutschen eine zugrundeliegende Systematik vorhanden ist und welche anderen phonetische Varianten als Füllpartikel zu klassifizieren sind.

⁵³Ob Füllpartikeln auch sprachliche Register charakterisieren, ist eine offene Frage, die ich hier nicht beantworten kann, weswegen ich den Begriff der Äußerungssituation verwende. Eine Registerperspektive kombiniert die Analyse linguistischer Charakteristika, die in einer bestimmten Textsorte [sic] häufig sind, mit einer Analyse der Gebrauchssituation der Textsorte (vgl. Biber & S. Conrad 2009: 2). Die linguistischen Charakteristika können hierbei jedwede lexiko-grammatischen Merkmale sein, die häufig in einer bestimmten Textsorte zu finden sind und somit eine kommunikative Funktion in diesem Register übernehmen (ebd.: 12).

2. **Dauer:** Mit welcher Dauer werden die Varianten produziert?

Die Dauer von Füllpartikeln ist sowohl sprach- als auch sprecherspezifisch – ist sie auch kontextspezifisch in einer bestimmten Äußerungssituation?

3. **Grundfrequenz:** Mit welcher Grundfrequenz werden Füllpartikeln produziert?

Die bisherige Forschung beschreibt die Tendenz zur intonatorischen Anpassung an die Umgebungskontur.

4. **Vokalqualität:** Welcher Vokal wird produziert?

Die Vokalqualität schwankt erheblich. Unklar ist, ob die Vokalrealisierung frei oder systematisch variiert.

5. **Phonation:** Welche Stimmqualität wird produziert?

Glottalisierte Füllpartikeln unterscheiden sich perceptiv von modalen und behauchten Varianten. Bisher ist nur wenig über ihre Verwendung und Variabilität bekannt.

6. **Intensität:** Mit welcher Intensität werden Füllpartikeln produziert?

Hierüber ist bisher wenig bekannt.

Aus den Sätzen (i)–(ii) leitet sich die in dieser Arbeit untersuchte Hypothese der Formvariabilität ab, nämlich dass Füllpartikeln phonetisch unterschiedliche segmentale und suprasegmentale Formen annehmen können. In einem zweiten Schritt wird dann die Hypothese aufgestellt, dass einige dieser Formen systematisch in bestimmten linguistischen Kontexten auftreten (vgl. Kapitel 3). Dieses kleinteilige *Bottom-up*-Vorgehen ist die Voraussetzung für nachfolgende Studien, die dann wiederum prüfen können, ob die gefundenen Form-Kontext-Muster mit bestimmten linguistischen Funktionen korrelieren.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





3 Funktionen und Kontexte

In diesem Kapitel erörtere ich in Kürze die Funktionen und anschließend zur Motivation des weiteren Vorgehens die Kontexte von Füllpartikeln. Abschnitt 3.1 zeigt die Multifunktionalität von Füllpartikeln anhand fünf häufig genannter Funktionen und stellt die Haltung dieser Arbeit zur Signal-Symptom-Debatte dar. Die postulierten Funktionen entstehen fast ausschließlich aus dem linguistischen Kontext von Füllpartikeln. Konsequenterweise müssen Funktionen daher als Interpretation der Kontexte hintan gestellt werden. Abschnitt 3.2 beschreibt folglich die bisher beschriebenen Kontexte und, falls vorhanden, ihre Interaktion mit der phonetischen Realisierung von Füllpartikeln, woraus schließlich die Form-Kontext-Hypothese entwickelt wird. Ferner wird in Abschnitt 3.3 ein Überblick über sprecher- und situationsspezifische Distributionen und Formen von Füllpartikeln gegeben.

3.1 Multifunktionalität von Füllpartikeln

3.1.1 Symptom versus Signal

Bevor ich die mutmaßlichen Funktionen von Füllpartikeln diskutiere, werde ich zur Einordnung der von dieser Arbeit eingenommenen Perspektive die Debatte um die beiden am häufigsten in der Literatur diskutierten Hypothesen zur Produktion von Füllpartikeln, die Symptomhypothese und die Signalthypothese, in ihren wesentlichen Zügen rekapitulieren. Dies ist notwendig, um die hier vorliegende Arbeit von diesen Hypothesen freizumachen.

Im Rahmen der Symptomhypothese werden Füllpartikeln als Epiphänomene der Sprachproduktion betrachtet und häufig als *Häsitation* oder *gefüllte Pause* bezeichnet. Ursächlich für ihre Verwendung seien Planungsansprüche an die Sprachproduktion (Beattie & Butterworth 1979; Beattie & Shovelton 2002; Blackmer & Mitton 1991; Corley & Stewart 2008; Finlayson & Corley 2012; Lay & Paivio 1969; Levelt

1983, 1989; Maclay & Osgood 1959; Martin & Strange 1968; Oviatt 1995; Rochester 1973). Neben dieser psycholinguistischen Erklärung haben Schönle & B. Conrad (1985) versucht, eine physiologische Erklärung zu geben. Sie untersuchen 16 deutsche Muttersprachler (im Mittel 26,4 Jahre alt), die fünf Minuten Spontansprache produzieren und deren Atmung mithilfe eines Atmungsmessgürtels aufgezeichnet wird, um zu ermitteln, an welchen Stellen ihrer Ausatemungsphase Füllpartikeln auftreten, wobei die Ausatemungsphase in vier äquidistante Bereiche aufgeteilt wird.⁵⁴ 62,5 % aller Füllpartikelvokale treten im ersten Bereich auf, 20,8 % im zweiten, 2,1 % im dritten und 14,6 % im vierten (ebd.: 294f.). Aus diesen Ergebnissen schließen Schönle & B. Conrad, dass „hesitation vowels“ als phonomotorische Subroutinen des Sprachproduktionssystems verwendet werden, um für fehlendes Sprachmaterial zu kompensieren und das respiratorische System vor unökonomischem Luftverlust zu bewahren. Obwohl beide Erklärungen, die psycholinguistische und die physiologische, Ähnlichkeiten hinsichtlich ihres symptomatischen Charakters aufweisen, sind sie bisher nicht zusammen betrachtet worden.

Im Rahmen der Signalthypothese wird argumentiert, dass Füllpartikeln ein intentionales Signal an einen Hörer oder eine Hörerin darstellen (Clark & Fox Tree 2002; Eklund 2004; Goodwin 1981; Jefferson 1974; Smith & Clark 1993; E. J. Walker et al. 2014; Willkop 1988). Beispiel für ein solches Signal sei die Verwendung (im Englischen) für „comments on the on-going performance“ (Clark & Fox Tree 2002: 107) oder die Verwendung als Turnhalte- oder Turnabgabesignal („I want to keep the floor“ und „I want to give up the floor“), die Verwendung als Unsicherheitssignal („I’m uncertain about what I want to say“) oder zur Anzeige von Wortfindungsanstrengungen („I’m hunting for the next word“) (ebd.: 104). In Abschnitt 3.1.2 werden diese Funktionen eingehender betrachtet. Hier scheint schon die Schwierigkeit einer klaren phänomenalen und epistemologischen Trennung durch; der gleiche Sachverhalt wird mit subtilen terminologischen Unterschieden von der signalthypothetischen Seite als „hunting for the next word“, von der symptomhypothetischen Seite hingegen als Häsitiation analysiert. Auf der einen Seite steht die aktive, signalisierende Interpretation des Phänomens im Vordergrund, auf der anderen Seite die passive, durch Schwierigkeiten in der Sprachplanung entstandene Verzögerung.

Ein vorgeschlagener Kompromiss für das Dilemma von Symptom und Signal be-

⁵⁴Fünf der Probanden verwenden keine Füllpartikeln (Schönle & B. Conrad 1985: 294).

steht darin, Füllpartikeln als Reflex der Sprachplanung zu betrachten, die dann (quasi indexikalisch)⁵⁵ von der Hörerseite verarbeitet werden (Arnold et al. 2003, 2004; Brennan & Williams 1995; Corley et al. 2007; Ferreira & Bailey 2004; Finlayson & Corley 2012; Watanabe et al. 2008). Für das Deutsche hat Keseling (1989) versucht, die Symptom- und die Signalperspektive zu vereinen.

Wenn es zutrifft, dass ÄH vor einer wichtigen Aussage oder einem wichtigen Wort plaziert ist und wenn es zugleich an Stellen auftritt, die vom Sprecher (verstärktes) Nachdenken erfordern, dann scheint die Hauptfunktion dieses Lautgebildes in der Koordination von planerischem Nachdenken auf der Sprecherseite und (erhöhter) Aufmerksamkeit auf der Hörerseite zu bestehen (ebd.: 589).

In diesem Sinne schreibt auch Reitbrecht (2017) Füllpartikeln „Sprachverarbeitung stützende Funktionen“ zu und folgert, dass der Ansatz, Füllpartikeln jeweils apodiktisch mittels einer einzigen dieser beiden funktionalen Hypothesen erklären zu wollen, scheitern muss. Reitbrecht nimmt die „sprechplanerische, kognitive Funktion“ als „potenziell gegeben [Im Original kursiv – M. B.]“ an, und versteht die Symptom- und die Signalthypothese als „potenziell simultan verlaufende Komplementärfunktionen [Im Original kursiv – M. B.] [...], die einander nicht ausschließen“ (ebd.: 27).

Ich schließe mich hier der Sicht von Reitbrecht an, da der Beitrag dieser Arbeit nicht die Lösung der Symptom-Signal-Debatte anstrebt. Vielmehr möchte ich eine gebrauchsbasierte Erklärung der verwendeten phonetischen Formen von Füllpartikeln in ihren linguistischen Produktionskontexten geben. Ich nehme an, dass Form-Kontext-Muster existieren (vgl. Abschnitt 3.2.1), die dann in zukünftigen Arbeiten zu den oben beschriebenen Erklärungsansätzen weiter diskutiert werden können. Bevor diese Hypothese jedoch genauer ausgeführt wird, beschreibe ich im nächsten Abschnitt einige Funktionen, die Füllpartikeln zugeschrieben werden. Die Unauflösbarkeit der verschiedenen *Top-Down*-Funktionszuschreibungen und ihre Rückführung auf die Auftretenskontexte motiviert die weitere Betrachtungsweise im Verlaufe dieser Arbeit.

⁵⁵Fischer (1999) votiert sogar für einen ikonischen Charakter von Füllpartikeln.

3.1.2 Funktionen

Ohne also weiter auf die Symptom-Signal-Debatte einzugehen, bespreche ich hier in gebotener Kürze einige häufig identifizierte sprachliche Funktionen von Füllpartikeln. Im linguistischen Sinne ist der Begriff ‚Funktion‘ polysem und kann je nach linguistischer Disziplin variieren (beispielsweise „Intonation [als] satztypendifferenzierende F[unktion]“ oder kommunikative Aspekte in der linguistischen Pragmatik, vgl. Glück 2005: 206). Hier verstehe und verwende ich den Begriff der Funktion zunächst bewusst grob als „interne oder externe Beziehung zwischen ling[uiistischen] Sachverhalten“ (ebd.).

Kjellmer (2003) votiert für die Multifunktionalität von Füllpartikeln im Englischen und gegen eine rein häsitatorische Funktion, denn „[h]esitation may not always be involved“ (ebd.: 171). Als Argument für die Multifunktionalität gibt Kjellmer Beispiele, dass Füllpartikeln *thought units* (‚Denkeinheiten‘) auf Wort-, Phrasen- und Satzebene einführen (ebd.: 174-181), Redewechsel steuern (ebd.: 186), die Salienz nachfolgender Wörter erhöhen (ebd.: 187), als Korrektureinheiten dienen (ebd.: 188f.), und als strukturelle Einheiten verwendet werden können (ebd.: 190).

Im Kern sind dies also fünf Funktionen, welche auch in der englischen und deutschen Literatur (vgl. unten) immer wieder genannt werden, nämlich die Verwendung als Verzögerungs- beziehungsweise Häsitationsmarker (*f.hes*), die Verwendung als redewechselbezogener Marker beziehungsweise Turnmarker (*f.turn*), die Verwendung als Reparaturmarker (*f.rep*), die Verwendung als prominenz- beziehungsweise salienzbezogener Marker (*f.sal*) und die Verwendung als struktureller beziehungsweise einführender Marker, was sich in den Grenzen der Dialogbetrachtung (vgl. Abschnitt 3.2.3.3) als dialogstrukturelle Verwendung (*f.dia*) auffassen lässt. Diese Funktionen stehen in keiner hierarchischen Abhängigkeit zueinander. Während die Funktionen *f.turn* und *f.dia* auf einer dialogischen Struktur aufbauen, können *f.hes*, *f.sal* und *f.rep* sowohl in Dialogen als auch in Monologen auftreten.

Im Folgenden wird jede dieser Funktionen aus Erwähnungen in der Literatur zum Englischen und Deutschen (sofern vorhanden) heraus näher beschrieben, bevor nochmals zwei deutschsprachige Studien skizziert werden, welche die fünf Funktionen in ähnlicher Weise fürs Deutsche postulieren. Generell ist fraglich, ob die Ergebnisse der Füllpartikelstudien im Englischen auf die Verwendung von Füllpartikeln im Deutschen eins zu eins übertragen werden können. Des weiteren ist die Ermittlung

einer Funktion keineswegs eindeutig, sondern kann in hohem Maße mit anderen Funktionen überlappen. Dies ist einerseits der Vielfalt der verwendeten Methoden geschuldet, die von psycholinguistischen und konversationsanalytischen bis hin zu korpuslinguistischen Ansätzen reichen, andererseits fehlt auch eine Entscheidungshilfe. Obwohl die funktionale Analyse in vielen Fällen durch ihren Kontext geleitet werden kann, bleibt sie hochsubjektiv.

Funktion *f.hes* – Verzögerungsmarker: Die Verwendung von Füllpartikeln als Häsitations- oder Verzögerungsmarker (*f.hes*) lässt sich größtenteils aus den Untersuchungen zu Füllpartikeln als Symptom ableiten (Blackmer & Mitton 1991; Clark & Fox Tree 2002; Corley & Stewart 2008; Finlayson & Corley 2012; Goldman-Eisler 1961; Lay & Paivio 1969; Levelt 1989; Lickley 2015; Martin & Strange 1968; Rochester 1973). Füllpartikeln in dieser Funktion werden demnach produziert, wenn die Belastung durch die Sprachplanung zu groß wird („when speech is interrupted at or close to trouble“, Levelt 1983: 74). Dies kann beispielsweise durch den erhöhten Schwierigkeitsgrad einer zu lösenden Aufgabe induziert werden (vgl. Rochester 1973: 70). So produzierten 24 männliche englischsprachige Versuchspersonen Füllpartikeln bei Selbstbeschreibungen in 2,8 % aller Wörter, bei Cartoonbeschreibungen in 5,7 % und bei der Erklärung von Sprichwörtern in 9 % aller Wörter (Lay & Paivio 1969: 31). *f.hes* ist quasi der Prototyp für die funktionale Verwendung von Füllpartikeln. Die Funktion *f.hes* basiert auf Annahmen über die Sprachplanung eines Sprechers oder einer Sprecherin. Es lässt sich in der Literatur keine Präferenz für eine spezifische Position innerhalb von linguistischen Kontexten feststellen.

Funktion *f.rep* – Reparaturmarker: Diese Funktion beschreibt die Verwendung von Füllpartikeln in Verbindung mit Reparaturen oder Neustarts (Belz 2013; Belz et al. 2017; Heeman & Allen 1999; Levelt 1983; Seyfeddinipur et al. 2008; Shriberg 1994). Auf den ersten Blick ist die Funktion einer Füllpartikelentität relativ einfach zuweisbar, steht sie doch innerhalb einer Reparatur, deren Struktur bekannt ist (vgl. bspw. Shriberg 1994). Prinzipiell wäre jedoch auch die Zuschreibung von *f.hes* zu solch einer Füllpartikel innerhalb einer Reparatur denkbar – dass Füllpartikeln innerhalb von Reparaturen die Funktion *f.rep* erhalten, ist also vor allem durch den gut definierbaren Kontext und das Auftreten in eben diesem bedingt. Während – vielleicht auch aufgrund dieser nicht uneindeutigen Zuweisbarkeit – über die Häufigkeit von *f.hes* bisher keine Aussagen getroffen worden sind (es fehlt

eine funktionsbasierte Korpusstudie), kann die Häufigkeit von *f.rep* gezielt ermittelt werden. Beispielsweise finden sich in BeMaTaC⁵⁶ (Sauer & Lüdeling 2016), einem Map-Task-Korpus mit 12 aufgabenbasierten Dialogen 16 deutscher Muttersprachlerinnen und Muttersprachler (11 199 Token), insgesamt 210 Reparaturen⁵⁷ (= 1,9 % aller Token), in denen wiederum 29 Füllpartikeln⁵⁸ enthalten sind. Bei insgesamt 274 Füllpartikeln⁵⁹ (*äh* und *ähm*) beträgt demnach der Anteil von Füllpartikeln mit der Funktion *f.rep* in diesem Korpus 10,6 %.

Funktion *f.sal* – Salienz und Informationsstatus: Diese Funktion beschreibt die Verwendung von Füllpartikeln als Signal für etwas Neues, zur Hervorhebung und zur Herstellung von Aufmerksamkeit. Der Status von Füllpartikeln im Zusammenhang mit Informationsstruktur⁶⁰ ist bislang noch umstritten, wenn auch die Hypothese, dass Disfluencies (und in Folge Füllpartikeln) im Allgemeinen eher vor nicht-orerwähnten Referenten vorkommen, schon lange im Raum steht (Maclay & Osgood 1959: 34). Auch wurde in psycholinguistischen Eye-Tracking-Studien für die englische Sprache festgestellt, dass Füllpartikeln häufiger vor neuer Information als vor gegebener Information produziert und erwartet werden (Arnold et al. 2003; Arnold & Tanenhaus 2011; Arnold et al. 2004). Für das Deutsche lässt sich mithilfe des eben schon herangezogenen Korpus BeMaTaC zeigen, dass etwa die Hälfte, nämlich 15 der 29 Füllpartikeln in Reparaturen, vor einem „neuen“, im linken Teil der Reparatur nicht vorerwähnten Token stehen.⁶¹ Faure (1980: 290) hingegen findet keinen Zusammenhang zwischen Thema-Rhema-Struktur und Füllpartikeln in französischen und deutschen Narrativen, Diskussionen und spontansprachlichen

⁵⁶Berlin Map Task Corpus v. 3.0, <http://u.hu-berlin.de/bematac>, besucht am 20.05.2019.

⁵⁷Da Reparaturen symmetrisch aus Reparandum und Reparas bestehen, werden für diesen Zweck nur die Reparaturanfänge (Reparanda) herangezogen, vgl. <https://korpling.org/annis3/?id=bb9aaf91-630c-490f-9ecd-3f82557ec7b6>, besucht am 24.04.2019.

⁵⁸<https://korpling.org/annis3/?id=4e88a086-da66-42e3-ad72-7f844bf2ceec>, besucht am 24.04.2019.

⁵⁹<https://korpling.org/annis3/?id=56f9b849-fd67-40aa-8519-bd4090516b22>, besucht am 24.04.2019.

⁶⁰Informationsstruktur untersucht „die Art, wie Sätze an den Kontext angepasst werden hinsichtlich Dimensionen wie Bekanntheit/Unbekanntheit, Topik/Kommentar und Fokus/Hintergrund“ (Musan 2010: 3; vgl. auch Krifka & Musan 2012). Diese Dimensionen beziehen sich dabei auf verschiedentlich große Konstituenten und werden besonders im Kontext geschriebener Sprache angewendet.

⁶¹Da der rechte Teil einer Reparatur auch aus mehreren Token bestehen kann, wird aufgrund der grammatischen Struktur des Deutschen das letzte Token in der Abfolge auf seine Vorerwähtheit hin evaluiert, <https://korpling.org/annis3/?id=9c932fe4-551c-4e2e-84a5-716c1166af0d>, besucht am 20.05.2019.

Konversationen.

Ein weiterer Aspekt von *f.sal* ist das Herstellen von Aufmerksamkeit, wie es (anekdotisch) beim Ansprechen eines Menschen verwendet werden kann, beispielsweise in *äh Entschuldigung*. Tatsächlich treten hier keine starken Belastungen durch die Sprachplanung auf, wie sie für *f.hes* postuliert werden. Dennoch überschneidet sich die Funktion mit *f.turn* und *f.dia* (s. unten). Ein ähnliches Beispiel für Aufmerksamkeitserheischung gibt Lickley (2015) für das Englische mit „*um, excuse me!*“ (ebd.: 463) und erwähnt auch den steigenden Tonhöhenakzent auf *um*, der als Intention gemeint sein kann, dass ein Sprecher etwas sagen möchte. Das ist mit einiger Wahrscheinlichkeit aber nicht allein das Verdienst des *um* mit steigendem Ton, sondern sicherlich auch induziert von dem Fragekontext, der oft mit steigendem Ton belegt ist.

Funktion *f.turn* – Turn-Marker: In der Diskussion ihrer Studie von 163 Äußerungen 13 englischsprachiger männlicher Sprecher auf einer wissenschaftlichen Konferenz schlagen Maclay & Osgood (1959: 41) vor, dass der Hauptunterschied zwischen stiller und gefüllter Pause in der Dauer des Intervalls liegt, in dem nicht gesprochen wird. Sie nehmen an, dass der momentane Sprecher solange den „conversation ‚ball““ behalten möchte, „until he has achieved some sense of completion“ (ebd.). Da ungefüllte Intervalle einer gewissen Länge zum Verlust dieses Rederechts führen könnten, verwendet der Sprecher neben anderen Strategien eine Füllpartikel (ebd.). Seit dieser Studie wird Füllpartikeln eine turnmarkierende Funktion zugeschrieben. Der Begriff *Turn* wird in der Konversationsanalyse verwendet und beschreibt einen Redebeitrag (Levinson 2000: 323) im Kontext des Sprecherwechsels (Sacks et al. 1974; Schegloff et al. 1977), wobei jedoch in den von mir rezipierten Arbeiten keine genauere Definition über die Ausdehnung des Begriff zu finden war. Dass 56 % aller Füllpartikeln im London-Lund-Corpus (Svartvik 1990) am Anfang einer Intonationsphrase stehen, interpretieren Clark & Fox Tree als turnmarkierende Verwendung von Füllpartikeln (Clark & Fox Tree 2002: 94). Kritischerweise ist hinzuzufügen, dass zwar jeder neue Turn mit dem Beginn einer Intonationsphrase zusammenfällt, der Beginn eine Intonationsphrase jedoch nicht immer einen neuen Turn markiert.

Lallge & Cook (1969) finden keine Indizien für die von Maclay & Osgood (1959) postulierte *Turn-holding*-Funktion von Füllpartikeln, da sie in Interviews mit 14 englischen Muttersprachlern mit Interviewern, die wenig oder viel Druck auf das

Rederecht des Interviewten erzeugten, keine signifikanten Unterschiede in der Füllpartikelanzahl finden (Lalljee & Cook 1969: 27). Aufgrund des hohen Drucks auf das Rederecht wurde erwartet, dass die Interviewten häufiger Füllpartikeln nutzen, um den Turn zu halten. Möglicherweise nutzen sie aber andere Strategien. In einer weiteren Studie elizitieren Lalljee & Cook (1974) Füllpartikeln von 30 männlichen Universitätsstudenten des Englischen unter drei Bedingungen. In der ersten Bedingung monologisiert der Proband zu Themen, die er vor sich auf Karten notiert findet. In der zweiten Bedingung unterhält er sich mit einem Interviewer über die gleichen Themen in der gleichen Reihenfolge. Schließlich diskutiert der Proband die Themen nochmals mit dem Interviewer in einer Art freiem Gespräch. Die Anzahl der produzierten Füllpartikeln je Bedingung unterscheidet sich nicht signifikant (Monolog 1,81 %, Interview 3,67 %, Konversation 2,83 %), woraus die Autoren schließen, dass in diesem Experiment keine Evidenz für die *Turn-holding*-Hypothese gefunden wird. Allerdings wird vorausgesetzt, dass die Verwendung im Sinne von *f.turn* sich auf die Produktion von Füllpartikeln überdurchschnittlich stark auswirkt, da Füllpartikeln auch mit anderen Funktionen verwendet werden können. Mangels einer detaillierten Analyse ist somit nicht klar, was die Häufigkeitsergebnisse aussagen. Lalljee & Cook (ebd.) schlagen als Alternativerklärung zur *Turn-holding*-Funktion *f.turn* Aufmerksamkeitserheischung („attention-getting“) vor, damit der frühe Teil der zu übermittelnden Message nicht verloren geht (ebd.: 224). Dies entspräche der Funktion *f.sal*. Tatsächlich wird die Multifunktionalität von Füllpartikeln und ihre mögliche Auswirkung auf die phonetische Realisierung bei Lalljee & Cook (ebd.) schon angedacht:

Perhaps filled pauses performing different functions are different in certain other respects. Some means could be sought to distinguish between various kinds of filled pauses, and grammatical location and extra-linguistic factors such as pitch or volume may be relevant criteria. (ebd.: 224)

Auch in einer rezenteren Studie im Englischen wird kein Indiz für die Verwendung von Füllpartikeln im Sinne von *f.turn* erbracht. Tottie (2015) analysiert das erste Viertel des *Santa Barbara Corpus of American English* (SBC1) mit 62 330 Wörtern aus 14 verschiedenen Konversationen von 20–25 Minuten Länge. Zwar kommen Füllpartikeln im SBC1 am häufigsten in turn-medialer Position vor (73 %), jedoch gibt es kaum Fälle in diesen Instanzen, in denen andere Teilnehmer den Turn ergei-

fen möchten (sofern sie dies akustisch angezeigt haben und dies auch transkribiert wurde). Somit dienen die Füllpartikeln in turn-medialer Position eher der Planung des verbleibenden Turns (vgl. Tottie 2015: 402–403). Am zweithäufigsten stehen Füllpartikeln in turn-initialer Position (15%), am dritthäufigsten in turn-finaler Position (6%) und in turn-zweiter Position (6%) (ebd.: 393). Sie schlägt als ergänzende Funktion vor, dass man in turn-finalen Positionen signalisiert, nicht weitersprechen zu wollen („diminished intent to continue“) (ebd.: 392). 38% der Füllpartikeln in turn-initialer Position kommen in Antworten vor (ebd.: 394), was nahelegt, die Dialogstruktur als erklärende Variable genauer zu betrachten. Zusammenfassend wird in der Füllpartikelforschung zwar immer wieder auf die Funktion *f.turn* rekurriert, ihr Status ist aber weniger gut belegt als die Literatur suggeriert.

Funktion von *f.dia* – Dialogstruktur: Neben der Verwendung von Füllpartikeln zu Beginn, am Ende oder innerhalb von dialogstrukturierenden Einheiten (Dialogzüge, vgl. Abschnitt 3.2.3.3) treten Füllpartikeln im Deutschen auch „unmittelbar nach der Begrüßung oder nach der Selbstidentifikation auf, vor dem Einstieg in das Thema“ (Rasoloson 1994: 61f.). Dies deckt sich mit der Beobachtung, dass ein *uhm* zu Beginn eines Telefonanrufes als strukturierendes Mittel steht, „which gives no indication here of uncertainty, reluctance, hesitation, trouble, or the like“ (Schegloff 2010: 147f.). Somit gibt es eine Schnittmenge zwischen *f.dia* und *f.sal*. In der Funktion *f.dia* müssen Füllpartikeln jedoch nicht zwangsläufig Aufmerksamkeit erregen, da in in Frage-Antwort-Paarsequenzen die Aufmerksamkeit ja schon hergestellt wurde – analog muss hier auch kein Turn beansprucht oder gehalten werden (*f.turn*). Dennoch stehen Füllpartikeln in englischen aufgabenbasierten Dialogen häufiger vor Antworten als vor anderen dialogstrukturellen Einheiten (Lickley 2001). Für das amerikanische Englisch liegt in ähnlichen Dialogen keine ausreichende Bestätigung vor (Nicholson et al. 2010). Auf beide Studien gehe ich in Abschnitt 3.2.3.3 näher ein. Im Deutschen dienen Füllpartikeln laut Schwitalla (2002) „dem prospektiven Sprecher zur Redeplanung und signalisieren, dass er/sie bereit ist, Antwortobligation zu übernehmen“ (ebd.: 274). Die funktionale Beschreibung von Füllpartikeln als „attention-getting signal“ (Lalljee & Cook 1974: 224), was vor dem neuen Sprechensatz stehen kann⁶², kann sowohl dialogstrukturierend als *f.dia* wie auch als *f.sal* oder als *f.turn* interpretiert werden.

⁶² „A person may say ‚er‘ before he begins speaking in order to draw attention to himself either to signal that he would like to speak or that he is about to speak.“ (Lalljee & Cook 1974: 224).

Nachdem die oben aufgeführten fünf Funktionen vornehmlich englischsprachige Studien wiedergeben, seien kurz die zwei deutschen konversationsanalytischen Studien erwähnt, die maßgeblich mit einer eigenen Analyse zu ähnlichen Schlussfolgerungen gekommen sind. Willkop (1988) untersucht Füllpartikeln vornehmlich in Alltagsgesprächen⁶³ und schreibt Füllpartikeln eine „primär reorganisierende Funktion“ zu; sie können in diesem Sinne Unterbrechungen seitens des Gegenübers verhindern, Formulierungsschwierigkeiten überbrücken, eine Turn-Übernahme anmelden und seien ein „Signal für eine nicht fremdinitiierte Eigenkorrektur“ (ebd.: 251). Rasoloson (1994) untersucht Füllpartikeln in 132 deutschen „Informationsdiskursen“ (Alltagssituationen, in welchen eine Information eingeholt wird) zwischen 1975 und 1986 (ebd.: 28 f.). Rasoloson (ebd.: 79) analysiert Füllpartikeln als Steuerungsmittel (Bearbeitung des Turn-/Musterabbruchs, Anspruch auf Sprecherrolle), als Korrekturausdruck, als Gliederungsmittel (Ankündigung eines neuen Inhaltsabschnitts, Markierung der Verteilung der Sprecher- und Hörerrolle), als deliberatives Ausdrucksmittel⁶⁴ (zur Lexemsuche und (Teil)Satzbearbeitung, vor Informationsfrage und vor positiver Antwort) und als Indikator für einen Erwartungsbruch (vor negativer Antwort).

In Tabelle 3.1 fasse ich die oben zusammengefassten meistbeschriebenen⁶⁵ Funktionen verkürzt unter ihren Abkürzungen zusammen. Die fünf dort beschriebenen Funktionen sind jedoch nicht immer disjunkt zueinander, im Gegenteil, sie haben große Schnittmengen. Des weiteren werden sie methodisch unterschiedlich übermittelt. Indikatoren für die Feststellung einer Funktion sind hauptsächlich die linguistischen Kontexte, in denen sie auftreten. So kann eine Füllpartikel vor einer Frage auch als Füllpartikel vor Turn-Beginn klassifiziert werden (*f.turn* und *f.dia*), wenn an dieser Stelle ein Turn beginnt. Zusammenfassend sind die Funktionen von Füllpartikeln im Sinne der von Fedriani & Sansò (2017) entwickelten Terminologie zur Beschreibung von diskursmarkierenden Einheiten prozedural (*f.hes*), indexikal (*f.turn*, *f.sal*, *f.dia*) und metalinguistisch (*f.rep*). Die Funktionen *f.turn* und *f.dia* beschreiben

⁶³Mindestens zwei Alltagsgespräche, ergänzt mit Telefondialogen und verschiedenen weiteren Daten. Der Gesamtumfang der Daten war nicht auszumachen (vgl. Willkop 1988: 14 f.).

⁶⁴Zum Verwendung des Terminus *deliberativ* im Sinne von *überlegt*, *bedacht*, *erwogen* in diesem Kontext vgl. „a deliberate signal chosen from a range of interjections“ (Smith & Clark 1993: 27) und „als deliberatives Ausdrucksmittel“ (Rasoloson 1994: 77).

⁶⁵Weitere Funktionen sind nicht ausgeschlossen. Weniger häufig erwähnt wird beispielsweise die Verwendung von *uh* im Englischen als pragmatischer Markierer zur Zurückweisung einer Implikatur (Norrick 2009: 875).

Funktionen, die besonders in interaktionalen Situationen auftreten. Die Funktionen *f.hes*, *f.rep* und *f.sal* können sowohl in monologischen als auch dialogischen Situationen auftreten.

Tabelle 3.1: Nichtexhaustive Zusammenstellung und Beschreibung der meistangenommenen Funktionen von Füllpartikeln in bestimmten Positionen, wie sie in der Literatur postuliert werden. Die Funktionen sind nicht hierarchisch organisiert.

Abkürzung	Funktion	Position
<i>f.hes</i>	Markiert Verzögerung	Häufige Interpretation einer Füllpartikel, wenn keine spezifischere Funktion zutrifft.
<i>f.rep</i>	Markiert Reparaturen	Steht innerhalb von Reparaturen.
<i>f.turn</i>	Markiert Turns	Steht am Anfang (Redeübernahme), in der Mitte (Erhalt des Redebeitrags) oder am Ende eines Turns (Abgabe oder Erhalt des Redebeitrags).
<i>f.sal</i>	Markiert Aufmerksamkeit, Saliens oder Informationsstatus	Steht zu Beginn der Gesprächseröffnung, vor neuen Inhaltsabschnitten, vor neuen Referenten.
<i>f.dia</i>	Markiert Dialogstruktur	Steht vor Fragen, vor Antworten, innerhalb von Narrativen.

Dieses Kapitel und Tabelle 3.1 zeigen, dass eine Funktionsbestimmung von Füllpartikeln nicht losgelöst von den sie umgebenden Kontexten durchgeführt werden kann. Tatsächlich ergeben sich die herausgearbeiteten Funktionen in den meisten Fällen gerade erst durch die Berücksichtigung ihres Kontextes, oder anders formuliert, der Kontext einer Füllpartikel stellt das definitorische Kriterium für ihre funktionale Interpretation. Die Funktion *f.rep* wird durch den Reparaturkontext, *f.turn* durch den Turnkontext, *f.dia* durch den diskursstrukturellen Kontext und *f.sal* durch den informationsstrukturellen Kontext definiert. Falls sich keine der spezifischeren Funktionen ableiten lässt, bleibt immer noch die Funktion *f.hes*. Eine spezifische Ausprägung einer Füllpartikel kann dabei auch mehrere Funktionen erfüllen. Nachdem beispielsweise ein Turn von Gesprächspartner A mit einer Frage beendet wurde, kann eine Füllpartikel vor der Antwort des Gesprächspartners B sowohl als Turnfunktion (*f.turn*) als auch als dialogstrukturierende Funktion (*f.dia*) interpretiert werden – mangels klarer ‚Beweise‘ muss sie es sogar; sie könnte aber natürlich auch die Funktion *f.hes* erhalten. Wenn – als zweites Beispiel – eine Füllpartikel innerhalb eines ansonsten flüssigen Redebeitrags vorkommt, so wird ihr oftmals (mangels anderer Indizien) die Verwendung als *f.hes* zugeschrieben. Findet sich die Füllpartikel allerdings innerhalb einer Reparatur, so erhält sie in dem

vorherrschenden *top-down*-funktionsbasierten Ansatz die Funktion *f.rep.* Dies findet schon Shriberg (1994) problematisch:

An alternative solution (which represents the implicit view in some past accounts) proposes that filled pauses that occur in the IM [interregnum – M.B.] of a longer DF [disfluency – M.B.] function as editing phrases, whereas those occurring elsewhere function as hesitations. There is to date no hard evidence for this view. (ebd.: 64)

Zusammenfassend wurde gezeigt, dass sich Funktionszuschreibungen für Füllpartikeln *nolens volens* an dem Kontext der Füllpartikel orientieren und dass die Festlegung einer bestimmten Funktion hochgradig interpretativ, methodisch divers und subjektiv ist. Ich werde daher davon Abstand nehmen, Funktionen für Füllpartikeln *top down* oder auf andere Weise zuzuweisen. Im folgenden Abschnitt stelle ich die Form-Kontext-Hypothese vor und diskutiere, wie sich dem Desiderat der fehlenden „hard evidence“ (ebd.) für Funktionszuschreibungen über den Umweg der kontextbasierten Analyse angenähert werden kann.

3.2 Kontext und phonetische Realisierung

In diesem Abschnitt stelle ich die Form-Kontext-Hypothese auf und erläutere sie mit Beispielen und einem Literaturüberblick.

3.2.1 Die Form-Kontext-Hypothese

Only when we look at the way speakers employ different phonetic means to organize the most common form of spontaneous speech, conversation, do patterns occur which are exclusive to spontaneous speech. (Simpson 2013: 169)

Simpson betont die immanente phonetische Relevanz für die Beschreibung des Forschungsgegenstands der spontanen gesprochenen Sprache und ihrer verwendeten Gebrauchsmuster. Ohne eine phonetische Beschreibungsebene bleibt die Beschreibung gesprochener Sprache und der in ihr von Sprecher*innen verwendeten Phänomene defektiv. Gerade für die Betrachtung genuin gesprochensprachlicher Phänomene, wie sie Füllpartikeln darstellen, können mithilfe phonetischer Merkmale bisher unbekannte Verwendungsmuster beschrieben, erklärt und eingeordnet werden.

Mit der Form-Kontext-Hypothese bezeichne ich die Vermutung, dass es eine systematische Beziehung zwischen phonetischer Realisierungsform und linguistischem Kontext dergestalt gibt, dass die Variabilität der verwendeten phonetischen Füllpartikelformen mit ihren linguistischen Auftretenskontexten korreliert. Mit Kontexten bezeichne ich die linguistischen Auftretenskontexte von Füllpartikeln – diese können sowohl sequenzieller als auch hierarchischer Art sein, also beispielsweise die die Füllpartikel umgebenden, adjazenten Lautsegmente (sequenziell) oder das Vorkommen in einer Antwort (hierarchisch). Eine Füllpartikel nimmt in ihrem Kontext dann eine bestimmte Position ein – sequenziell zwischen Segmenten, Lauten oder Wörtern, oder hierarchisch an den Grenzen oder innerhalb von Phrasen oder höheren Einheiten.

Wie in Abschnitt 3.1.2 erörtert, dient in den meisten Fällen der Auftretenskontext beziehungsweise die Position einer Füllpartikel als Indiz für ihre unterstellte Funktion. Weitere Evidenzen für diese Annahme fehlen bisher jedoch. Fraglich bleibt zudem, inwiefern Füllpartikeln in den verschiedenen Kontexten dann *tatsächlich* die ihnen zugeschriebenen Funktionen übernehmen. Es bleibt offen, ob dies überhaupt mit Sicherheit beantwortet werden kann. Daher besteht hier ein Desiderat der Beschreibung kontextueller Verwendungen von Füllpartikeln, welches ich zudem mit dem Desiderat der Beschreibung phonetischer Realisierungen verknüpfe. Wenn Füllpartikeln verschiedene phonetische Formen zeigen (wofür es starke Indizien gibt), kann überprüft werden, ob bestimmte Formen charakteristisch für bestimmte Kontexte sind und sie dort über- oder mindergebraucht werden. Meine Hypothese ist, dass sich aufgrund der hohen Formvariabilität von Füllpartikeln (vgl. Kapitel 2) möglicherweise phonetische Korrelate bestimmter Auftretenskontexte finden lassen, die die These stützen, dass es kontextuell unterschiedliche Verwendungsmuster und damit unterschiedliche Form-Kontext-Paare gibt. Mithilfe dieses innovativen Ansatzes wird zunächst getrennt voneinander die Form und der Kontext von Füllpartikeln beschrieben. Dies bedeutet, dass Kontexte nicht nur in der Umgebung einer Füllpartikel beschrieben werden, sondern sämtliche in den Daten enthaltenen ausgesuchten Kontexte, sodass eine nachfolgende Analyse auch Aussagen darüber treffen kann, in welchen Kontexten und wie häufig keine Füllpartikel vorkommt. Durch die Vermeidung einer holistischen *Top-down*-Funktionszuweisung und die Wahl des umgekehrten Weges, eines auf Oberflächenmerkmalen basierenden *Bottom-up*-Ansatzes,

wird die Analyse von Form und Kontext separat nachprüfbar und replizierbar. Dies ist insofern eine adäquate Herangehensweise an die Beschreibung dieses Phänomens, als eine eindeutige Funktionszuordnung nicht möglich ist. Die produzierten Formen können so jedoch unabhängig von ihren Kontexten beschrieben und systematisiert werden. Erst in der anschließenden korpusbasierten Analyse wird dann eine kombinierte Auswertung von phonetischer Form und linguistischem Kontext vorgenommen. Die Analyse des Audiosignals in den Kontexten ist ein zusätzliches, jedoch unerlässliches Novum und ermöglicht eine detaillierte Beschreibung der Füllpartikeln in Bezug auf die phonetisch-akustischen Korrelate ihrer Verwendung.

Darüber hinaus könnte man evaluieren, inwiefern die angenommenen Funktionen aus Abschnitt 3.1.2 mit den emergierten Form-Kontext-Mustern korrelieren. Die hier vorgeschlagene Analyse lässt auch die Analyse von Kontextdistributionen zu, in denen keine Füllpartikel auftritt, und ermöglicht so die Theoriebildung komplementärer Muster. Beide Forschungsfragen führen über das hinaus, was diese Arbeit leisten kann – der Fokus hier liegt auf der detaillierten phonetischen Formbeschreibung und ihrer potenziellen Kontextabhängigkeit. Jedoch wird hierdurch weiterführenden Forschungsfragen die methodische und, da die Annotationsdaten größtenteils zur Verfügung gestellt werden dürfen, auch eine empirische Grundlage bereitet. Insgesamt führt der hier entwickelte Ansatz im besten Falle zu einer weniger holistischen und subjektiven Interpretation von Füllpartikeln und damit zu einer höheren Systematisierung der Füllpartikelforschung.

Um Form-Kontext-Muster zu finden, benötigt man eine Formalisierung der relevanten phonetischen Formen und linguistischen Kontexte. Tabelle 3.2 wiederholt das in Kapitel 2 erarbeitete Schema zur Darstellung der Formrepräsentationen von Füllpartikeln und erweitert es um ihre Kontextstruktur, die verschiedene Arten von Kontexten beschreibt, nämlich sequenzielle Kontexte (wie segmentale oder wortbasierte) und hierarchische Kontexte (wie phrasale, prosodische oder dialogstrukturelle), die ich in den folgenden Abschnitten 3.2.2 und 3.2.3 beschreibe und diskutiere.

In Tabelle 3.2 ist die orthographische und symbolphonetische Transkription von *und ähm* (*Einatmen*) zusammen mit ihren phonetisch-akustischen Beschreibungsparametern gegeben. Akustisch können mindestens die in Abschnitt 2.2 diskutierten phonetischen Parameter gemessen werden. Dabei wird die Füllpartikel in ihre kategorial trennbaren Laute segmentiert. Für verschiedene Segmente kann die Dauer

Tabelle 3.2: Schema verschiedener Repräsentationsebenen von Füllpartikeln und ihrer Kontextstruktur anhand des fiktiven Beispiels *und ähm*, gefolgt von einem hörbaren Einatmungsgeräusch. Die akustischen Maßeinheiten werden in Kapitel 4 beschrieben. Die Füllpartikel ist in sequenzielle (se.) und hierarchische (hi.) linguistische Kontexte eingebettet.

		Antezedentia	Füllpartikel			Postzedentia
	Orthographisch Symbolphonetisch und	ähm ?ɛ:m			...
Akustik	Segmentierung		?	Vokal	m	
	Dauer		ms	ms	ms	
	Grundfrequenz			Hz		
	Formanten			Hz		
	Phonation			modal		
	Intensität			H ₁ -A ₃	H ₁ -A ₃	
Kontexte	Segmental (se.)		t			...
	Wortbasiert (se.) und				...
	Prosodisch (hi.)	...	Intonationsphrase			Einatmen ...
	Dialogstrukturell (hi.)		Erzählung			

in Millisekunden, die Grundfrequenz und Formanten in Hertz und die Intensität in Dezibel gemessen werden. Die Phonation wird aufgrund methodischer Schwierigkeiten nur perceptiv kategorisiert (hier als modal bezeichnet), vgl. Abschnitt 4.4. Der sequenzielle Kontext von *ähm* besteht auf lautsegmentaler Ebene aus antezedentem [t] und postzedenter Pause, auf wortsegmentaler Ebene aus antezedentem *und* und postzedenter Pause. Der hierarchische Kontext besteht auf prosodischer Ebene aus einer Intonationsphrase, die aus *und ähm* besteht und die vor dem Einatmungsgeräusch endet; auf dialogstruktureller Ebene ist *und ähm* in eine Erzählung eingebettet. Da Tabelle 3.2 zur besseren Darstellung ein fiktives Beispiel wiedergibt, werden die Kontexte nun nochmals kurz an einem tatsächlichen Beispiel aus dem Berlin Dialogue Corpus compiled v. 1 (BeDiaCo_c) exemplifiziert. Beispiel (2) illustriert die verschiedenen Ausprägungen der sequenziellen und hierarchischen Auftretenskontexte und ihre Problematiken für einen Ausschnitt aus einem spontansprachlichen aufgabenfreien Dialog.⁶⁶ Die bisherigen Erkenntnisse zu den jeweiligen Kontexten

⁶⁶Ich gebe hier die aussprachebasierte Transliteration der diplomatischen Ebene in BeDiaCo_c wieder, da ich für jegliche Interpretation immer das akustische Signal hinzuziehe. Andere Transliterationmöglichkeiten sind HIAT (Ehlich & Rehbein 1976, 1979) oder GAT2 (Selting et al. 2009).

werden anschließend ausführlicher in Abschnitt 3.2.2 und 3.2.3 zusammengefasst.

- (2) a. A: ⟨P, 961 ms⟩ **äh** wo kommst du ursprünglich her
 b. B: aus X aus sonem ⟨P, 372 ms⟩ ja aus kein Kuhkaff aber ⟨lacht⟩
 c. A: okay
 d. B: also ne kleinere Stadt is das ⟨P, 59 ms⟩ **äh** (Einatmung, 280 ms) das kennt auch niemand (BeDiaCo_c frei_f1f2_ch2, 257,6–267,3s)

Der sequenzielle segmentale Kontext von *äh* in Beispiel (2a) besteht aus einer langen stillen Pausen von 961 ms und einem [v] in *wo*, der von *äh* in Beispiel (2d) besteht aus einer kurzen stillen Pause von 59 ms und einem Einatmungsgeräusch von 280 ms Dauer. Die Festlegung eines Schwellenwertes für die minimale Ausdehnung einer stillen Pause ist nicht trivial; kategorisiert man erst ab 250 ms eine stille Pause (Goldman-Eisler 1958, vgl. auch die Kritik von Hieke et al. 1983), so besteht der linkssegmentale Kontext von *äh* in Beispiel (2d) aus [s].⁶⁷ Allerdings führt die ‚Wegkategorisierung‘ einer stillen Pause zu keiner Veränderung des akustischen Signals.

Der sequenzielle Wortkontext in (2a) besteht aus dem nachfolgenden Wort *wo*, der in (2d) aus *das* und *das*.

Der hierarchische syntaktische Kontext ist in den beiden Beispielen identisch mit den beiden Intonationsphrasen (IP_n), es ist jedoch nicht eindeutig, ob *äh* in die hierarchische Phrasenstruktur von *also ne kleinere Stadt is das* integriert wird oder ob es eine eigene Konstituente bildet (für eine syntaktische Analyse von *äh* in hierarchischen Strukturen vgl. Ferreira & Bailey 2004).

Der prosodische Kontext in (2a) ist nur hierarchisch zu fassen, als Einbettung in die IP *äh wo kommst du ursprünglich her*. In (2d) ist zu diskutieren, ob *äh* Bestandteil der IP *also ne kleinere Stadt is das* *äh* ist oder eine eigene IP konstituiert, da die Pause vor *äh* sehr kurz ist und die Intonationskontur der IP auf *äh* ohne Auffälligkeiten fortgeführt wird (vgl. Abschnitt 4.2.2).

Die dialogstrukturellen hierarchischen Kontexte lassen sich in unterschiedlicher Granularität und aus verschiedenen theoretischen Perspektiven beschreiben. Im Rahmen der Turn-Taking-Theorie (Sacks et al. 1974) müsste der Ausschnitt in Turns segmentiert werden. Während für Sprecherin A ein Turn in Beispiel (2a) relativ

⁶⁷Die untere angenommene Grenze für stille Pausen wird in der Gesprochenen-Sprache-Forschung je nach Ziel der Studie und Hintergrund der Forschenden unterschiedlich gehandhabt und reicht von 10 ms bis 550 ms (vgl. den Überblick in Fors 2015: 40) oder gar bis 1000 ms (Leeuw 2007: 95).

einfach und intuitiv als solcher zu erkennen ist und *äh* hier demnach turn-initial stünde, müssten für die Analyse von Beispiel (2d) weitreichende Annahmen über Turns getroffen werden, die dann Voraussetzung für eine Segmentierung in entweder einen einzigen Turn (2b mit 2d) oder zwei separate Turns (2b) und (2d) sind. In beiden möglichen Fällen stünde *äh* turn-medial, im zweiten Fall jedoch ‚medialer‘ (tatsächlich etwa in der Mitte des Turns) als im ersten Fall. Die Feststellung der Turn-Medialität ist also nur bedingt aussagekräftig und wird in dieser Arbeit nicht betrachtet.

Dialogstrukturell (Carletta et al. 1997, 1996) stünde *äh* in (2a) am Anfang einer W-Frage, *äh* in (2d) inmitten einer Antwort auf eine W-Frage. Auch in dieser Betrachtung existieren Segmentierungsprobleme, die sich auf die Interpretation der Füllpartikel auswirken. So ist beispielsweise unklar, ob der Satz *das kennt auch niemand* noch zur Antwort gezählt werden sollte oder nicht; falls der Satz als zusätzliche oder eigenständig motivierte Information betrachtet wird, stünde *äh* am Ende der Antwort (zu dieser Diskussion vgl. Abschnitt 4.2.3).

Beispiel (2) und Tabelle 3.2 zeigen den Bedarf für eine systematische Analyse der Kontextstrukturen von Füllpartikeln. Auch wird gezeigt, dass eine solche Analyse keineswegs trivial ist und je nach Segmentierungsentscheidung verschiedene Ergebnisse zustande kommen. Mithilfe der Annotationsrichtlinien in Kapitel 4 wird versucht, dem entgegenzuwirken.

3.2.2 Sequenzielle Kontexte

In diesem Kapitel fasse ich Studien zu sequenziellen Füllpartikelkontexten auf segmentaler und wortbasierter Ebene zusammen, in welchen unterschiedliche phonetische Realisierungen beobachtet werden.

3.2.2.1 Segmente

Der segmentale Kontext beschreibt die den Füllpartikeln vorangehenden und nachfolgenden Segmente. Eine Sonderform phonetischer Segmente sind dabei stille Pausen (für einen Überblick zu stillen Pausen vgl. Fletcher 2010 und Fors 2015). Über Laute als segmentale Umgebung ist unabhängig von Füllpartikeln bekannt, dass sich in Konsonant-Vokal- (CV) und Vokal-Konsonant-Strukturen (VC) Transitionen in den Vokalformanten an den Übergängen von vorangehenden oder folgenden Kon-

sonanten zeigen (Lehiste & Peterson 1961). Dies sollte *mutatis mutandis* auch für Füllpartikeln gelten. Da jedoch über den lautsegmentalen Kontext von Füllpartikeln keine Studie auffindbar war, diskutiere ich nun ausgewählte Studien zu Füllpartikeln und stillen Pausen. Beide Kontexte, Laute und Pausen, untersuche ich auch in den beiden Korpusstudien (vgl. Kapitel 5 und 6).

Stille Pausen sind als Abwesenheit stimmhafter Komponenten im akustischen Signal definiert (vgl. Fletcher 2010: 573). Im Englischen ziehen Vokal-Nasal-Formen (VN-Formen) längere stille Pausen nach sich als dies bei vokalischen Formen (V-Formen) der Fall ist (Clark & Fox Tree 2002). Clark & Fox Tree analysieren als Hauptquelle ihrer Studie das London-Lund-Korpus (Svartvik 1990), das aus 50 Gesprächen (170 000 Wörter) britischer Erwachsener mit hauptsächlich akademischem Hintergrund besteht, die zwischen 1961 und 1976 aufgenommen wurden (Clark & Fox Tree 2002: 80). Nach *um* sind Verzögerungen wie Pausen, Füllpartikeln und Pausen plus Füllpartikeln (ebd.: 83) sowohl signifikant häufiger (61 %) als nach *uh* (29 %), als auch signifikant länger (ebd.: 82). Die Pausendauer orientiert sich an den Transliterationen, die selbst wiederum eine kategorial-perzeptive Einschätzung seitens der Transkribierenden darstellen, keine akustische Messung.⁶⁸ Da oft auch andere prosodische Signale wie beispielsweise Segmentlängungen als subjektive Pause perzipiert werden (Duez 1993), ist die Schlussfolgerung aus den Daten kritisch zu sehen. Vor *um* sind Verzögerungen signifikant häufiger (41 %) als vor *uh* (34 %), die Verzögerungsdauer ist hingegen nicht unterschiedlich lang. Hieraus folgern Clark & Fox Tree, dass sich *uh* und *um* hauptsächlich darin unterscheiden, dass sie unterschiedlich lange Verzögerungen einleiten (Clark & Fox Tree 2002: 86).

Ein ähnliches Ergebnis beschreibt die Studie von Smith & Clark (1993), in der 25 englische Versuchspersonen je 40 Fragen beantworten, wobei 39mal *uh* und 38mal *um* produziert wird. Stille Pausen vor *um* sind im Mittel mit 1,1s länger als vor *uh*; Pausen nach *um* sind im Mittel 3,1s länger als nach *uh* (ebd.: 35, Tab. 6). Die Position der Füllpartikel in der Antwort wird allerdings nicht näher beschrieben. Zudem handelt es sich um eine interviewähnliche Methode, die mit gleichberechtigten Dialogen nur schwerlich vergleichbar ist.

Rose (2015) untersucht eine Stichprobe von 165 min des Santa Barbara Corpus Teil

⁶⁸„Brief pauses ‚of one light foot‘ are marked with periods (.), and unit pauses ‚of one stress unit‘ with dashes (-). When we need a measure of pause length, we treat the unit pause as 1 unit long, and the brief pause as 0.5 units long [...]“ (Clark & Fox Tree 2002: 80).

1 (Du Bois et al. 2000) mit 149 Füllpartikeln in aufgabenfreien Alltagssituationen, die nicht im Labor stattfinden. Während sich *uh* und *um* nicht signifikant in der Dauer unterscheiden, folgen jedoch nach *um* mit 36,9% der Fälle signifikant häufiger stille Pausen nach als nach *uh* (16,9%). Die mittlere Dauer der stillen Pause ist nach *um* (+ 910 ms) nur marginal länger als nach *uh* (+ 744 ms). Rose fasst die Dauer der V-/VN-Form und die Dauer der nachfolgenden stillen Pause unter dem Maß des „total delay“ zusammen. Für VN + stille Pause ist die Gesamtdauer signifikant länger als für V + stille Pause (Rose 2015: 74).

Jessen (2012) fasst unveröffentlichte Ergebnisse⁶⁹ zu Füllpartikeln in segmentalen Kontexten in deutscher Spontansprache („erschwerte Bildbeschreibung“, ebd.: 158) von 72 Männern danach zusammen, ob vor oder nach einem *äh* oder einem *ähm* eine stille Pause (in seiner Studie alles über 100 ms, ebd.: 148) steht, oder ob sich die Füllpartikel (FP) „nahtlos an den vorangehenden bzw. folgenden Kontext anschließt“ (ebd.: 149). Jessen gibt die prozentuale Verteilung von allen *äh*- und *ähm*-Entitäten für die vier Kontexte Laut-FP-Laut, Pause-FP-Laut, Laut-FP-Pause und Pause-FP-Pause wieder (ebd.: 154). 41,6% aller *ähm* stehen zwischen Pausen, aber nur 19,6% aller *äh*, während 37% aller *äh* zwischen Wörtern stehen, aber nur 16,4% aller *ähm* (meine Berechnungen).⁷⁰ Jessen schließt daraus, dass der „Typ *äh* [...] also meistens in einen lautlichen Kontext eingebettet [ist], während [der] Typ *ähm* meistens isoliert in Erscheinung [tritt]“. Diese Verteilung spiegelt die Tendenz der Studie von Clark & Fox Tree (2002) wider. Clark & Fox Tree stellen zusätzlich die Hypothese auf, dass VN-Formen eine längere nachfolgende Pause anzeigen als V-Formen. Dies ist aus den Daten von Jessen nicht zu entnehmen, allerdings sind beide Formen in Pause-FP-Pause-Kontexten länger ($\bar{x}_{\text{äh}} = 447$ ms, $s = 152$ ms; $\bar{x}_{\text{ähm}} = 560$ ms, $s = 202$ ms) als in Laut-FP-Laut-Kontexten ($\bar{x}_{\text{äh}} = 317$ ms, $s = 132$ ms; $\bar{x}_{\text{ähm}} = 443$ ms, $s = 141$ ms) (Jessen 2012: 154).

Sowohl fürs Englische als auch fürs Deutsche lässt sich also tentativ sagen, dass VN-Formen häufiger zwischen stillen Pausen stehen als V-Formen, während sich

⁶⁹Basierend auf Studien von Trouvain, Tschäpe et al. (2005) und einer unveröffentlichten Masterarbeit von Bauer (2007): Sprecherspezifisch der Stimmtonhöhe in gefüllten Pausen und am Ende von Intonationsphrasen. Masterarbeit, Universität des Saarlandes, Fach Phonetik und Phonologie.

⁷⁰Tabelle 5-2 in (Jessen 2012: 154) gibt Prozentzahlen aller *äh* und *ähm* gemeinsam auf alle Kontexte verteilt wieder, d. h. alle Zahlen zusammen ergeben 100%. Allerdings kommt *äh* in 68,3%, *ähm* nur in 31,7% aller Fälle vor. Zur Evaluierung der Vorkommen je *äh*- und *ähm*-Gesamtheit für jeden Kontext normalisiere ich die prozentuale Verteilung je Form, so dass jede Formspalte 100% ergibt.

das Verhältnis in Laut-FP-Laut-Kontexten eher umkehrt. Auch ist die Dauer einer Füllpartikel zwischen zwei Pausen länger als in anderen Kontexten, während für die Länge der Pause im Zusammenhang mit der Länge der Füllpartikel keine konsistenten Ergebnisse vorliegen.

3.2.2.2 Wörter

Dieser Abschnitt betrachtet nicht mehr die Ebene der füllpartikeladjazenten Laute, sondern die der füllpartikeladjazenten Wörter, vergrößert also den Kontext. Auch in dieser Ebene werden Pausen betrachtet, auch wenn sie nun nicht mehr mit Lauten, sondern eben mit Wörtern verglichen werden. Die Vergleichbarkeit ist in diesem Fall dennoch gegeben, da Laute im vorherigen Abschnitt meistens auch Anfänge oder Enden von Wörtern darstellen.

Analog und zu den oben beschriebenen Pausen- und Laut-Kontexten in Jessen (2012) analysiert Leeuw (2007) in Interviews mit 21 deutschsprachigen Versuchspersonen Pausen- und Wort-Kontexte (vgl. Tabelle 3.3). Nur Pausen, die länger als eine Sekunde waren, gehen in die Kategorie Pause ein, kürzere Dauern sind der Kategorie Wort zugeordnet (ebd.: 95). VN-Formen erreichen im Vergleich zu V- und N-Formen die höchsten prozentualen Werte besonders in Wort-FP-Pause- und Pause-FP-Wort-Kontexten. Trotz ihrer ähnlichen Vorgehensweise lassen sich diese Kontexte jedoch nicht mit den von Jessen (2012) ermittelten Werten vergleichen, da Jessen alle Kontexte und Formen zu 100 % addiert, während Leeuw die Formen je Kontext zu 100 % addiert und sich die Pausenschwellen stark unterscheiden (100 ms bei Jessen vs. 1000 ms bei Leeuw).

Tabelle 3.3: Vergleich der Anteile vokalischer, vokalisch-nasaler und nasaler Füllpartikelformen bei deutschsprachigen Versuchspersonen in Wort-Pause-Kontexten in der Studie von (Leeuw 2007: 104).

Kontext	Füllpartikelform		
	vokalisches	vokalisch-nasal	nasal
Wort-FP-Wort	42 %	58 %	–
Wort-FP-Pause	17 %	82 %	1 %
Pause-FP-Wort	6 %	75 %	19 %
Pause-FP-Pause	27 %	58 %	15 %

Zu Wörtern beziehungsweise Wortarten in der Umgebung von Füllpartikeln exis-

tieren einige weitere Studien, die allerdings nur vereinzelt die phonetische Form von Füllpartikeln betrachten und eher auf Wortarten abzielen.⁷¹

Ob nun von Laut- oder Wortkontexten die Rede ist, ist also für die phonetische Form vorerst nicht von Bedeutung, da aufgrund inkonsistenter beziehungsweise in Deutschen nicht vorhandener Ergebnisse zu Wortarten keine Aussage getroffen werden kann.

3.2.3 Hierarchische Kontexte

Mit hierarchischen Kontexten beschreibe ich alle Kontexte, innerhalb derer weitere, über laut- und silbensegmentale Einheiten hinausgehende hierarchische Abhängigkeiten bestehen können, wie syntaktische Phrasen, Intonationsphrasen, Äußerungen, Turns und dialogstrukturelle Einheiten. Dieser Abschnitt diskutiert die bisher beobachteten unterschiedlichen phonetischen Realisierungen in solchen Kontexten. Bisweilen ist nicht klar, ob eine Füllpartikel, die am Rand einer solchen Konstituente steht, auch Teil der Konstituente ist. In solchen Fällen würde auch eine sequenzielle Betrachtung infrage kommen, da die Füllpartikel dann zwischen zwei Einheiten stünde. Daher werde ich in diesem Kapitel auch Studien diskutieren, die solche Positionen am Rande von größeren Konstituenten erwähnen, ohne tiefgehend die genaue theoriegeleitete Einbettung in ebendiese Strukturen zu diskutieren.

3.2.3.1 Syntaktische Phrasen und Äußerungen

Shriberg (1994) untersucht drei Korpora des amerikanischen Englisch.⁷² In allen drei Korpora tendieren V-Formen eher dazu, satzmedial aufzutreten und VN-Formen eher dazu, satzinitial aufzutreten (ebd.: 154). Über alle Korpora hinweg liegt der Anteil satzinitialer Füllpartikeln an allen vorkommenden Füllpartikeln bei 41,5%–64,3%, der Anteil satzmedialer Füllpartikeln entsprechend bei 35,7%–58,5% (ebd.: 154–155). In allen drei Korpora ist die V-Form häufiger als die VN-Form (ebd.: 155).

⁷¹Fürs Englische beschreiben Maclay & Osgood (1959) & Martin & Strange (1968), dass Füllpartikeln eher vor Inhalts- als vor Funktionswörtern zu finden sind. Andere Studien finden keinen Unterschied (Beattie & Shovelton 2002; Cook 1971; Rose 2015) oder mehr Füllpartikeln vor Funktionswörtern (Hawkins 1971) oder mehr Füllpartikeln nach Funktionswörtern (Schneider 2014).

⁷²66 Sprecherinnen und Sprecher mit Reiseplanungsdialogen in AMEX (ca. 12 000 Wörter, Kowtko & Price 1989), 15 Sprecherinnen und 15 Sprecher aufgabenfreier Telefongespräche in Switchboard (ca. 40 000 Wörter, Godfrey & Holliman 1993), und Mensch-Computer-Dialoge aus ATIS (D. A. Dahl et al. 1994) mit ca. 180 000 Wörtern (Shriberg 1994: 35–40).

Satzinitiale Disfluencies (hierunter definiert Shriberg nicht nur Füllpartikeln) werden wahrscheinlicher, je länger ein Satz ist (Shriberg 1994: 106–107). Auch Schneider (2014) findet für einen Teil des Switchboard Corpus des Amerikanischen Englisch, dass besonders am Beginn oder in der Nähe des Beginns von Sätzen eine ‚Häsitation‘ (in dieser Studie „filled and unfilled pauses“, ebd.: 1) produziert wird (ebd.: 243), was sich mit den Ergebnissen von Shriberg (1994) deckt. Die Formen *uh* und *uhm* haben ein ähnliches Verteilungsmuster innerhalb von Präpositionalphrasen (Schneider 2014: 63), eine genaue Aufschlüsselung der Formvarianten über die Positionen zu Beginn von Sätzen wird jedoch nicht gegeben.

O’Shaughnessy (1992) untersucht Mensch-Computer-Interaktionen, in denen 42 Versuchspersonen dem Computer Anweisungen zur Reiseplanung geben.⁷³ Er unterscheidet ‚grammatische‘ Füllpartikeln an „major syntactic boundaries“ von ‚ungrammatischen‘ Füllpartikeln „within small syntactic units, such as noun groups“ (ebd.: I-521). Die erste Kategorie zeigt eine Dauer von 200–500 ms, die zweite Kategorie eine Dauer von 170–320 ms. Auch hier wird keine Unterscheidung der phonetischen Realisierungen vorgenommen.

Zur Dauer der Einzelformen findet Rose (2015) in der oben schon erwähnten Stichprobe des Santa Barbara Corpus, dass die mittlere Dauer von *uh* und *um* an größeren Grenzen wie Diskurs- und Satzgrenzen kürzer [sic!] ist ($\bar{x}_{uh} = 552\text{ ms}$, $\bar{x}_{um} = 680\text{ ms}$) als an kleineren Grenzen wie dem Beginn subordinierender Sätze oder Präpositionalphrasen ($\bar{x}_{uh} = 641\text{ ms}$, $\bar{x}_{um} = 963\text{ ms}$) (ebd.: 75). Dieses Ergebnis steht konträr zu den Ergebnissen von O’Shaughnessy (1992).

Zusammenfassend zeigen die Studien, dass Füllpartikeln häufiger an syntaktischen Konstituentengrenzen auftreten als innerhalb von Konstituenten, insbesondere an oder in der Nähe von Satzgrenzen. Bisher wird jedoch nur in wenigen Studien Hinweisen auf unterschiedliche phonetische Realisierungsformen in diesen syntaktischen Kontexten nachgegangen. In englischen Dialogen kommen VN-Formen satzinitial häufiger vor als satzmedial, für kleinere Konstituenten wie beispielsweise Präpositionalphrasen wird kein Formhäufigkeitsunterschied festgestellt.⁷⁴ Die Ergebnisse

⁷³Die untersuchte Sprache ist sehr wahrscheinlich Englisch in Kanada. Von 36 Sprecherinnen und Sprechern wurden jeweils die ersten drei Äußerungen in diesem experimentellen Setting analysiert, hinzu kommen nochmal je drei Äußerungen von sechs zusätzlichen Sprecherinnen und Sprechern, die Pausen enthielten.

⁷⁴Ähnliche Ergebnisse finden sich auch in monologischen Nacherzählungen. Fraundorf & Watson (2013) untersuchen 15 englische Versuchspersonen zwischen 18 und 22 Jahren, die drei Passagen

zur Dauer von Füllpartikeln an unterschiedlichen Positionen sind inkonsistent. Syntaktische Phrasen werden in dieser Arbeit daher nicht weiter betrachtet.

3.2.3.2 Intonationsphrasen

Eine Intonationsphrase (IP) ist eine prosodische Konstituente beziehungsweise prosodisch definierte Phrase, „which [...] is the domain of a perceptually coherent intonational contour, or tune“ (Shattuck-Hufnagel & Turk 1996: 210). Im Rahmen der prosodischen Planung einer Äußerung kann eine syntaktische Konstituenten- oder Satzgrenze gleichzeitig auch eine Intonationsphrasengrenze darstellen (Ferreira 1993; Krivokapić 2007; für einen Überblick vgl. Fletcher 2010). Andererseits können syntaktische und prosodische Grenzen auch auseinanderfallen (vgl. Shattuck-Hufnagel & Turk 1996: 200). Intonationsphrasen werden durch verschiedene phonetische Merkmale markiert, vgl. Vaissière (1983), Peters et al. (2005) und Abschnitt 4.2.2.

An Intonationsphrasengrenzen und innerhalb von Intonationsphrasen können Füllpartikeln stehen. Clark & Fox Tree (2002) definieren in ihrer Korpusstudie drei Positionen an und in Intonationsphrasen, die sie aus sukzessive geringeren Anforderungen an die Sprachplanung ableiten: „(I) at the boundary; (II) after the first word (ignoring *uh* and *um*); and (III) later“ (ebd.: 94). Sie beziehen die Möglichkeit einer Füllpartikel an Position I–III auf die Wortlänge einer IP und ermitteln signifikant unterschiedliche Füllpartikelraten von 4,3% für Position I, von 2,7% für Position II und von 1,3% für Position III; d. h. Füllpartikeln sind in Position III weniger wahrscheinlich, obwohl es in dieser Position mehr Möglichkeiten für sie gäbe, aufzutreten. 56% der Füllpartikeln in Position I, aber nur 39% beziehungsweise 37% der Füllpartikeln in den Positionen II und III sind VN-Formen (ebd.). VN- und V-Formen stehen demnach in diesem Korpus etwa gleich häufig an der Grenze, also vor einer IP. Nach dem ersten Wort und innerhalb einer IP überwiegen die V-Formen.

Swerts (1998) untersucht mithilfe von zwölf monologischen Bildbeschreibungen zweier Sprecherinnen im Niederländischen (insgesamt 46,5 min und 310 Füllpartikeln, ebd.: 487, also 6,7 FP/min), ob „FPs [filled pauses – M. B.] may be informative

aus *Alice im Wunderland* anhand einer zusätzlichen Liste mit *plot points* nacherzählen sollen, die sie nach dem Lesen des Textes erhalten. Sowohl der Beginn einer Äußerung, der Beginn einer grammatischen Phrase, der Beginn einer Reparatur und der Beginn eines neuen Plots erhöhen die Wahrscheinlichkeit, dass eine Füllpartikel auftritt, signifikant (Fraundorf & Watson 2013: 1089). Äußerungen sind definiert als „subject and predicate together separated by a discernible prosodic break“ (ebd.: 1087).

about [...] units that exceed the level of the sentence“ (Swerts 1998: 486). Genauer untersucht Swerts, ob die strukturellen Eigenschaften eines Diskurses die Distribution von (bestimmten) Füllpartikeln voraussagen können (ebd.: 487). Hierzu lässt er die zwölf erhobenen Monologe von 19 Annotatoren anhand von Pausen, Tonhöhenreset, Grenztontyp und weiterer, nicht näher benannter Parameter in „prosodische Phrasen“ segmentieren (ebd.: 487 f.), um auf diese Weise Einblicke in die Verteilung von Füllpartikeln an den Rändern und in der Mitte dieser Phrasen zu erhalten. Phrasengrenzen, die von über 75 % der Annotatoren als Grenze markiert werden, definiert Swerts als ‚stark‘, die anderen als ‚schwach‘ (ebd.: 488). Ausgewertet werden initiale und mediale Positionen; mögliche finale Positionen werden als initiale kategorisiert (ebd.). Von 502 IPs kommen in 226 IPs Füllpartikeln vor (45 %), davon enthalten 109 IPs mediale Füllpartikeln, 117 initiale Füllpartikeln. Sieben der 109 Füllpartikeln in medialen IPs werden als ‚stark‘ kategorisiert; von den initialen sind es 43 der 117. Von allen 64 als stark kategorisierten Grenzen enthalten demnach 50 Stück (78 %) eine Füllpartikel (ebd.). Swerts schließt daraus, dass eine IP-Grenze die Wahrscheinlichkeit für Füllpartikeln erhöht (ebd.). Diese Schlussfolgerung ist kritisch zu sehen – ebenso denkbar ist, dass das Vorkommen einer Füllpartikel von den Annotatoren als Hinweis für eine IP-Grenze interpretiert wurde.

In initialer Position sind 75 % aller Füllpartikeln VN-Formen und 25 % V-Formen (*in toto* $n = 117$), in medialer Position sind 12 % aller Füllpartikeln VN-Formen und 88 % V-Formen (*in toto* $n = 193$) (ebd.: 490). Obwohl unklar bleibt, ob die Vergleichbarkeit von Swerts’ medialer Position zu den IP-Positionen II und III von Clark & Fox Tree (2002) gegeben ist beziehungsweise ob die Kategorien „prosodische Phrase“ und „Intonationsphrase“ übereinstimmen, lässt sich doch für den Beginn von diesen prosodisch orientierten Phrasen eine ähnliche Tendenz feststellen, nämlich dass VN-Formen mit 75 % bei Swerts 1998 und 56 % bei Clark & Fox Tree 2002 überwiegen (allerdings mit der Unschärfe, dass die Einteilung bei Swerts auch potenziell finale Füllpartikeln in dieser Position enthält, s. oben). Füllpartikeln in initialer Position sind zudem signifikant länger ($\bar{x} = 58,7s$) als in medialer Position ($\bar{x} = 36,2s$) (Swerts 1998: 491).

Tabelle 3.4 zeigt, dass Kontexte mit vorangehender Pause und beidseitiger Pause in der Studie von Swerts (ebd.) eher symptomatisch für initiale Positionen und Kontexte mit vorangehendem Wort oder beidseitigem Wort eher symptomatisch

für mediale Positionen sind.⁷⁵ Dies ist nicht allzu verwunderlich, da einerseits das Auftreten stiller Pausen ein möglicher Hinweis auf die Grenzen einer prosodischen Phrase sein kann (vgl. Vaissière 1983) und es andererseits in medialer Phrasenposition aufgrund ihrer potenziell längeren Spanne eine höhere Wahrscheinlichkeit für Wort-FP-Wort- und Wort-FP-Pause-Kontexte gibt. In initialer Position zeigen die niederländischen Füllpartikeln eine signifikant höhere Tonhöhe und eine signifikant längere Dauer als in medialer Position (Swerts 1998: 491).

Tabelle 3.4: Vergleich der Anteile initialer und medialer Füllpartikeln in prosodischen Phrasen (PP) bei niederländischsprachigen Versuchspersonen in Wort-Pause-Kontexten in der Studie von Swerts (1998: 491) (Berechnung von mir – M.B.).

Kontext	Position in der prosodischen Phrase	
	initial	medial
Wort-FP-Wort	10 % ($n = 7$)	90 % ($n = 66$)
Wort-FP-Pause	10 % ($n = 11$)	90 % ($n = 94$)
Pause-FP-Wort	69 % ($n = 33$)	31 % ($n = 15$)
Pause-FP-Pause	79 % ($n = 66$)	21 % ($n = 18$)

Von 204 „Häsitationspartikeln wie *äh* und *ähm*“ (Peters 2005b: 310) an turn-internen Phrasengrenzen im *Lindenstraßenkorpus* (sechs aufgabenbasierte Dialoge von vier männlichen und acht weiblichen deutschsprachigen Versuchspersonen, 80 Minuten, ca. 13 000 Wörter, vgl. Peters 2005a) zeigt der größte Teil (80,9 %) nur eine schwache intonatorische Bewegung, die zumeist als eben oder leicht fallend kategorisiert wird (Peters 2005b: 312). Von den Füllpartikeln mit starken melodischen Ausprägungen sind 26 Stück stark fallend. Nur 8 Stück werden als fallend-steigend und nur 3 Stück als stark steigend klassifiziert (ebd.). Peters analysiert Füllpartikeln augenscheinlich nur an Phrasenenden und nicht an anderen Positionen (ebd.: 206).

Die Betrachtung des prosodischen Kontext der Intonationsphrase zeigt also nicht nur die Distribution von Füllpartikeln (sie sind zu Beginn von IPs wahrscheinlicher als später in der IP) sondern auch den Zusammenhang mit der realisierten phonetischen Form von Füllpartikelformen. So sind VN-Formen IP-initial im Englischen und Niederländischen häufiger als V-Formen; in nicht-initialen Positionen dreht sich das Verhältnis um. Füllpartikeln zeigen eine höhere Grundfrequenz (f_0) und eine längere Dauer IP-initial im Englischen im Vergleich zu IP-medial (Swerts 1998) sowie eine eher ebene, fallende oder stark-fallende IP-finale Intonationskontur im Deutschen

⁷⁵Über den unteren Schwellenwert einer stillen Pause gibt Swerts (1998) keine Auskunft.

(Peters 2005b). Diese Merkmale passen zu den prosodischen Merkmalen von Intonationsphrasen, die im Allgemeinen eine absinkende f_0 über die Dauer einer IP aufweisen (vgl. Vaissière 1983). Überhaupt ist zu erwarten, dass sich prosodische Effekte auf die Realisierung der phonetischen Füllpartikelform auswirken, wie beispielsweise Sprechgeschwindigkeit, finale Längung und initiale Stärkung von Segmenten an Phrasenenden (für einen Überblick s. Cho 2015). Diese prosodischen Effekte werden aber meines Wissens nicht im Zusammenhang mit Füllpartikeln berücksichtigt. So hat der Beginn einer Intonationsphrase eine verstärkende Wirkung auf die phonetischen Merkmale von Segmenten an dieser Position (initiale Stärkung). Beispielsweise wird der Vokal /a/ im Amerikanischen Englisch zu Beginn von Intonationsphrasen länger artikuliert als zu Beginn von (hierarchisch niedrigeren) Wortgrenzen (Cho 2008). Auch in intonationsphrasenfinaler Position zeigen sich Effekte, insbesondere Längungseffekte (finale Längung). Lange (gespannte) und kurze (ungespannte) Vokale in Silbennuklei sowie Nasale in Silbenkoda, wie sie in VN-Formen vorkommen, werden in einsilbigen deutschen Wörtern in IP-finaler Position signifikant länger produziert als in IP-medialer Position (vgl. Belz et al. [eingereicht]). Die längere Dauer einer Füllpartikel könnte also zumindest in Teilen durch ihre Position zu Beginn oder Ende einer IP und damit durch das Wirken initialer Stärkung oder finaler Längung auf ihre Segmente erklärt werden (was sich letztlich auch gezeigt hat, vgl. Abschnitt 6.2.3).

3.2.3.3 Dialogstruktur

Über Intonationsphrasen lassen sich verschiedene weitere Strukturen beobachten: Äußerungen, Redebeiträge (*Turns*) und Dialogstruktur, deren Argumentbereiche sich teilweise überschneiden. In dieser Arbeit untersuche ich die Einheit mit dem maximalen Ausdehnungspotenzial in dyadischen Dialogen, nämlich Dialogstruktur.⁷⁶ Als *Dialog* verstehe ich nach Imo den

interaktive[n] Austausch von mindestens zwei Personen [...], die von Angesicht zu Angesicht, zeitlich und räumlich koprsäent, mit verbalen, paraverbalen und nonverbalen Mitteln kommunizieren (Imo 2016: 341 f.).

⁷⁶Ich vermeide den Terminus *Diskurs*, da sich damit vielfältige andere Forschungsrichtungen verbinden (Grosz et al. 1995; Kamp et al. 2011; Mann & Thompson 1988; Stutterheim 1997).

Das Verständnis von Dialog in dieser Arbeit ist demnach der direkte interaktive verbale Austausch von zwei zeitlich und räumlich kopräsenten Personen von Angesicht zu Angesicht. Ferner wird nur die aufgabenfreie Variante eines spontan-sprachlichen Dialogs berücksichtigt.

Arbeiten zur Untersuchung von Dialogstruktur existieren hauptsächlich als überwiegend qualitative Studien innerhalb der Konversationanalyse (Goodwin 1981; Schegloff et al. 1977) und der Interaktionalen Linguistik (Selting 2007). Grundlegende Arbeiten für die quantitativ-qualitative Beschreibung von dialogischen Strukturen sind die Studien von Carletta et al. (1997, 1996). Die dort entwickelten Schemata legen zugrunde, dass sich Dialoge sequenziell-interaktional entwickeln, und dass sich diese Strukturen beschreiben lassen. Sie basieren auf der Theorie zu ‚Dialogspielen‘ („dialogue games“ in der gleichnamigen Monographie von Carlson 1985). Dialogspiele bestehen aus „utterances starting with an initiation and encompassing all utterances up until the purpose of the game has been either fulfilled [...] or abandoned“ (Carletta et al. 1997: 14). Somit ist der Begriff des Dialogspiels als ein übergeordnetes Thema zu verstehen, welches im Dialog behandelt wird. Dialogspiele bestehen wiederum aus Sequenzen von Dialogzügen (Engl. ‚dialogue moves‘),⁷⁷ „which are [...] initiations and responses classified according to their purposes“ (ebd.). Mit Initiierungen und Antworten sind beispielsweise Dialogzüge gemeint, in denen eine Frage gestellt wird (initiativer Dialogzug, s. Beispiel 3a), die dann vom Gegenüber beantwortet wird (responsiver Dialogzug, s. Beispiel 3b). Dialogzüge (DZ) stellen also gewissermaßen Dialogstrukturkonstituenten dar.

- (3) a. A: Wie findest du das Essen in der Mensa?
 b. B: **äh** ich hab (P, 180 ms) hier im Campus no nie in der Mensa gegessen weil ich
 eigentlich in X bin (BeDiaCo_c: a_f8f9, 5,6–11,6 s)

Ich werde im Folgenden die Ergebnisse von Studien zu Füllpartikeln in dialogstrukturellen Kontexten wiedergeben und anschließend die sich daraus ergebende weitere Fragestellung diskutieren. Die bisherige Forschung konzentriert sich besonders auf das Auftreten von Füllpartikeln in bestimmten Positionen des Dialogs ohne Beschreibung der phonetischen Form.

Lickley (2001) untersucht Füllpartikeln im aufgabenbasierten Map-Task-Korpus

⁷⁷ *Dialogue move* wird synonym zu *conversational move* verwendet (Carletta et al. 1997: 19).

HCRC (Anderson et al. 1991) fürs Englische.⁷⁸ Er annotiert Dialogzüge nach dem Dialogstrukturannotationsschema von Carletta et al. (1997, 1996). Im Korpus sind 1492 Füllpartikeln enthalten (Lickley 2001: 94), was ca. einem Prozent aller Wörter entspricht. Am häufigsten sind Füllpartikeln in Antworten auf W-Fragen (ca. 15%) und in instruierenden Zügen (ca. 12%) enthalten, bei einer mittleren Wortlänge der Züge von 8,2 (vgl. Abb. 2, ebd.: 95). Responsive Dialogzüge wie Antworten auf W-Fragen und negative Antworten auf Entscheidungsfragen zeigen sowohl insgesamt als auch in einem Subset von 4–6 Wörtern je Dialogzug mit die höchsten Füllpartikelraten (ca. 1,75–2,5 je 100 Wörter, vgl. Abb. 3, ebd.). Erklärt wird dies damit, dass in diesem aufgabenbasierten Format Bestätigungen, wie sie in positiven Antworten auf Entscheidungsfragen vorkommen, weniger aufwendig zu planen sind, während für negative Antworten und komplexe Antworten auf W-Fragen ein höherer Suchaufwand in der Map Task angenommen werden kann (ebd.: 96).

Nicholson et al. (2010) testen für aufgabenbasierte Dialogen von drei Sprechern und drei Sprecherinnen (sechs gleichgeschlechtliche Dialoge) des amerikanischen Englisch im Korpus Indiana CReST (Eberhard et al. 2010), ob responsive Züge mehr Füllpartikeln enthalten als initiiierende Dialogzüge. Sie nutzen die Dialogzugkategorisierung von Carletta et al. (1997). Initiiierende Züge enthalten „Instructions, Explains, or Queries“, responsive Züge enthalten Antworten auf Fragen, „[a]cknowledgments that a previous move was heard and understood, or a clarification of an answer“ (Nicholson et al. 2010: 90). Über die Segmentierung wird nichts gesagt. Eine Unterscheidung nach Position in einer ‚Äußerung‘ wird auch vorgenommen, nach einer Definition, die sowohl Position I also auch Position II nach Clark & Fox Tree (2002) entspricht (vgl. Abschnitt 3.2.3.2).⁷⁹ Füllpartikeln sind in initiiierenden Zügen signifikant häufiger als in responsiven Zügen (Nicholson et al. 2010: 91), was demnach nicht zu den Ergebnissen der Studie von Lickley (2001) passt.

Die Berücksichtigung der Dialogzüge bringt also mehr Wissen um die dialogstrukturellen Besonderheiten dyadischer Kommunikation in die Positionsanalyse von Füllpartikeln mit ein, als es eine rein prosodische oder eine übergeordnete turnsequenzielle Betrachtung könnte. Zur Veranschaulichung möchte ich nochmals kurz auf

⁷⁸128 Dialoge mit 64 schottischen Studierenden, insgesamt 15 Stunden Sprachmaterial und ca. 150.000 Wörter (Lickley 2001: 94).

⁷⁹„A FP was deemed utterance-initial if it was the first word in a move or if it was preceded only by a single discourse marker (*and, so, now*). All other FPs were considered utterance-medially.“ (Nicholson et al. 2010: 90).

Beispiel (3) zurückkommen, hier wiederholt als Beispiel (4). Am Anfang des beantwortenden Dialogzugs von Person B in (4b) steht ein *äh*. Intonationsphrasen werden in runden Klammern mit subskribiertem IP markiert, Dialogzüge mit Raute und subskribiertem DZ. Verschiedene Segmentierungen in unterschiedlichen Granularitätsstufen führen zu einer unterschiedlichen Anzahl an zu analysierenden Kategorien. Auf diese Weise enthält Beispiel (4) 12 Turns/Sprecherwechsel (6 von A, 6 von B), 9 Dialogzüge (5 von A, 4 von B) und 17 Intonationsphrasen (10 von A, 7 von B). Intonationsphrasengrenzen können mit Dialogzügen beziehungsweise Konstituenten der Dialogstruktur zusammenfallen, genauso wie syntaktische Grenzen mit Intonationsphrasen zusammenfallen können (für Diskussionen zu den Annotationsrichtlinien vgl. Abschnitt 4.2.3 für Dialogzüge und Abschnitt 4.2.2 für Intonationsphrasen).

- (4) a. A: #(Wie findest du das Essen in der Mensa?)_{IP}#_{DZ} Initiierung Frage
 b. B: #(äh ich hab ⟨P, 180 ms⟩ hier im Campus no nie in der Mensa gegessen weil ich eigentlich in *beep* bin)_{IP}
 c. A: #(mhm)_{IP}#_{DZ} Feedback
 d. B: (an dem Campus dort)_{IP}
 e. A: #(mhm)_{IP}#_{DZ} Feedback
 f. B: (deshalb)_{IP} ⟨P, 860 ms⟩ (das bei uns is ziemlich gut aber ich hör immer dass so generell die Leute nich so zufrieden sin mit den Mensen)_{IP}#_{DZ} Antwort
 g. A: #(mhm)_{IP}#_{DZ} Feedback
 h. # (also ich bin sehr)_{IP} ⟨P, 340 ms⟩ (positiv überrascht)_{IP}
 i. B: #(ja?)_{IP}#_{DZ} Feedback
 j. A: (von der Mensa)_{IP} (also ich hab davor mein Bachelor in *beep* gemacht)_{IP}
 k. B: #(mhm)_{IP}#_{DZ} Feedback
 l. A: (un da is die Mensa)_{IP} ⟨P, 286 ms⟩ (überhaupt nich gut)_{IP}#_{DZ} Initiierung
 m. B: #(okay)_{IP}#_{DZ} Feedback (BeDiaCoC: a_f8f9, 5,6-31 s)

In (4b) beginnt der Dialogzug von B in Form einer Antwort, wird mehrmals von A mit Feedback-Dialogzügen komplementiert und endet schließlich in (4f). Innerhalb ihrer Antwort produziert B vier IPn. Im Anschluss gibt A Feedback (4g), bevor sie in (4h) einen neuen Dialogzug und eine neue IP initiiert. B gibt in (4i) kurz Feedback, dann setzt A ihren Dialogzug in einer weiteren IP fort (4j). Ihr initiierender Dialogzug ist in diesem Beispiel nach (4l) beendet (*in realiter* geht der DZ noch weiter).

Das *äh* in Beispiel (4b) steht am Anfang eines beantwortenden Dialogzuges, nämlich einer W-Frage. Zugleich steht es auch zu Beginn einer IP. In seinem linksadjazenten Kontext steht je nach Betrachtung entweder das Wort *Mensa* von Versuchsperson A oder eine Pause in der Dialogspur von Versuchsperson B, in seinem rechtsadjazenten Kontext steht das Wort *ich*. Nur aus der Transkription heraus lässt sich die phonetische Realisierung dieses *äh* nicht beschreiben, womit nach jetzigem Stand der Forschung ein Zusammenhang zwischen seiner Realisierungsform und seiner Position im Dialog weder angenommen noch ausgeschlossen werden kann. Eine solche Analyse ist jedoch – wie ich in diesem Kapitel argumentiere – geboten, um die auftretenden Muster in spontaner gesprochener Sprache linguistisch adäquat beschreiben und daraus kommunikative Gebrauchsmuster ableiten zu können. Mit der Form-Kontext-Hypothese (vgl. Abschnitt 3.2.1) postuliere ich, dass die dialogische Struktur mit der Form von Füllpartikeln, i. e. ihrer phonetischen Realisierung, interagiert.

Beispielsweise kann das *äh* in Beispiel (4b) initiale glottale Plosive zeigen und sich in seiner Vokalqualität, Dauer, Grundfrequenz und Phonationsart von anderen Füllpartikeln der gleichen Sprecherin unterscheiden (vgl. Abschnitt 2.2). In seiner spezifischen Position zwischen einer Pause und einem Laut in Beispiel (4b) ist die vokalische Form *äh* nach der Distributionsanalyse von Jessen (2012) in etwa 18 % aller *äh*-Vorkommen vorgesehen (vgl. Abschnitt 3.2.2.1). Nach der Distributionsanalyse von (Leeuw 2007) steht eine *äh*-Form in nur 6 % aller Pause-FP-Wort-Kontexte (vgl. Abschnitt 3.2.2.2). In seinem hierarchischen Kontext (zu Beginn einer Intonationsphrase) könnte es eine eher erhöhte f_0 zeigen als an anderen Positionen (vgl. Abschnitt 3.2.3.2).

Über die Einbindung von Dialogzügen als hierarchisch übergeordnete und damit potenziell verstärkend wirkende prosodisch-linguistische Grenzen ist nichts bekannt, es ist daher durchaus vorstellbar, dass auch der Dialogzug (Antwort auf eine W-Frage) als höhergeordneter Kontext eine verstärkende Wirkung auf die phonetische Realisierung (in diesem Fall die Dauer) von Segmenten an dieser Position ausübt.

Zusammenfassend ist über die phonetische Realisierung der Füllpartikeln im Zusammenhang von Dialogzügen – zumal im Deutschen – kaum etwas bekannt. Bisherige Studien haben vor allem aufgabenbasierte Dialoge untersucht und geben keine nähere Beschreibung der phonetischen Realisierung in bestimmten Positionen des

Dialogzugs. Aus den zusammengefassten Ergebnissen bisheriger Studien über die phonetischen Realisierungen in anderen hierarchischen Kontexten wie syntaktischen Phrasen und Intonationsphrasen sowie in sequenziellen Kontexten wie Pausen, Lauten und Wörtern lässt sich die Hypothese aufstellen, dass auch Dialogzüge bestimmte systematisch vorkommende Ausprägungen phonetischer Realisierungen zeigen, da beispielsweise der Beginn einer Intonationsphrase mit dem Beginn eines Dialogzuges zusammenfallen kann und somit analog zu anderen additiven prosodischen Stärkungsprozessen (vgl. initiale Stärkung, finale Dehnung) ausgeprägtere phonetische Muster erwartbar sind. Die Beschreibung dieser Muster für solche hierarchisch höhergeordnete Chunks in einem Dialog, nämlich Dialogzüge, ist bisher nur rudimentär für die Beschreibung der Distribution von Füllpartikeln erfolgt – ihre phonetische Form wurde nicht beschrieben. Die weitere Beforschung dieses Zusammenhangs ist jedoch notwendig, um die hohe Variabilität der phonetischen Füllpartikelformen (vgl. Kapitel 2) einordnen zu können.

3.3 Paralinguistische Variation

Als paralinguistische Variation bezeichne ich diejenigen Freiheitsgrade in der Realisierung von Füllpartikeln, die nicht von bestimmten linguistischen Auftretenspositionen und -kontexten abhängig sind (vgl. Abschnitt 3.2), sondern sprecher- und situationsspezifisch erklärt werden können. Trotz sprecherspezifischer Unterschiede kann die Beantwortung der Forschungsfrage durch die Wahl geeigneter statistischer Methoden dennoch gewährleistet werden. Bei situationsspezifischen Unterschieden gilt es jedoch, die Aufnahmesituation *a priori* zu kontrollieren.

3.3.1 Sprecherspezifität

Sprecher*innen zeigen eine große Formvariabilität in der Produktion von Füllpartikeln in bestimmten Positionen sowie große Frequenzunterschiede. So finden Clark & Fox Tree (2002: 97) eine Spanne von 1,2–88,5 Füllpartikeln je 1000 Wörter für die 65 untersuchten Versuchspersonen (Median 17,3). Andere Studien berichten über idiosynkratische Präferenzen bei der Wahl und Positionierung von V- und VN-Formen (Leeuw 2007: 109; Shriberg 1994: 156).

Auch Alters- und Geschlechterunterschiede scheinen sich sowohl auf die Frequenz

als auch auf die Präferenz für bestimmte Füllpartikelformen auszuwirken. Männliche Englischsprecher zeigen eine signifikant höhere Füllpartikelrate pro Wort als Frauen (ca. 2,7% vs. 1,4% bei Shriberg 1994: 148, Abb. 29, 3,04% vs. 2,07% bei Bortfeld et al. 2001: 139). In aufgabenbasierten Dialogen drei verschiedener Altersgruppen (im Mittel 28, 47 und 67 Jahre alt) produziert die ältere Gruppe signifikant häufiger Disfluencies⁸⁰ als die mittlere und jüngere Gruppe, während die beiden letzteren sich nicht unterscheiden (ebd.). Alterseffekte können in dieser Arbeit nicht evaluiert werden, da alle Versuchspersonen aus einer ähnlichen Kohorte stammen (20–31 Jahre).

Die männlichen Probanden in den deutschsprachigen Interviews von Jessen (2012: 152) verwenden die V-Form insgesamt häufiger als die VN-Form (2,48 V/min vs. 1,18 VN/min), jedoch verwenden einige Sprecher fast nur die V-Form, andere fast nur die VN-Form. Auch im Englischen wird in Interviews beobachtet, dass Männer mehr V- als VN-Formen produzieren (McDougall & Duckworth 2017: 22). Passend dazu beobachten Wieling et al. (2016), dass Frauen und jüngere Menschen jeweils ein größeres VN-zu-V-Verhältnis aufweisen als Männer und ältere Menschen (vgl. auch Tottie 2013). Im Amerikanischen Englisch liegt das VN-zu-V-Verhältnis im Speed Dating Corpus (Jurafsky et al. 2009), je nach Transkriptionsquelle, für Frauen bei 1,8–2,4, für Männer bei 0,5–0,7 (Acton 2011: 4). Andererseits finden Laserna et al. (2014) in einer Studie alltäglicher Konversationen englischer Muttersprachlerinnen und Muttersprachler außerhalb des Labors mit 263 Männern und Frauen keinen Geschlechtereffekt beim Gebrauch der V-/VN-Formen und nur einen schwachen Alterseffekt ($r = -0,12$, $p = 0,05$) beim Vergleich von vier Probandengruppen (17–19, 20–22, 23–34 und 35–69 Jahre) (ebd.: 333).

Akustisch gibt es eine starke Korrelation ($r = 0,86$) zwischen dem F_1 -Mittelwert von *uh* und dem F_1 -Mittelwert von *uhm* sowie zwischen dem F_2 -Mittelwert von *uh* und dem F_2 -Mittelwert von *uhm* der einzelnen Sprecher in fiktiven Polizeibefragungen mit 60 männlichen englischsprachigen Sprechern (Hughes et al. 2016: 116). Dies deutet darauf hin, dass sich die Qualität des in der Mitte gemessenen Füllpartikelvokals in V-Formen innerhalb eines Sprechers nicht sonderlich von VN-Formen unterscheidet (vgl. ebd.), für die Formanttrajektorien ist das Bild jedoch eher divers (ebd.: 124). In dieser Arbeit wird sich herausstellen, dass der Vokal in V-Formen im

⁸⁰Hierunter zählen in ihrer Studie auch Wiederholungen und Neustarts (Bortfeld et al. 2001: 134).

Deutschen systematisch tiefer und anteriorer als der Vokal in VN-Formen produziert wird.

Keine der besprochenen Studien berücksichtigt direkte akustische oder weitere dialogstrukturelle Kontexte, um die etwaige Abhängigkeit der Distributionen von anderen Variablen zu überprüfen (zu diesem Kritikpunkt vgl. auch Hughes et al. 2016: 102 f.). Es ist beispielsweise nicht ausgeschlossen, dass sich die Dialogstrukturen oder -strategien der Sprecherinnen und Sprecher voneinander und untereinander unterscheiden. Die Herausforderung liegt also darin, über die sprecherspezifischen Frequenzen, Distributionen und Formen zu generalisieren, um systematische Faktoren ausfindig zu machen.

3.3.2 Situationspezifität

Die produzierte Füllpartikelanzahl kann von der Situation abhängen, in der sich ein*e Sprecher*in befindet. Der Einfluss von Alkohol reduziert die Füllpartikelanzahl (Christenfeld & Creager 1996). Auch eine von außen festgelegte Sprechgeschwindigkeit reduziert Füllpartikeln (Christenfeld 1996: 4). Universitätsvorlesungen und Map-Task-Dialoge wirken sich im Portugiesischen nicht auf die Häufigkeit von Füllpartikeln aus (Moniz et al. 2014). Im Englischen steigt die Füllpartikelanzahl, wenn man mit einem Menschen statt zu einem Computer spricht (E. J. Walker et al. 2014: 265), im Schwedischen sinkt sie (Eklund 2004).

Manchmal reicht die bloße Anwesenheit anderer Personen, um die Disfluencyrate zu erhöhen (Broen & Siegel 1972: 224). Schnadt & Corley (2006) hingegen finden keinen Effekt einer zweiten Person im Raum. Das von Bell (1984) beschriebene *audience design* besagt, dass sich Sprecherinnen und Sprecher bei der Wahl ihres Stils oder Registers nicht nur am Gegenüber, sondern zu einem gewissen Grad auch an anderen Personen (Dritten) orientieren, die nicht direkt am Gespräch teilnehmen, die aber bekannt und ratifiziert sind (*auditors*) oder unbekannt sind und zufällig das Gespräch mithören könnten (*overhearers*) (ebd.: 159). Dieser Effekt kann in der vorliegenden Arbeit nicht ganz ausgeschlossen werden, allerdings war der Experimentator nicht direkt mit in der Aufnahmekabine und auch während des Dialogs zumindest in BeDiaCo_c nicht zu sehen.

3.4 Fazit und Forschungsdesiderat

In diesem Kapitel habe ich die (multi-)funktionalen Aspekte von Füllpartikeln, ihre möglichen Auftretenskontexte und einige schon bekannte damit verbundene phonetische Realisierungen vorgestellt. Diese Arbeit wird die phonetische Realisierung von Füllpartikeln in ihren Auftretenskontexten eingehender analysieren (Form-Kontext-Hypothese) und somit prüfen, ob die Formvariabilität durch ihre Kontexte erklärt werden kann. Verschiedene Vorarbeiten deuten darauf hin, dass dies eine fruchtbare Hypothese ist. Aufgrund der hohen Variabilität der Realisierungen und Kontexte ist ein detailliertes methodisches Vorgehen notwendig. Die Hypothese, dass die akustische Realisierung von ihrem Auftretenskontext abhängig ist, macht die Betrachtung sequenzieller, füllpartikeladjazenter Laute und Wörter nötig, um Auswirkungen des segmentalen Kontextes auf die Lautrealisierung auszuschließen. Somit orientiere ich mich an einem der von Local & G. Walker (2005) aufgestellten Imperative für die methodologische Betrachtung phonetischer Strukturen in spontaner Sprache.

Treat all details at all levels as of potential relevance to the participants. [Kursivierung lt. Quelle – M. B.] We simply do not know, from the outset, which details might be of relevance to the participants and might have a communicative function. (ebd.: 122)

Neben der epistemischen Frage nach den ‚Details‘, die jeder Fachbereich, der sich mit gesprochener Sprache beschäftigt, anders beantworten wird, kann trotz der sorgfältigen Beschreibung und Analyse der phonetischen Korrelate in dieser Arbeit keineswegs davon ausgegangen werden, dass andere als die hier betrachteten phonetischen Korrelate relevant für die Verbindung zwischen Form und Kontext sind oder dass andere als die untersuchten Faktoren einen Einfluss auf die phonetischen Korrelate von Füllpartikeln haben. Gerade der Detailreichtum der beeinflussenden Variable Kontext muss notwendigerweise – linguistisch begründet – reduziert werden, um zu einer abstrahierenden Aussage über den Zusammenhang von Kontext und Form zu gelangen.

Die Auswahl der potenziell beeinflussenden Kontexte auf die phonetische Form von Füllpartikeln priorisiere ich unter Betrachtung der bisherigen Forschung und unter den zeitlichen Restriktionen dieser Arbeit daher wie folgt. Als sequenzielle Kontexte werden die füllpartikeladjazenten Laute und Wörter analysiert (Mikrokontext). Hierunter fallen auch Pausen und Atmungsgeräusche. Für diese Kontexte

existieren (zumindest fürs Englische) zahlreiche Studien, die einen Zusammenhang zwischen Kontext und Form nahelegen (vgl. Abschnitt 3.2.2). Als hierarchischer Kontext wird in den beiden folgenden Korpusstudien die Dialogstruktur betrachtet (Makrokontext, da diese Struktur höhergeordneter zusammengehörender Einheiten jedem spontanen Dialog innewohnt und in der Literatur erste Hinweise auf einen erklärenden Aspekt dieser Struktur vorhanden sind, die bisher nicht vertieft wurden (vgl. Abschnitt 3.2.3.3). Nach der explorativen Studie (vgl. Abschnitt 5.1.1) ziehe ich in der konfirmatorischen Studie zusätzlich Intonationsphrasen (vgl. Abschnitt 3.2.3.2) als hierarchischen Kontext hinzu, quasi als Mesokontext zwischen Mikro- und Makrokontexten, um eine weitere Verfeinerung der dialogstrukturellen Kontexte zu erreichen (vgl. hierzu auch Kapitel 5).

Während für die Überprüfung der in Abschnitt 2.3 aufgestellten Hypothese der Formvariabilität eine kontextunabhängige phonetische Analyse von Füllpartikelinstanzen ausreicht (was die Schwierigkeit nicht mindert), kann die Form-Kontext-Hypothese nur mithilfe linguistischer Kontexte überprüft werden. Hierbei sind klar unterscheidbare Kontexte vonnöten, um die vorkommenden Formen distributiv fassen zu können. Mithilfe des hier verwendeten kleinteiligen und unabhängig aufeinander aufbauenden Ansatzes der Mehrebenenannotation dieser Kontexte in gesprochenen Korpora spontansprachlicher aufgabenfreier Dialoge kann jeder Kontext sowohl für sich selbst analysiert als auch in den Zusammenhang anderer Kontexte gestellt werden. Die hier von mir aufgestellten Hypothesen, die Form-Kontext-Hypothese (vgl. Abschnitt 3.2.1) und die Hypothese der Formvariabilität (vgl. Abschnitt 2.3), emergieren aus dem aktuellen Forschungsstand heraus. Ihre Überprüfung ist daher ein notwendiger und fruchtbarer Schritt hin zu einem besseren Verständnis von Füllpartikeln. Die im Folgenden verwendete Methode der oberflächenbasierten Vergabe von Annotationswerten ermöglicht eine transparente, nachvollziehbare und gebrauchsbasierte Analyse der phonetischen Füllpartikelform in ihrem Kontext, da hier keine weitgreifenden funktionalen Interpretationen gezogen werden. Im Ergebnis der folgenden Studien ist es so möglich festzustellen, ob systematische Form-Kontext-Paare angenommen werden können. Erst im Anschluss entstünde somit eine solide und gebrauchsbasiert emergierte Wissensbasis, um diesen Form-Kontext-Paaren (und dies wäre ein offenes Desiderat für weitere Forschung außerhalb dieser Arbeit) kommunikative Funktionen zuzuschreiben.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





4 Methode

Für diese Arbeit kombiniere ich phonetische und korpuslinguistische Methoden, indem akustische Messwerte an den von meinen Annotationen ausgezeichneten Stellen aus dem Audiosignal extrahiert und weiter analysiert werden. Zusätzlich definiere ich Annotationswerte, die das akustische Signal beschreiben und in die Analyse einfließen. Dies geht über die ‚klassische‘ Korpusauswertung „gesprochener Korpora“ (Wichmann 2008) hinaus, die sich zwar auf Transkripte von Sprachaufnahmen bezieht (vgl. Kübler & Zinsmeister 2015: 4), aber die in ihnen enthaltenen „sound patterns“ außen vor lässt (Wichmann 2008: 189).

Diese Arbeit enthält eine explorative und eine konfirmatorische Korpusstudie mit vergleichbaren Daten (spontane gesprochene Dialoge). Da während der Exploration von Hypothesen sogenannte „researcher degrees of freedom“ (Roettger 2019) entstehen, d. h. die Hypothesenüberprüfung nicht auf einer einzigen Analyse fußt, sondern mit mehreren Entscheidungen über die Art und Weise ihrer Auswertung verflochten ist, überprüfe ich die Ergebnisse, die sich nach der explorativen Auswertung ergeben, mit anderen Daten in einer konfirmatorischen (also hypothesengeleiteten) Studie.

Die explorative Korpusstudie verwendet eine Stichprobe aus GECO v. 1.1 (Schweitzer & Lewandowski 2013), in welcher die pilotierten Annotationsrichtlinien evaluiert und gegebenenfalls vor der Annotation des konfirmatorischen Korpus geändert werden. Die konfirmatorische Korpusstudie verwendet Daten aus dem Berlin Dialogue Corpus v. 1 (BeDiaCo v. 1) (Belz & Mooshammer 2020), das im Rahmen dieser Arbeit erstellt wurde. Dort halte ich die in der explorativen Studie erstellten Analyseheuristiken so konstant wie möglich, um die Replizierbarkeit der Ergebnisse aus der explorativen Studie zu testen. Gegebenenfalls neu hinzugekommene Hypothesen tragen jedoch notwendigerweise auch in der zweiten Studie wieder explorativen Charakter.

Zusatzmaterial online

Zusätzliche Informationen sind in der Online-Version dieses Kapitels (https://doi.org/10.1007/978-3-662-62812-6_4) enthalten.

© Der/die Autor(en) 2021

M. Belz, *Die Phonetik von äh und ähm*,

https://doi.org/10.1007/978-3-662-62812-6_4

4.1 Korpora

4.1.1 Studie I: GECO v.1.1

Für die explorative Studie I verwende ich einen Teil der Daten aus dem Korpus GECO v. 1.1⁸¹ (Schweitzer & Lewandowski 2013). Das GECO-Korpus wird gewählt, da es bis dato eines der wenigen deutschsprachigen Korpora ist, welches spontane, aufgabenfreie dialogische deutsche Daten enthält, deren komplette Datenbasis zugänglich und schon vorsegmentiert verfügbar ist. Spontansprachlichkeit, Aufgabenfreiheit und Dialogizität stellen genau die situativen Parameter dar, deren Untersuchung ich für Füllpartikeln in dieser Arbeit motiviere (vgl. Abschnitt 1.1). GECO v. 1.1 besteht aus Gesprächen von Studierenden der Universität Stuttgart in guter akustischer Qualität im Labor (Abtastrate 48 000 Hz); jedes Gespräch dauert 25 min. Für diese Untersuchung verwende ich nur Daten aus dem „multidimensionalen“ Teil des Korpus, in welchem sich die Probandinnen durch eine Glasscheibe sehen. Zur Unterscheidung von GECO v. 1.1 heißt der hier verwendete Datensatz GERman CONversations-Füllpartikeln v. 1 (GECO-FP) (Belz 2019b). In acht Dialogen kommt jede der acht Sprecherinnen zweimal vor.⁸² Die Versuchspersonen sind nicht befreundet und kennen sich zum größten Teil nicht, in einigen Fällen sind sie sich für ein Vorgängerexperiment jedoch schon einmal begegnet. Teilweise enthält ihre Intonation und Lexik Merkmale des schwäbischen Regiolektivs. Weitere Metadaten beschreibt Tabelle 4.1.

4.1.2 Studie II: BeDiaCo v.1

Für die konfirmatorische Studie II nutze ich das Korpus BeDiaCo v. 1 (Berlin Dialogue Corpus v. 1) (Belz & Mooshammer 2020), welches im Rahmen dieser Arbeit erhoben wurde, da zum Zeitpunkt der Studie keine mit GECO-FP vergleichbaren deutschsprachigen Dialogdaten mit einer ausreichenden Zahl von Versuchspersonen zur Annotation verfügbar waren. Für diese Studie werden die spontanen aufgabenfreien Dialoge (ca. 15 min) sowie die gelesenen Wortlisten (vgl. Abschnitt 4.4.3.2) zu

⁸¹<http://hdl.handle.net/11858/00-247C-0000-0023-5137-2>

⁸²Sprecherin A spricht mit C und K, C mit A und D, D mit C und M, F mit H und K, H mit F und J, J mit H und M, K mit A und F, M mit D und J.

Anfang und Ende des Experiments verwendet.⁸³ Die Versuchspersonen waren frei in ihrer Themenwahl, wurden aber angeleitet durch folgenden von mir zu Beginn eingesprochenen Text.

Sie dürfen sich jetzt 15 Minuten frei unterhalten, über beliebige Themen. Ich werde mich nicht einmischen, außer Sie möchten das Experiment abbrechen. Ich gebe Ihnen gerne eine kleine Starthilfe: Wie bewerten Sie denn das Essen in der Mensa?

BeDiaCo v. 1 enthält 16 Versuchspersonen. In dieser Studie kombiniere ich die Daten von BeDiaCo v. 1 mit Daten, die im gleichen Experiment entstanden sind, für die aber keine Einverständniserklärungen zur Veröffentlichung in einem Repository vorliegen (s. auch Abschnitt 4.6). Diese restringierten Daten erhalten die Bezeichnung BeDiaCo_r v. 1 (Berlin Dialogue Corpus restricted v. 1). Betroffen sind vier Dialoge (gleichgeschlechtliche Dyaden) mit acht Versuchspersonen (sechs Frauen, zwei Männer). Das so neu kompilierte Korpus und im folgenden hier verwendete Korpus mit 24 Versuchspersonen erhält zur Desambiguierung die Bezeichnung BeDiaCo_c (Berlin Dialogue Corpus compiled v. 1).

Die Aufnahmen wurden mit Audacity (Audacity Team 2017) mit einer Abtastrate von 44 100 Hz durchgeführt. Jede Versuchsperson trug eine Sprechgarnitur mit Kugelmikrofon (Headset Opus 54 beyerdynamics), die über einen Verstärker zusammenliefen (Tascam 2x2). Die Aufnahme fand in einer schallisolierten Kabine statt, die Versuchspersonen saßen sich gegenüber.

Jede der 24 Versuchspersonen nimmt nur an genau einem Dialog teil. In vier Dialogen reden Männer mit Männern (m/m), in vier Dialogen Frauen mit Frauen (f/f) und in weiteren fünf Dialogen je vier Frauen mit vier Männern (f/m).⁸⁴ Die Versuchspersonen kannten sich nicht. Sie befinden sich größtenteils im Studium und erhielten

⁸³Der komplette Experimentablauf hat den folgenden chronologischen Experimentaufbau: (i) einzelnes Lesen einer Wortliste; (ii) Lösen einer Diapix-Aufgabe (ca. 5 min); (iii) freies Gespräch (ca. 15 min); (iv) nochmalige Diapix-Aufgabe (5 min); (v) nochmalige Wortliste. BeDiaCo v. 1 verwendet die von Alina Zöllner ins Deutsche übersetzten Diapix *Street 1* und *Farm 1* (Baker & Hazan 2011). Eine Ansicht der Originale findet sich hier: <https://www.phon.ucl.ac.uk/project/kidLUCID/diapix.php>, besucht am 23.10.2018.

⁸⁴Ursprünglich wurden 14 Dialoge mit 28 Versuchspersonen aufgenommen. Ein Sprecher in einem m/m-Dialog spricht durchgehend Kiezdeutsch (vgl. Wiese 2010), der Dialog wird daher ausgeschlossen. Ein m/f-Dialog wird ausgeschlossen, um eine einheitliche Stichprobe von 12 Dialogen (4 m/m, 4 f/f, 4 m/m) mit 24 Versuchspersonen (12 männlich, 12 weiblich) zu erhalten.

10 Euro für ihre Teilnahme.⁸⁵ Sie stammen aus Bayern (2), Berlin (9), Brandenburg (2), Hamburg (2), Hessen (2), Mecklenburg-Vorpommern (2), Niedersachsen (2) und Nordrhein-Westfalen (3). Tabelle 4.1 beschreibt weitere Metadaten und vergleicht sie mit denen von GECO-FP.

Tabelle 4.1: Korpusdesign und Metadaten der verwendeten Korpora.

Korpus	Dialoge	VP	Altersspanne	Token ^a	Minuten ^{a, b}
GECO-FP	8	8 ♀	20–30	44 209	197
BeDiaCo _c	12	12 ♂, 12 ♀	18–31	37 263	167

^a Ohne Pausen oder andere in spitzen Klammern (<>) annotierten Token.

^b Summe aller Versuchspersonen (monophon).

Nach der Aufnahme wurden je Versuchsperson mit CAT⁸⁶ stille Pausen detektiert⁸⁷, automatisch gechunkt (Poerner & Schiel 2018) und aussprachebasiert transliteriert. Anschließend wurden die Chunks in Praat (Boersma 2001) ggf. manuell korrigiert und dann mit WebMAUS (Kisler et al. 2017) in Wörter segmentiert und signalaligniert.

Die Dialoge wurden manuell durch Anhören pseudonymisiert, indem auf einer Ebene **pseudo** Eigennamen und andere Merkmale zur Identifizierung Betroffener mit einem *x* markiert und das entsprechende akustische Signal an dieser Stelle mithilfe eines R-Skripts (Belz 2019c) mit einem Sinuston von 200 Hz ersetzt wurde.

4.1.3 Diplomatische Transliteration und Segmentierung

In GECO-FP existiert bereits eine Transliterationsebene (**words**), die automatisch mit dem Signal aligniert ist und eine orthographiebasierte Transliteration enthält.⁸⁸ Diese Ebene verwende ich als korrigierte Transliterationsebene (**wordscor**) weiter, um die Alignierung von Transliteration und Signal in Umgebungen von Füllpartikeln zu korrigieren.⁸⁹

Auch BeDiaCo v. 1 enthält eine Transliterationsebene, die hier jedoch anders als

⁸⁵Gefördert aus dem Forschungsfonds des Instituts für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin.

⁸⁶Chunked Audio Transcription, Version 3.1.2, Simon Sauer, Humboldt-Universität zu Berlin.

⁸⁷Matlabskript *fu_pause_detector_bracket*, Uwe Reichel, Ludwig-Maximilians-Universität München.

⁸⁸Die Namen von Annotationsebenen werden in äquidistanter Schriftart gesetzt.

⁸⁹Die Ebene **words** und **wordscor** bleiben so trotz der Korrekturen vergleichbar.

in GECO-FP eine diplomatische, an der Orthographie orientierte, aber aussprachebasierte Transliteration enthält (**dip1**). Die phonetische Form [isəs] wird beispielsweise als *isses* wiedergegeben. Für jegliche Interpretation der transliterierten Daten ist das akustische Signal verfügbar, so dass eine zusätzliche Transliteration in einem der konversationsanalytischen Verfahren wie HIAT (Ehlich & Rehbein 1976, 1979) oder GAT2 (Selting et al. 2009) nicht mehr zwingend notwendig ist.

4.2 Annotation

Für die Durchführung einer Korpusstudie bedarf es einer guten Definition der untersuchten Variablen und ihrer Varianten (Lüdeling 2017). Jede Entscheidung für eine bestimmte Kategorie muss nachvollziehbar sein und jede Annotationsentscheidung ist zugleich eine Interpretation, die mit einer zwingenden Reduktion der Wirklichkeit einhergeht (vgl. Lüdeling 2011). Die Wiederverwendung eines Korpus „erfordert unter anderem eine Dokumentation der jeweiligen Korpora mit den verwendeten Annotationskonzepten in den genutzten Formaten“ (Odebrecht 2018: 39). Um diesen Ansprüchen an eine nachvollziehbare, nachhaltige und wiederverwendbare Korpusstudie genüge zu tun, erläutere ich im Folgenden meine Annotationsrichtlinien.⁹⁰

4.2.1 Füllpartikelannotation

Die Ebenen **dip1** (BeDiaCo_c) beziehungsweise **wordscor** (GECO-FP) enthalten eine Transliteration als Annotation auf dem akustischen Signal. Hier können also orthographische Annotationen der Füllpartikelexponenten hinzugefügt werden. Füllpartikeln werden in beiden Korpora durch Anhören der einzelnen Tonspuren eines Dialoges in Praat (Boersma 2001) identifiziert und gegebenenfalls in die Transliterationsebene eingefügt, vgl. hierzu auch die Ebenendokumentation von **dip1** in BeDiaCo_c (Tabelle B.1) beziehungsweise **wordscor** in GECO-FP (Tabelle A.1). Für die wichtigsten beiden Schritte (Identifizierung und Segmentierung) wird zur Messung der Übereinstimmung ein Inter-Annotator-Agreement vorgenommen (s. Abschnitt 4.3). Anschließend werden Füllpartikeln in dem hier entwickelten Schema auf verschiedenen Ebenen in Praat annotiert, die für sich genommen unabhängig sind. Somit werden alle Annotationsebenen sowohl getrennt voneinander als auch getrennt von den

⁹⁰Teile von Abschnitt 4.2.1 sind als technischer Bericht in Belz (2019a) veröffentlicht worden.

Primärdaten gespeichert, die Ebenen können sich aber aufeinander beziehen (Sauer & Lüdeling 2016: 423). Tabelle 4.2 demonstriert die Mehr-Ebenen-Annotation für das fiktive Beispiel *und ⟨P⟩ ähm ja dann*, wobei ⟨P⟩ für eine stille Pause steht. Den unteren Schwellenwert für die Länge stiller Pausen lege ich in beiden Studien auf 50 ms fest. Für alle Intervallgrenzen innerhalb der Füllpartikelannotation gilt, dass sie am Nulldurchgang aufsteigender Flanken im Oszillogramm gesetzt werden.

Tabelle 4.2: Mehr-Ebenen-Annotation von Füllpartikeln. Der senkrechte Trennstrich ‚|‘ stellt eine Intervallgrenze dar. **dip1** 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⟩		ähm		ja	dann	dip1 ^a	Signal	Signal
	ap		fv		ps		fp	dip1	dip1 und Signal
	_	?	E	m	j		segm	fp	fp ^b und Signal
			mod				phon ^c	segm	segm

^a **dip1** heißt in den von mir annotierten GECO-Daten **wordscor**.

^b **fp** heißt in den von mir annotierten GECO-Daten **hes**

^c **phon** heißt in den von mir annotierten GECO-Daten **coart**.

Die Ebene **fp** in BeDiaCo_c (beziehungsweise **hes** in GECO-FP) dient der Annotation von Füllpartikeln in ihren verschiedenen Ausprägungen und von ihrem direkten segmentalen Kontext.⁹¹ Weitere Details zu **fp** finden sich in der Ebenendokumentation im Anhang, Tabelle A.2 für GECO-FP sowie Tabelle B.2 für BeDiaCo_c mit einer verbesserten und reduzierten Modifizierung der Richtlinien).⁹² Die Annotationswerte sind hierarchisch aufgebaut und zweistellig. Der Annotationswert beginnt mit dem Buchstaben *f*. Besteht die vorliegende Variante aus einem Vokal, einem Nasal oder einer Vokal-Nasal-Sequenz, so wird der Annotationswert *fv* vergeben (vgl. Abbildung 4.1). Besteht die vorliegende Variante aus einer Abfolge glottaler Plosive, wird der Annotationswert *fg* vergeben (vgl. Abbildung 4.2).

Die Ebene **segm** enthält die segmentale Annotation der auf der Ebene **fp** beziehungsweise **wordscor** identifizierten Füllpartikeln und Füllpartikelkontexte (vgl. für

⁹¹Im Falle von Pausen entspricht der direkte Kontext auch nicht-segmentalem Kontext.

⁹²So enthält die Annotationsrichtlinie zur Ebene **fp** in GECO-FP zusätzlich noch Anteante- und Postpostzedentia.

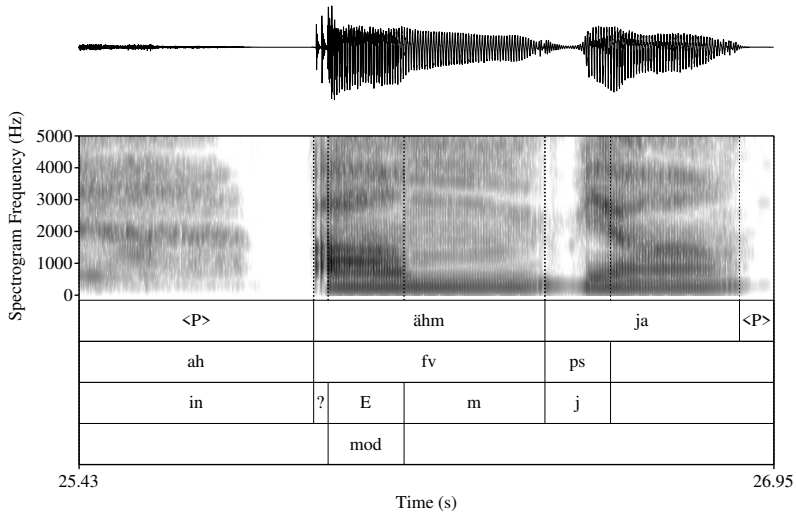


Abbildung 4.1: Nicht-glottale Füllpartikel von Versuchsperson fl in BeDiaCo_c, annotiert auf *dip1* mit *ähm*; auf *fp* mit *ah* (linker Kontext, Atmung), *fv* und *ps* (rechter Kontext, Segment); auf *segm* mit *in* (Inhalation), *?* (ein bis drei glottale Plosive), *E* (Vokal), *m* (Laut nach SAMPA), *j* (Laut nach SAMPA); auf *phon* mit *mod* (modale Phonation).

GECO-FP Tabelle A.3 und für BeDiaCo_c Tabelle B.3 im Anhang). Die Segmentierung orientiert sich daran, ob ein Laut gerade schon oder gerade nicht mehr als Laut einer bestimmten Klasse erkennbar ist. Die Laute werden ohrenphonetisch mithilfe des Speech Assessment Methods Phonetic Alphabet (SAMPA)⁹³ annotiert. Abweichend hiervon annotiere ich alle Vokale innerhalb von Füllpartikeln mit *E*. *E* fällt somit die Rolle eines reinen Platzhalterannotationswertes zu, um keine *a-priori*-Kategorisierung der Vokalqualität vorzunehmen.⁹⁴ Auch die Kontexte erhalten die Transkription nach SAMPA. Stille Pausen erhalten den Wert „_“, Pausen mit hörbarem Atmungsgeräusch werden nach Ein- (*in*) oder Ausatmung (*ex*) klassifiziert.

⁹³<http://www.phon.ucl.ac.uk/home/sampa/index.html>, besucht am 23.07.2018. Ein Vergleich von IPA und SAMPA findet sich in Tabelle D.1 im Anhang.

⁹⁴Die Wahl fällt auf den Wert *E*, da dieser in der Beschreibung von Füllpartikeln zu einer Art Quasi-Standard der symbolphonetischen Transkription geworden ist (vgl. Abschnitt 2.1.2).

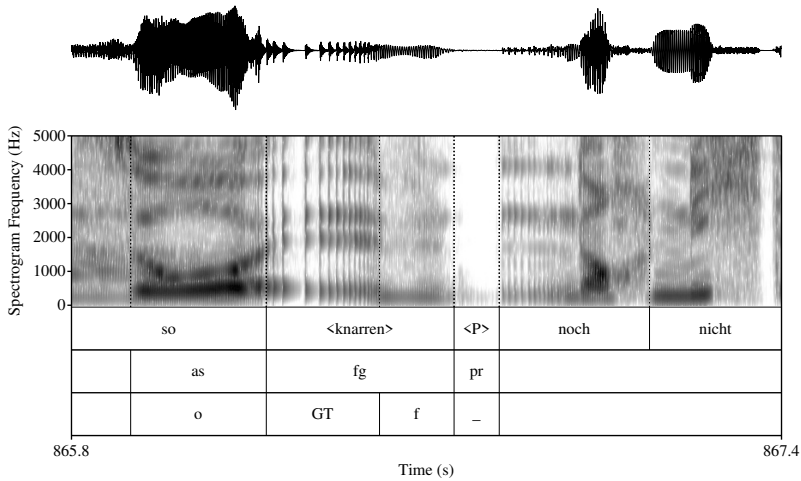


Abbildung 4.2: Glottale Füllpartikel von Versuchsperson f1 in BeDiaCo_c, annotiert auf *dip1* mit *<knarren>*; auf *fp* mit *as* (linker Kontext, Segment), *fg* und *pr* (rechter Kontext, überwiegend stille Pause); auf *segm* mit *o* (Laut nach SAMPA), *GT* (glottale Transitionen antezedenter Segmente in den *fv*-Vokal), *f* (Laut nach SAMPA), *–* (Laut nach SAMPA). Die Ebene *phon* bleibt leer.

In BeDiaCo v. 1 füge ich zusätzlich noch die Annotationswerte *ar* (Antezedensreflex) beziehungsweise *pr* (Postzedensreflex) hinzu, die angibt, dass Teile der Pause artikulatorisch unbestimmbar Reflexe enthalten, die Pause also nicht komplett still ist (vgl. Belz & Trouvain 2019).

Auf der Ebene *phon* in BeDiaCo_c beziehungsweise *coart* in GECO-FP wird die Phonationsart des Füllpartikelvokals annotiert (in BeDiaCo_c auch behauchte Phonation, vgl. die Richtlinien in Tabelle A.4 für GECO-FP und Tabelle B.4 für BeDiaCo v. 1 im Anhang). Glottalisierung wird mit *glott* markiert, modale Phonation mit *mod*. Dabei orientiere ich mich an den von Praat angezeigten Pulsen und der Periodenschwingung im Oszillogramm sowie dem Plosivsignal im Sonagramm (zur Erkennung von Glottalisierung vgl. Abschnitt 2.2.5). Im Laufe der Annotation hat sich jedoch herausgestellt, dass die perzeptive Annotation der Phonation sehr subjektiv ist. Passend hierzu zeigt Davidson (2018), dass bei männlichen Sprechern

perzeptiv oftmals nicht zwischen tiefer Grundfrequenz und Glottalisierung unterschieden werden kann. Für diese Annotationsebene muss demnach die Frage der Validität gestellt werden und es ist fraglich, ob ihre Auswertung reliabel ist. Alternativ kann die Phonationsart des Füllpartikelvokals mit anderen phonetischen Parametern approximiert werden (vgl. Abschnitt 4.4.4).

Sequenzielle Kontexte vor und nach Füllpartikeln (Ante- und Postzedentia) können aus Lauten, hör- und sichtbaren Atmungsgeräuschen, Clicks, stillen Pausen, oder extralinguistischen Geräuschen bestehen. Innerhalb von hör- und sichtbaren Atmungspausen werden stille Pausen nicht separat annotiert, da Ein- und Ausatmungsgeräusche sich häufig nicht direkt der Sprachproduktion anschließen, sondern von kurzen stillen Pausen umgeben sein können (vgl. Belz & Trouvain 2019). Zudem können zu Beginn und am Ende der Einatmung glottale Plosive entstehen; auch diese werden nicht in die Annotation einbezogen (vgl. Abbildung 4.3).

4.2.2 Intonationsphrasenannotation

Intonationsphrasen (IPs) sind intonatorisch definierte und perzeptiv kohärente prosodische Konstituenten (vgl. Shattuck-Hufnagel & Turk 1996: 210). Sie enthalten mindestens eine intermediäre Phrase, die wiederum mindestens einen Tonakzent enthält (vgl. Grice et al. 2005: 13; außerdem Beckman & Pierrehumbert 1986; Nespor & Vogel 1986). Phonetische Merkmale von Intonationsphrasen sind, sowohl sprachübergreifend (vgl. Vaissière 1983) als auch für deutsche Spontansprache (Peters et al. 2005), „segmentelle Längung im Vergleich zur lautlichen Umgebung“, „Intonationsmuster, die stark fallen, hoch steigen oder Fallen und Steigen in einer Satzakkzentposition kombinieren“, „Neueinsatz der F₀-Deklinaton“ und „Pausen und Atmen“ (ebd.: 145). Weitere Merkmale sind Sprechgeschwindigkeitswechsel, Veränderungen der Intensität sowie Laryngalisierung (vgl. ebd.). Auch das DIMA-Modell⁹⁵ nennt als „[p]honetisch-perzeptive Kriterien für eine Grenze [...] eine Pause, phrasenfinale Längung, eine phrasenfinale tonale Bewegung, Tonhöhen-Reset sowie segmentale Phänomene wie Laryngalisierung oder *creaky voice*“ (Kügler & Baumann 2019; vgl. auch Kügler et al. 2015).

Trotz der Annahme, dass die Grenzen einer IP perzeptiv gut erkennbar seien (Shattuck-Hufnagel & Turk 1996: 211; Himmelmann et al. 2018), werden insbeson-

⁹⁵ *Deutsche Intonation, Modellierung und Annotation.*

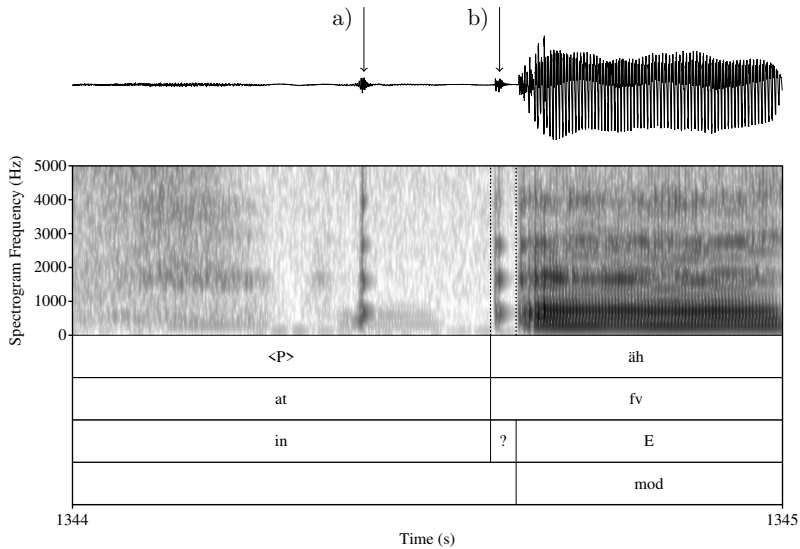


Abbildung 4.3: Nicht-glottale Füllpartikel in GECO-FP von Versuchsperson H in multi_H-J_left mit stiller Pause (<P>) und *äh* *wordscor*, Redewechsel (*at*) und vokalischer Füllpartikel (*fv*) auf *hes*, Einatmung (*in*), glottalem Plosiv (? nach SAMPA) und Vokal (*E*) auf *segn* und modaler Phonation (*mod*) auf *coart*. Der mit (a) markierte glottale Plosiv nach der Einatmungsphase wird anders als der mit (b) markierte nicht zur Füllpartikel gezählt.

dere für die Annotation von IPs in spontaner gesprochener Sprache nachvollziehbare Annotationsrichtlinien benötigt. Als Kriterien für eine Intonationsphrase lege ich somit die (1) maximale Ausdehnung einer kohärenten Intonationskontur mit mindestens einem nuklearen Phrasenton fest, deren Grenzen (2) durch segmentale Längung, (3) tonale Bewegung, (4) Tonhöhenreset, (5) Laryngalisierung, (6) Pausen größer als 50 ms oder (7) Atmungspausen festgelegt werden. Nachrangig können zudem (8) syntaktische Merkmale einbezogen werden. IPs erhalten auf der Ebene *ip* den Wert *ip*.⁹⁶ Atmungspausen markieren immer eine IP-Grenze. Intonationsphrasen müssen nicht mit syntaktischen Phrasen übereinstimmen. Folgt eine längere Phase

⁹⁶Die Minuskelform des Annotationswertes wurde zur heuristischen Vereinfachung des Annotationsprozesses gewählt und steht hier nicht für die manchmal damit abgekürzte intermediäre Phrase.

schnellen Sprechens mit mehreren nuklearen Tonakzenten aufeinander, ohne dass Pausen, finale Längung oder Laryngarisierung festzustellen sind, kann die Syntax als Segmentierungsmerkmal herangezogen werden. Des weiteren schließe ich extralinguistische Phänomene wie Lachen oder Husten als eigenständige IP aus, solche Phänomene können jedoch im Falle von sprachbegleitendem Lachen auch kurzem Räuspern auch Teil einer IP sein.

Intonationsphrasen werden nur in BeDiaCo_c annotiert, da sich nach der explorativen Auswertung von GECO-FP das Desiderat einer feineren Unterteilung der Dialogzugannotation ergeben hat (vgl. Abschnitt 5.2.1.2, 5.2.2, 5.2.2.2 und 5.2.2.3).⁹⁷ Tabelle B.5 im Anhang enthält die Annotationswerte der Intonationsphrasenannotation für BeDiaCo v. 1. Die Werte *ip* und *ipp* werden im verbleibenden Abschnitt mit Beispielen illustriert und diskutiert.

Abbildung 4.4 zeigt die Annotation einer IP, in der stille Pausen die IP nicht unterbrechen. Obwohl stille Pausen ein Indiz für zwei vorliegende IPs sind (Himmelman et al. 2018: 239; Swerts & Geluykens 1994), können sie doch auch innerhalb von IPs vorkommen. So wird die Intonationskontur in Abbildung 4.4 zwar durch Pausen unterbrochen, jedoch in den folgenden Wörtern wieder aufgegriffen und mit einem Tonakzent auf *keiner* und einem Grenzton auf *mehr* und dessen segmentaler Längung abgeschlossen. Wenn die Kontur allerdings durch einen Tonhöhenreset unterbrochen wird, liegen wahrscheinlich zwei IPs vor (Himmelman et al. 2018: 239).

Besonders bei Füllpartikeln ist die Entscheidung, ob sie selbstständige IPs bilden oder in eine IP integriert sind, nicht trivial. Füllpartikeln werden nur dann als eigenständige IP annotiert, wenn sie sich klar intonatorisch (durch Bewegung) und durch eine stille Pause von mehr als 50 ms visuell und perzeptiv unterscheiden lassen; ansonsten werden sie in die Spanne der vorangehenden IP aufgenommen. Abbildung 4.5 zeigt ein *ähm*, nach welchem sich eine längere Pause anschließt. Es ist durch eine stille Pause von 75 ms visuell vom vorangehenden Sprachmaterial abtrennbar. Die segmentale Längung und Laryngalisierung des /n/ in *bisschen* ist ein weiteres Indiz für die Abgeschlossenheit der vorangehenden prosodischen Konstituente. Somit liegen hier zwei IPs vor.

⁹⁷So wird einerseits eine genauere Aussage über die Position von Füllpartikeln innerhalb von Dialogzügen möglich, andererseits können eventuell vorhandene prosodische Effekte abgefangen werden.

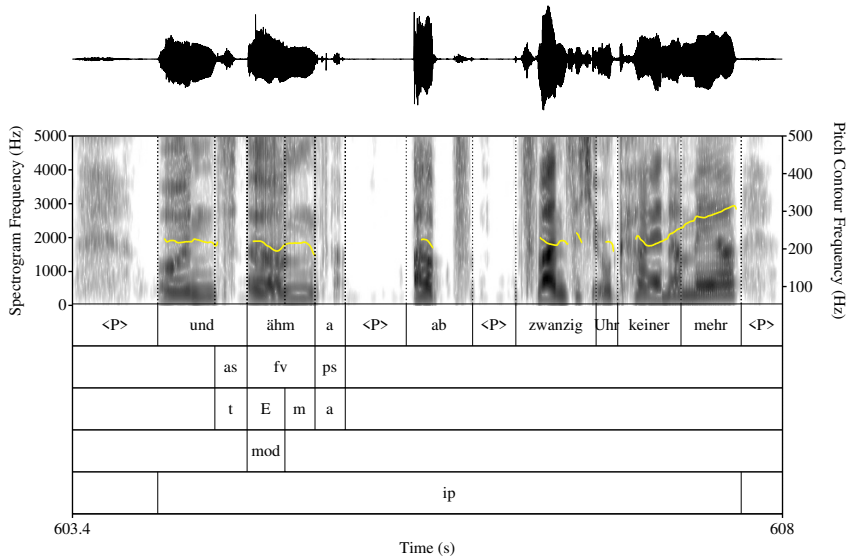


Abbildung 4.4: Durch Pausen unterbrochene Intonationsphrase in BeDiaCo_c *frei_f5f6_ch1*. Ohne Pausen ergibt sich eine durchgehende Intonationskontur (markiert mit der gelben Linie im Spektrogramm) mit einem Grenztone auf *mehr*.

Für Pausenschwellwerte unter 50 ms wird *ähm* nicht als eigenständige IP annotiert, sondern als rechter Rand und Teil der vorangehenden IP. Allerdings haben sich während der Annotation Fälle ergeben, in denen diese Füllpartikeln in IP-finaler Position keinen Grenztone zeigen. Wie in Abbildung 4.6 liegt der Grenztone der IP auf dem Wort davor, in diesem Fall auf *kann*. Der in Praat gemessene Unterschied der Grundfrequenz beträgt 4,3 Hz zwischen den letzten drei regelmäßigen Schwingungen von *kann* vor der Laryngalisierung und den ersten drei regelmäßigen Schwingungen von *ähm* nach der Laryngalisierung. Hier ist also ein tonaler Sprung zu verzeichnen, der eigentlich eine IP-Grenze impliziert. Auch die Laryngalisierung selbst ist ein Indiz für eine Grenze. Die hier zu treffende Entscheidung ist für eine spätere Auswertung von hoher Bedeutung. Entscheidet man sich für eine IP, so steht *ähm* IP-final. Entscheidet man sich für zwei IPs, so konstituiert *ähm* direkt im Anschluss an eine vorangehende IP eine neue IP. Es ergeben sich also unterschiedliche Struk-

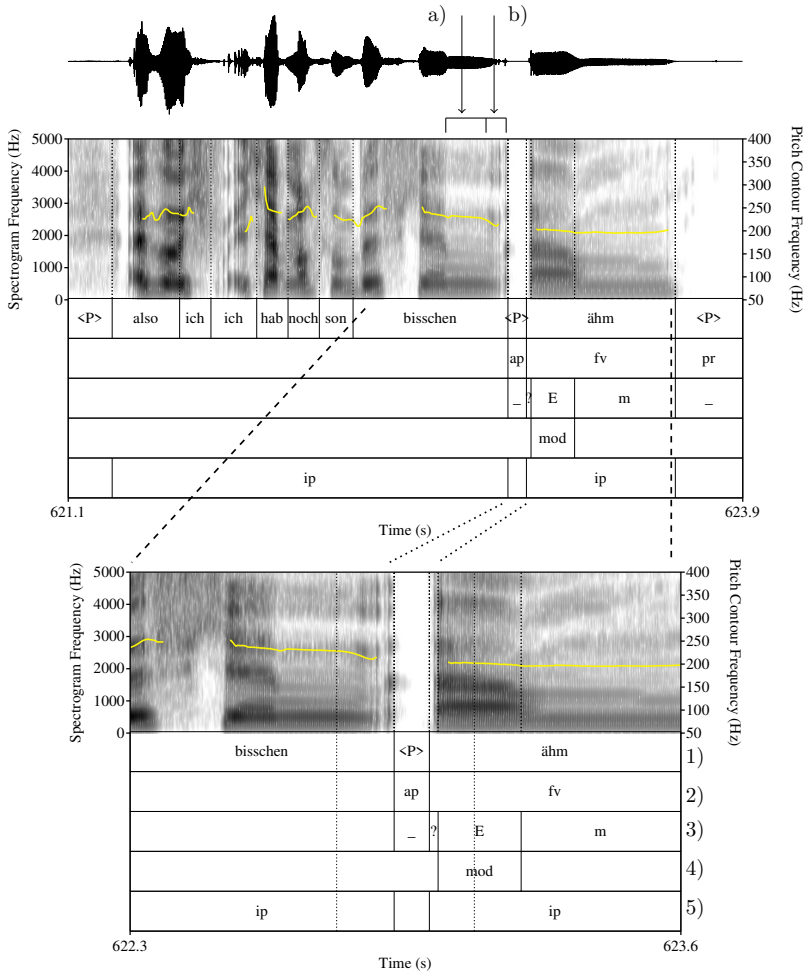


Abbildung 4.5: Ausschnitt aus BeDiaCo *frei_ffm1_ch1*, mit lokaler Vergrößerung im unteren Bild und diplomatischer Transliteration (Ebene 1), Füllpartikelannotation (Ebene 2–4), Intonationsphrasen (Ebene 5) und Intonationskontur (gelb). Die stille Pause von 75 ms vor *ähm* sowie die Längung (a) und Laryngalisierung (b) des /n/ in *bisschen* sind Merkmale einer IP-Grenze.

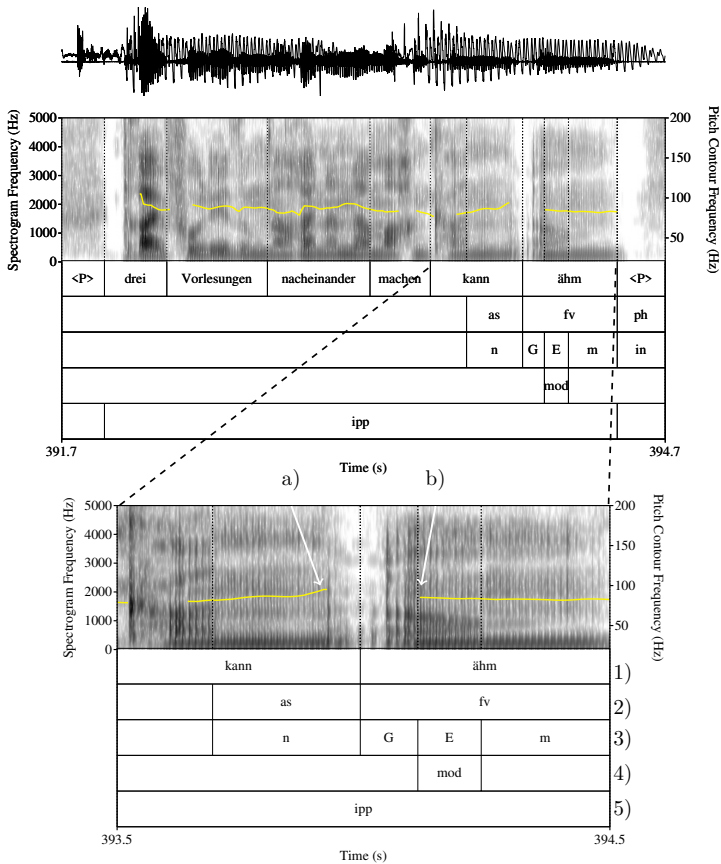


Abbildung 4.6: Ausschnitt aus BeDiaCo *frei_fm1_ch2*, mit lokaler Vergrößerung im unteren Bild und diplomatischer Transliteration (Ebene 1), Füllpartikelannotation (Ebene 2–4), Intonationsphrasen (Ebene 5) und Intonationskontur (gelb). Ebene 5 zeigt eine Intonationphrase mit Postposition (*ipp*). Position a) (weißer Pfeil) entspricht den letzten drei regelmäßigen Schwingungen (89,3 Hz) vor der Laryngalisierung, b) den ersten drei (85 Hz) nach ihr.

turen. Da die Entscheidung nicht eindeutig getroffen werden kann, wird eine neue Kategorie *ipp* vergeben. Wird also eine IP produziert, der nach einem Grenzton noch ein prosodisch merklich abgesetztes Wort ohne Pause größer als 50 ms folgt, welchem wiederum eine stille Pause oder Atmungspause folgt, so wird das abgesetzte Wort als Postposition zur IP hinzugezählt und die IP als *ipp* markiert. Diese Fälle können somit sowohl getrennt analysiert oder in eine generalisierte Auswertung aller IPs eingebunden werden.

Auch die Abgrenzung eigenständiger IPs von nachfolgenden IPs ist nicht trivial, hat jedoch große Auswirkungen auf die nachfolgenden Analyse. Abbildung 4.7 und 4.8 zeigen vier Beispiele einer Sprecherin für die Kategorisierung von Füllpartikeln als eigenständige IP oder als IP-initiale Position. In Abbildung 4.7 bilden die Füllpartikeln eigene IPs aus. Im oberen Bild von Abbildung 4.7 ist *ähm* trotz fehlender stiller Pause perzeptiv von *also* abgesetzt; das [m] von *ähm* hat eine Grundfrequenz (f_0) von etwa 183 Hz und zeigt am Ende Laryngalisierung, während die f_0 des Wortes *also* nach Beginn seiner regelmäßigen Schwingung etwa 189 Hz anzeigt. Auch im unteren Bild von Abbildung 4.7 hat das [m] von *ähm* eine f_0 von etwa 184 Hz, das nachfolgende *ich* etwa 202 Hz. Somit liegt in beiden Fällen der Neueinsatz einer höheren Intonationskontur vor. In Abbildung 4.8 ist die Intonationskontur in beiden Fällen durchgehend und fallend, und es treten keine Laryngalisierungen oder Pausen auf. Zusätzlich steigt die Intonationskontur nach den Füllpartikeln nicht merkbar an, die Füllpartikeln sind also in keiner Weise perzeptiv von dem nachfolgenden linguistischen Material abgesetzt. Diese Kriterien sprechen demnach nicht für die Annahme einer eigenständigen Intonationsphrase für die Füllpartikeln in Abbildung 4.8.

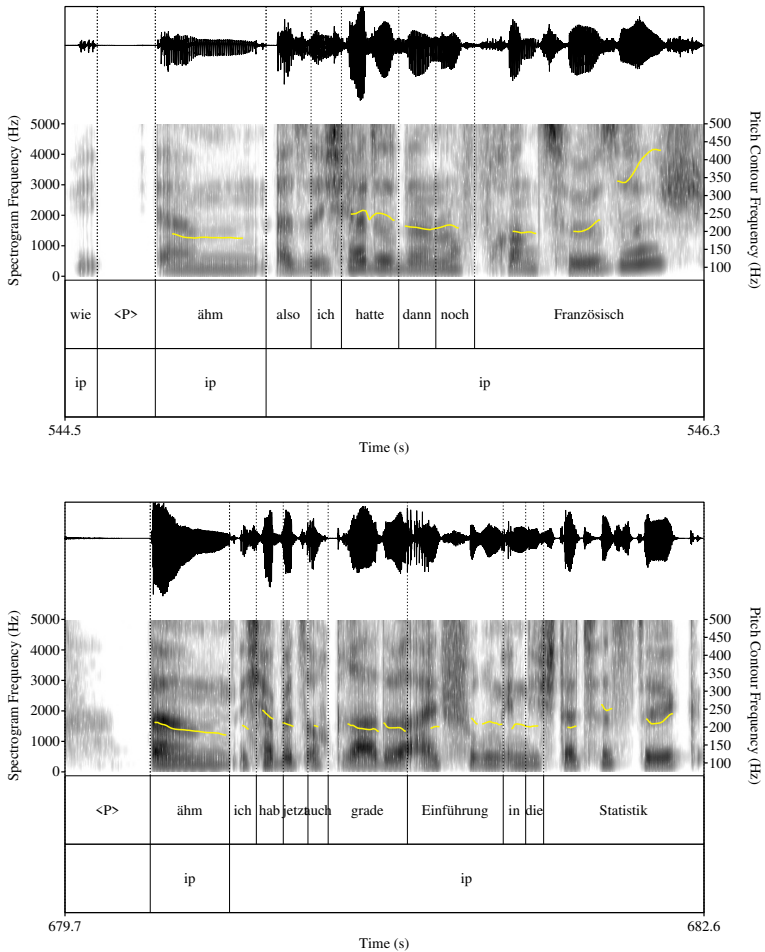


Abbildung 4.7: Ausschnitt aus BeDiaCo *frei_f3f4_ch1*, mit diplomatischer Transliteration (Ebene 1), Intonationsphrasenannotation (Ebene 2) und Intonationskontur (gelb). Die Füllpartikel bildet im oberen und unteren Bild eine eigenständige Intonationsphrase.

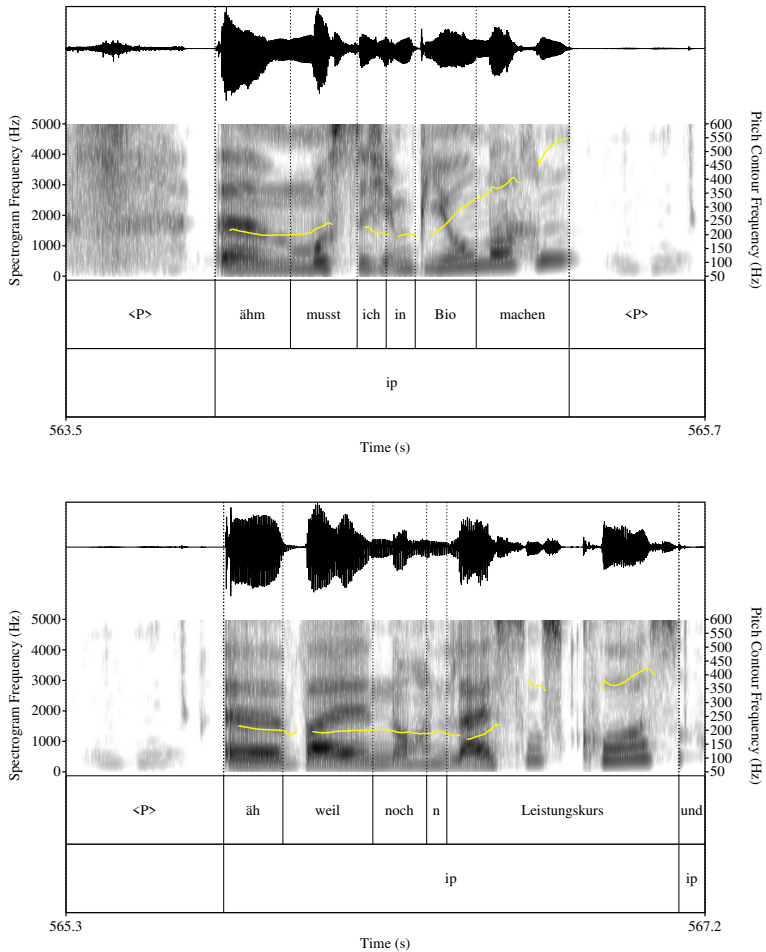


Abbildung 4.8: Ausschnitt aus BeDiaCo *frei_f3f4_ch1*, mit diplomatischer Transliteration (Ebene 1), Intonationsphrasenannotation (Ebene 2) und Intonationskontur (gelb). Die Füllpartikel steht im oberen und unteren Bild zu Beginn einer Intonationsphrase.

4.2.3 Dialogstrukturannotation

Als Dialogzüge werden hier strukturkonstituierenden Einheiten eines Dialoges verstanden, wie beispielsweise Fragen, Antworten oder Erzählungen. Nachdem in Abschnitt 3.2.3 gezeigt wurde, dass diese Strukturen mit Füllpartikeln in bestimmter Weise interagieren, also im Dialogkontext Muster ausbilden, wird für die weitere Quantifizierung ein Annotationsschema für Dialogstrukturen benötigt. Von Carletta et al. (1997, 1996) stammt ein Schema zur Dialogstrukturannotation im Rahmen aufgabenbasierter Dialoge. Diese Dialogzüge beschreiben funktionale Struktureinheiten in Dialogen, „different kinds of initiations and responses classified according to their purposes“ (Carletta et al. 1997: 14). Initiierende Züge haben gemeinsam, dass sie eine Reaktion oder Antwort erwarten (Carletta et al. 1996: 3). Beantwortende Züge folgen auf initiierende Züge. Mit diesem Schema lässt sich also die dyadische Struktur eines Dialoges abbilden. Im Folgenden erläutere ich die Segmentierung und Annotationswerte der Dialogstruktur mithilfe des von mir modifizierten Schemas von Carletta et al. (1997, 1996). Die Annotation erfolgt auf der Ebene *dia* in BeDiaCo_c beziehungsweise auf *func* in GECO-FP.

4.2.3.1 Segmentierung

Für Dialogzüge ist keine eindeutige Segmentierungsrichtlinie vorhanden („the move coders agree on how to segment a dialogue into moves“, Carletta et al. 1997: 25), was eine eigene Festlegung erforderlich macht. Wie in Abschnitt 3.2.3.3 schon erwähnt, ist es notwendig, sich für eine Segmentierungsgrundlage für die Annotation zu entscheiden. Hierbei könnte sich die Segmentierung sowohl an der minimalen Ausdehnung im Korpus orientieren, also an einer Tokenebene, oder an mittleren Ausdehnungen wie turnkonstruktionalen Einheiten oder Turns, oder als drittes an einer maximalen Ausdehnung.⁹⁸

Die explorative Anwendung in GECO-FP hat gezeigt, dass die Segmentierung keineswegs trivial ist. Problematisch sind besonders die qualitativen Unterschiede

⁹⁸Zwar wurden in der Konversationsanalyse *Redeübergabestellen* beziehungsweise *turnkonstruktionale Einheiten* (TCUs) als basale Einheiten eines Gesprächs postuliert, die Grenzen dieser Einheiten sind jedoch aufgrund mangelnder ‚harter‘ Definitionen oft unscharf (Selting 2000: 480) und sowohl von Syntax als auch von Prosodie abhängig (Selting 1996: 384). Im Prinzip könnte jeder syntaktische oder prosodische Abschluss (Sacks et al. 1974) eine mögliche Redeübergabestelle sein und damit eine TCU konstituieren.

zwischen den Annotationswerten, die verschieden große Segmentierungsgrundlagen hervorbringen. So sind Erzählungen typischerweise lang und werden als zusammenhängende Einheit wahrgenommen, die jedoch vom Gegenüber mit Backchanneleinheiten unterbrochen werden können. In diesen Fällen habe ich in GECO-FP nach der Backchanneleinheit eine neue Spanne für die fortgeführte Erzählung gesetzt – umgekehrt sind Backchanneleinheiten typischerweise kurz und liegen weit auseinander. Mehrere aufeinanderfolgende, aber vom Gegenüber durch eigene Beiträge unterbrochene Backchanneleinheiten wurden daher nicht als eine große Spanne markiert, sondern separat annotiert. Insgesamt ist die Segmentierungsgrundlage der Dialogstruktur in GECO-FP daher eher turnbasiert beziehungsweise der Segmentierung in größere Turneinheiten (Stirling et al. 2000: 114) nicht unähnlich.

Konzeptuell-methodisch sind hier jedoch einige Punkte problematisch. So ist die Bedeutung des Anfangs einer Struktureinheit nicht klar, wenn im Anschluss eine gleiche Struktureinheit folgt; im Prinzip müsste die erste Einheit einer Kette von gleichen Struktureinheiten gefunden werden, um Aussagen über ihren Anfang zu treffen; analog verhält es sich mit ihrem Ende. Für die Darstellung des gesamten Zuges wäre eine Aufaddierung der kleineren Ausdehnungen nötig. Weiterhin erlaubt die angewendete Strukturierung praktisch keinerlei quantitative Interpretation, da diese immer von der turnstrukturellen Beschaffenheit abhängt.

Die explorative Annotation ist demnach ein großer Gewinn für diese Arbeit, hat sie doch gezeigt, dass die Segmentierungsgrundlage für eine dialogstrukturelle Annotation hochgradig von der Beschaffenheit der Turnstruktur abhängt. Von dieser konfundierenden Variable muss Abstand genommen werden, um dann die Segmentierung auf eine methodisch strikt gleiche Weise zu behandeln – sie wird sonst weder einer korpuslinguistischen Definition von Segmentierung gerecht noch ist sie irgendwie quantitativ valide auswertbar.

Die Segmentierung in BeDiaCo_c erfolgt daher folgendermaßen. Annotiert wird die maximale Ausdehnung des Dialogzuges als Spannenannotation über alle in diesem Dialogzug enthaltenen Token. Somit gibt es keine Abfolgen gleicher Annotationswerte in der Ebene einer Versuchsperson. Damit bezieht sich die Segmentierung auf die strukturell-inhaltliche Ebene des Dialogs und nicht allein auf die Struktur des Redewechsels. Sie verbindet damit die Vorteile einer strukturellen Erfassung mit denen einer propositionalen Beschreibung. Dieses Vorgehen hat den Nachteil, dass

eine quantitative Analyse nicht intuitiv eine Menge an Dialogzügen eines Sprechers darstellt. Sie motiviert daher die zusätzliche prosodische Annotation von Intonationsphrasen, um so eine interpretierbare Quantifizierung je Dialogzug zu erhalten. Beispiel (5) wiederholt Beispiel (4) nun mit einer anderen Segmentierung für Dialogzüge (DZ, markiert als Rauten #). Intonationsphrasen (IP) stehen innerhalb von Klammern. Tabelle 4.3 stellt die hierarchische Beziehung zwischen Dialogstruktur- und Intonationsphrasensegmentierung und -annotation in Verbindung mit der Wortebene anhand Beispiel (5) schematisch in einem Mehr-Ebenen-Modell dar.

- (5) a. A: #(Wie findest du das Essen in der Mensa?)_{IP}#DZ Initiierung Frage
 b. B: #(äh ich hab ⟨P, 180 ms⟩ hier im Campus no nie in der Mensa gegessen weil ich eigentlich in *beep* bin)_{IP}
 c. A: #(mhm)_{IP}
 d. B: (an dem Campus dort)_{IP}
 e. A: (mhm)_{IP}
 f. B: (deshalb)_{IP} ⟨P, 860 ms⟩ (das bei uns is ziemlich gut aber ich hör immer dass so generell die Leute nich so zufrieden sin mit den Mensen)_{IP}#DZ Antwort
 g. A: (mhm)_{IP}#DZ Feedback
 h. #(also ich bin sehr)_{IP} ⟨P340 ms⟩ (positiv überrascht)_{IP}
 i. B: #(ja?)_{IP}
 j. A: (von der Mensa)_{IP} (also ich hab davor)_{IP} (mein Bachelor in *beep* gemacht)_{IP}
 k. B: (mhm)_{IP}
 l. A: (un da is die Mensa)_{IP} ⟨286 ms⟩ (überhaupt nich gut)_{IP}#DZ Initiierung
 m. B: (okay)_{IP}#DZ Feedback (BeDiaCo_c: a_f8f9, 5,6–31 s)

4.2.3.2 Annotationswerte

Meine Adaption des Dialogzugannotationschemas von Carletta et al. (1997) enthält eine dichotome Gliederung in initiiierende Züge (ERZÄHLUNG, ENTSCHEIDUNGSFRAGE, W-FRAGE, REDEBEREITSCHAFT) und responsive Züge (BACKCHANNEL, POSITIVE ANTWORT, NEGATIVE ANTWORT, UNKLARE ANTWORT, und KOMPLEXE ANTWORT).⁹⁹ Abbildung 4.9 gibt einen Überblick über die Entscheidungsfindung für eine bestimmte Kategorie in dem hier angewandten Dialogzugeschema. Das in GECO-FP

⁹⁹Weitere im ursprünglichen Schema von Carletta et al. vorhandene Kategorien wie INSTRUCT (Instruktion), ALIGN (Überprüfung der Übereinstimmung) und CHECK (Informationsbestätigung)

Tabelle 4.3: Schematische Darstellung der Dialogzüge (Ebene *dia*, Wert *DZ*) und Intonationsphrasen (Ebene *ip*, Wert *IP*) in Beispiel (5) mit jeweils erstem Wort der betreffenden Annotationsspanne auf der aussprachebasierten Transliterationsebene *dipl* über die Zeit *t*.

Ebene	Beispiele									
	5a	5b	5c	5d	5e	5f	5f	5g	5h	
A <i>dia</i>	DZ			DZ						DZ...
A <i>ip</i>	IP		IP		IP		IP		IP	IP
A <i>dipl</i>	Wie...		mhm		mhm		mhm		also...	
B <i>dia</i>	DZ									
B <i>ip</i>	IP		IP		IP		IP		IP	
B <i>dipl</i>	äh...		an...		deshalb		das...			
	Fortsetzung									
	5h	5i	5j	5j	5j	5k	5l	5m		
A <i>dia</i>	...DZ									
A <i>ip</i>	IP		IP		IP		IP		IP	
A <i>dipl</i>	positiv...		von...		also...		mein...		un...	
B <i>dia</i>	DZ		DZ		DZ		DZ		DZ	
B <i>ip</i>	IP		IP		IP		IP		IP	
B <i>dipl</i>	ja?		mhm		mhm		okay			

-----> t

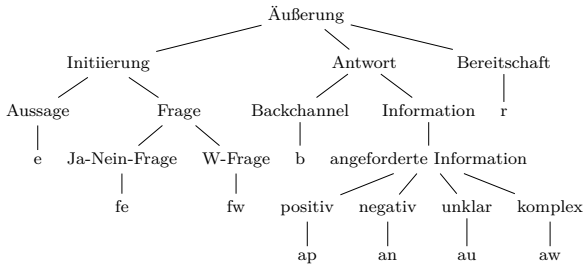


Abbildung 4.9: Entscheidungsbaum für die Dialogstrukturannotation.

verwendete Schema findet sich in Tabelle A.5 im Anhang. Nach der Exploration und Auswertung von GECCO-FP (vgl. Kapitel 5) mit dem Schema in Tabelle A.5 konnten nicht alle dort verwendeten Kategorien aufrechterhalten werden. Im Folgenden erläutere ich daher nur die in BeDiaCo_c annotierten Werte ausführlicher (vgl. hierzu die Richtlinien in Tabelle B.6 im Anhang).

ERZÄHLUNGEN (Annotationswert *e*) enthalten Informationen, die nicht direkt im Gespräch elizitiert werden („information that has not been directly elicited by the partner“, Carletta et al. 1997: 16). Dies sind Äußerungen, die nicht direkt von einer Frage des Gegenübers ausgelöst werden. Darunter zählen auch genauere Erläuterungen¹⁰⁰ oder Spezifizierungen, die auf die eigentliche Antwort einer Frage folgen, wie der Dialogzug in Beispiel (6c), der auf die Antwort in (6b) folgt. In Beispiel (6) steht *fw* für die Kategorie W-FRAGE, *aw* für KOMPLEXE ANTWORT und *e* für ERZÄHLUNG. ERZÄHLUNGEN können in meiner Adaption am Ende diskursmarkierende Partikeln aufweisen, die keine Frage evozieren, sondern kommentierend verwendet werden, wie *ne?*, *ja?*, *gell?* (vgl. Beispiel 9c).

- (6) a. A: #(wie bewertes du s Essen in der Mensa)_{ip}#*fw*
 b. B: #(äh überhaupt gar nicht gut)_{ip}#*aw*
 c. #(also ich ess auch nich mehr in der Mensa weil ich finde das is nich)_{ip} (also)_{ip}
 (P, 100 ms) (der Geruch un so viele Menschen ich mein des)_{ip} (P, 196 ms) (ess
 ich ja nich aber)_{ip} (des so für das Gesamtambiente und)_{ip} (P, 601 ms) (ähm)_{ip}
 ... #*e* (BeDiaCo_c: k_m10f12, 3,1–16,2s)
 Legende: #DZ# (IP) [Überlappung] (Pausen/etc.)

ENTSCHEIDUNGSFRAGEN (Annotationswert *fe*) sind Fragen, die eine positive oder negative Antwort erwarten lassen und mit *ja* oder *nein* beantwortbar wären, vgl. Beispiel (7a) und (7i). Merkmale von Entscheidungsfragen können eine steigende Frageintonation und Verberststellung sein. Im Deutschen können Fragen auch mit fallender Intonation realisiert werden (Peters et al. 2016) – falls in diesen Fällen keine Verberststellung vorhanden ist, wird die Struktur als ERZÄHLUNG gewertet. In Beispiel (7) steht *b* für BACKCHANNELING und *au* für UNKLARE ANTWORT.

für aufgabenfreie Dialoge nicht zu erwarten – sie sind zudem, wie auch die Kategorie CLARIFY (Erläuterung), auf die ich unten kurz weiter eingehe, anfällig für Verwechslung mit anderen Kategorien (Carletta et al. 1997: 26). Aus diesen Gründen schließe ich sie aus meiner Modifikation aus.

¹⁰⁰Erläuterungen (bei Carletta „CLARIFY“) sind Informationen, die über die eigentliche, strikte Beantwortung der Frage hinausgehen: „a reply to some kind of question in which the speaker tells the partner something over and above what was strictly asked“ (ebd.: 21).

- (7) a. A: #(gehst du überhaupt in die Mensa X)_{ip}#_{fe}
 b. B: #(also ich war hier vor Jahren das letzte mal in der Mensa)_{ip}
 c. A: #(okay)_{ip}#_b
 d. B: (weil ich bin geh auch nich hier zur Uni oder so)_{ip}#_{au} (lachen)
 e. A: #(ah okay)_{ip}#_b
 f. B: #([ne genau])_{ip}#_b
 g. A: #([wie kommst] du)_{ip} (wenn du nich hier zur Uni gehst wie kommst du dann überhaupt zu dem Experiment)_{ip}#_{fw}
 h. B: #([äh])_{ip}#_{aw}
 i. A: #([kennst du X] oder)_{ip}#_{fe}
 j. B: #(ja genau ich kenn X)_{ip}#_{ap}

(BeDiaCo_c: n_m15f13, 1,8–16,8 s)

Legende: #DZ# (IP) [Überlappung] (Pausen/etc.)

W-FRAGEN oder Set-Fragen (Annotationswert *fw*) sind Fragen mit steigender oder fallender Intonation, die ein ‚W-Wort‘ enthalten oder damit paraphrasierbar sind (*wer, wie, wo, wann, was, warum, weshalb, weswegen, wieso, worüber, woran, worauf, wohin* etc.). Diese Kategorie enthält alle Fragen, deren Antwort nicht *ja* oder *nein* sein kann, wie der gesamte Dialogzug beziehungsweise die ersten beiden IPs in Beispiel (7g).¹⁰¹

Responsive Züge bestehen in dem hier adaptierten Schema aus den Kategorien BACKCHANNEL, POSITIVE ANTWORT, NEGATIVE ANTWORT, KOMPLEXE ANTWORT und (von mir hinzugefügt) aus der Kategorie UNKLARE ANTWORT.

BACKCHANNEL-Züge (Annotationswert *b*) wie in Beispiel (7c), (7e) und (7f) sind Äußerungen, in denen ein Hörer signalisiert, dass er den Zug des Sprechers verstanden und akzeptiert hat (Carletta et al. 1997: 19). Carletta et al. nennen sie „ACKNOWLEDGE Moves“, sie ähneln aber vielmehr Backchanneleinheiten, wie beispielsweise kurzen Verbalisierungen (*mhm, aha, ja*), positiven oder negativen Bewertungen, kurzen Nachfragen oder Reformulierungen, Ausrufen oder versuchten Unterbrechungen seitens des Hörers (vgl. Heinz 2003: 1117; außerdem Schegloff 1982). Dies führt zu der auf den ersten Blick paradoxen Situation, dass ein Sprecher A in Beispiel (8a) den Backchannelzug *echt?* äußert, der von Sprecherin B in (8b) mit

¹⁰¹Hier bestünde auch die Möglichkeit, zwei Dialogzüge zu annotieren, eine trunkierte W-Frage (*wie kommst du*) und eine W-Frage (*wenn du nich hier zur Uni gehst wie kommst du dann überhaupt zu dem Experiment*). Allerdings sind trunkierte W-Fragen GECO-FP so selten, dass diese Kategorie für BeDiaCo_c nicht übernommen wurde.

einer UNKLAREN ANTWORT beantwortet wird. Tatsächlich wird das Schema durch diese Fokussierung auf die Struktur eines Dialogs überhaupt erst so flexibel, dass es sowohl das von Sprecher A produzierte Signal *echt?* auf der Ebene von Sprecher A als BACKCHANNELZUG erfassen als auch den daraufhin von B produzierten Dialogzug inhaltlich-relational als UNKLARE ANTWORT kennzeichnen kann.

- (8) a. A: #(*echt?*)_{ip}#_b
 b. B: #(weiß ich nicht)_{ip} (Einatmung, 232 ms) (bin eigentlich auch nicht so der Phonetiker)_{ip} (P, 374 ms) (also weiß nicht ob ich jemals da draußen sitzen werde)_{ip}#_{au}
 (BeDiaCo_c: k_m10f12, 107,2–112,8 s)
 Legende: #DZ# (IP) [Überlappung] (Pausen/etc.)

POSITIVE ANTWORTEN (Annotationswert *ap*) sind Züge, die auf eine Entscheidungsfrage mit einer Äußerung antworten, die einen positiven Gehalt hat – „a reply [...] that means ‚yes‘, however that is expressed“ (Carletta et al. 1997: 20), vgl. Beispiel (7i).

NEGATIVE ANTWORTEN (Annotationswert *an*) sind Züge, die auf eine Entscheidungsfrage mit einer Äußerung antworten, die einen negativen Gehalt hat – „a reply [...] that means ‚no‘“ (ebd.). Problematisch an positiven und negativen Antworten ist ihr rechtes Ende, besonders um diese Kategorie von der Kategorie ERZÄHLUNG abzugrenzen. Ich definiere daher POSITIVE und NEGATIVE ANTWORTEN so, dass sie linguistisches Material beinhalten können, welches noch mit „ja“ oder „nein“ paraphrasierbar ist, vgl. Beispiel (7b) und insbesondere die Fortsetzung der Antwort in (7d), die für sich genommen auch paraphrasierbar ist mit *nein, weil weil ich bin geh auch nicht hier zur Uni oder so*.

KOMPLEXE ANTWORTEN (Annotationswert *aw*) sind Dialogzüge, die auf eine Frage mit einer Äußerung antworten, die nicht ‚ja‘ oder ‚nein‘ oder ‚weiß nicht‘ bedeutet. Typischerweise sind das Antworten auf W-FRAGEN wie der Dialogzug in Beispiel (5b), (5d) und (5f). Strukturell können auch abgebrochene Züge, die nach W-Fragen stehen, als *aw* markiert werden.

UNKLARE ANTWORTEN (Annotationswert *au*) sind bei Carletta et al. nicht vorgesehen und werden von mir hinzugefügt. Eine Äußerung wird als UNKLARE ANTWORT markiert, wenn eine Sprecherin oder ein Sprecher eine Frage nicht mit Sicherheit beantworten kann, die Antwort also keinen klar positiven oder negativen Gehalt hat, und die Antwort zugleich keine Reaktion auf eine W-FRAGE ist (vgl. Beispiel 7d und 8b).

Als dritte große Dialogstrukturkategorie geben Carletta et al. die sogenannten „READY moves“ an, „moves that occur after the close of a dialogue game and prepare the conversation for a new game to be initiated“ (Carletta et al. 1997: 22). Ich kategorisiere diese anders, nämlich als initiiierende Züge, und nutze diese Kategorie somit für konversationseröffnende und -schließende Strukturen, wie *hallo*.

Der mithin schwierigste Punkte der Dialogstrukturannotation ist die Abgrenzung responsiver von initiiierenden Zügen. In Beispiel (9) ist dies noch relativ einfach. Während (9b) die Antwort auf eine vorangehende Entscheidungsfrage von B (*ach kennst du auch X?*) darstellt, erklärt Sprecher A direkt anschließend in (9c) den Hintergrund seiner Anwesenheit im Experiment, die aber nicht mehr direkt auf die Entscheidungsfrage von A beziehbar ist. In Beispiel (5b)–(5f) bezieht sich jede Intonationsphrase von B weiterhin direkt auf die Frage. In Beispiel (6) gibt B in ihrer ersten Intonationsphrase in (6b) eine Antwort auf eine W-Frage, geht dann aber in der nächsten IP über in eine initiiierende Erzählung (6c), die sich nicht mehr direkt auf die Frage beziehen lässt.

- (9) a. A: #(okay)_{ip}#_b
 b. #(jaja genau)_{ip}#_{ae}
 c. #(also ich hab nen paar Kurse mit X [zusammen und])_{ip} (äh genau wir ham uns im)_{ip} (Einatmung, 300 ms) (ersten Mastersemester kennengelernt))_{ip} (und irgendwie hatte sie mich einfach per Email)_{ip} (Einatmung, 400 ms) (angehauen)_{ip} (und hatte gefragt ob wir ähm oder ob ich bereit wäre an nem Experiment teilzunehmen)_{ip} (und äh)_{ip} (P, 330 ms) (ja)_{ip} (P, 746 ms) (ne)_{ip}#_e

(BeDiaCoC: n_m15f13, 17,7–35,1 s)

Legende: #DZ# (IP) [Überlappung] (Pausen/etc.)

Das Dialogstrukturschema kann nur eine grobe Strukturierung des Dialoges leisten. Heuristisch werden daher keine diskontinuierlichen Kategorien wie beispielsweise die Abfolge Antwort–Erzählung–Antwort annotiert. Sobald das Sprachmaterial als nicht mehr direkt elizitiert gilt beziehungsweise nicht mehr als Antwort auf eine W- oder Entscheidungsfrage paraphrasierbar ist, tritt der Zustand einer Erzählung ein. Das folgende Beispiel (10) macht dies deutlich. Dort antwortet B auf die Entscheidungsfrage von A in (10a) mit einem responsiven Zug in den Zeilen (10b, 10d, 10f) und (10h). Mit dem nachfolgenden erzählerischen Kommentar ab (10j) kann auch ungeachtet möglicher Retraktionen auf die Frage keine responsive Annotati-

onskategorie mehr vergeben werden. Eckige Klammern kennzeichnen überlappende Sprache.

- (10) a. A: #(und kann man danach dann Psychotherapeut werden)_{ip}#_{fe}
 b. B: #(ja)_{ip} (also man müsste man im (P, 90 ms) Master)_{ip} (P, 180 ms) (bestimmte Module machen)_{ip}
 c. A: #(mhm)_{ip}
 d. B: (zur zur Psychotherapie)_{ip} (Einatmung, 540 ms) (un dann muss man noch ne Ausbildung (P, 400 ms) machen)_{ip}
 e. A: (mhm)_{ip}
 f. B: (die dauert)_{ip} (P, 400 ms) (ich glaube drei Jahre)_{ip} (P, 380 ms) (un da verdient man au nich wirklich oder muss teilweise was draufzahlen deshalb)_{ipp} (Einatmen, 482 ms) (is das n bisschen)_{ip}
 g. A: (mhm)_{ip}
 h. B: ([nich so lukrativ])_{ip} (ja)_{ip}#_{ap}
 i. A: ([schwierig])_{ip}
 j. B: #(deshalb will ich das auch erstmal nich machen)_{ip}
 k. A: (mhm)_{ip}
 l. B: (aber das (P, 220 ms) ka man halt jederzeit noch nachholen also)_{ip} (auch wenn man sagt man arbeitet erstmal fünf Jahre oder so)_{ip} (und macht dann noch die Ausbildung)_{ip}
 m. A: (mhm)_{ip}#_b
 n. B: (des is au noch okay)_{ip}#_e

(BeDiaCo.c: a_m15f13, 154–188 s)

Legende: #DZ# (IP) [Überlappung] (Pausen/etc.)

Zusammenfassend werden die dyadischen Dialogzüge zweier Personen sowohl strukturell als auch inhaltlich informiert annotiert. Die Exponenten der Annotation und damit die zugrundeliegenden Struktureinheiten des Dialogs werden nicht *a priori* festgelegt, sondern emergieren sowohl strukturgestützt also auch inhaltlich *bottom-up* aus den interaktionalen Sprachhandlungen des Dialogs. Die Ziel der Dialogzugannotation ist somit, eine möglichst konsistente und nicht weiter unterstrukturierte Abfolge der Äußerungen zu erfassen.

4.3 Qualität der Annotationsrichtlinien

4.3.1 Füllpartikeln

„Counting filled pauses is about as simple as producing them.“ (Christenfeld & Creager 1996: 1235)

Obwohl das Zitat von Christenfeld & Creager suggeriert, dass das Zählen von Füllpartikeln einfach sei, spricht einiges gegen diese Behauptung. Trotz genauen Hörens werden häufig nicht alle Füllpartikeln bemerkt (vgl. Lickley & Bard 1998). Nach der Identifizierung einer Füllpartikel liegt die weitere Schwierigkeit in der Klassifizierung als nicht-glottale oder glottale Füllpartikel.

Zur Prüfung der Annotationsrichtlinien für die Identifizierung und Segmentierung von Füllpartikeln wird ein Inter-Annotator-Agreement-Experiment mit den Richtlinien in Anhang C.1 durchgeführt. Verwendet wird eine fünfminütige Stichprobe eines Dialogs aus GECO-FP. Die Tonspuren der Sprecherinnen werden nacheinander als Monosignal annotiert. Die Annotation erfolgt in Praat (Boersma 2001), das Abhören des Signals über Kopfhörer (Sennheiser HD 449). Die Annotatoren können sich visuell am Signal und an der in GECO-FP enthaltenen, unkorrigierten Ebene `word` orientieren. Zur Berechnung von Cohens κ (Cohen 1960) wird bei Überlappungen von Werten eine synthetische Trennung in Einzelwerte vorgenommen.

Getestet wurde die Übereinstimmung zwischen der Annotation von Füllpartikeln (nicht-glottale und glottale Formen) und keinem Vorkommen. Clicks sind ein Sonderfall und werden aus den Daten entfernt, da unklar ist, ob sie von der Füllpartikeldefinition abgedeckt werden. Die Übereinstimmung beider Annotatoren für die vergebenen Werte *fv* (nicht-glottale Form), *fg* (glottale Form) und *none* (keine Form vorhanden) liegt bei 75,93%, mit einem κ -Wert nach (ebd.) von $\kappa = 0,56$ ($z = 5,5, p < 0,001$, Token = 54), berechnet mit dem R-Paket *irr* (Gamer et al. 2012). Damit liegt die von κ gemessene Übereinstimmung nach Landis & Koch (1977: 165) in einem moderaten Bereich.

Das Erkennen und Annotieren von Füllpartikeln ist demnach keine triviale Aufgabe.

4.3.2 Intonationsphrasen

Studien gesprochener Sprache haben oftmals Schwierigkeiten mit der Annotation von Intonationsphrasen („highly problematic and unreliable“, Plag et al. 2017: 194), und verwenden bevorzugt andere linguistische Ebenen um die Daten prosodisch auszuzeichnen, wie beispielsweise syntaktische Phrasen (Plag et al. 2017; aber vgl. Himmelmann et al. 2018). Zur Evaluierung der in Abschnitt 4.2.2 beschriebenen Heuristik zur Annotation von Intonationsphrasen (IPs) wird ein Inter-Annotater-Agreement-Experiment mit 7,5 min der Versuchsperson f8 im freien Dialog von BeDiaCo (*a_f8f9_ch1*, 84,59–507,61 s) neben mir von einer anderen Annotatorin anhand der Richtlinien in Anhang C.2 annotiert. Zur Familiarisierung mit den Richtlinien wurde eine gemeinsame Übungssitzung mit den ersten 60 Sekunden des Dialogs durchgeführt. Die Annotation erfolgt in Praat (Boersma 2001), das Abhören des Signals über Kopfhörer (Sennheiser HD 449). Der Test fand statt, bevor der Annotationswert *ipp* hinzugefügt wurde.

Evaluert wird, ob die zugrundeliegende Wortsegmentierung identisch oder abweichend voneinander annotiert ist. Aufgrund der möglichen abweichenden Spannenannotationen der beiden Rater sowie nur eines einzigen Wertes kann keine Evaluation der vergebenen Werte auf einer vorgegebenen Tokenisierung getroffen werden. Als Lösung werden daher diejenigen Wortgrenzen der schon existierenden und beiden Ratern vorliegenden orthographischen Ebenen herangezogen und für jeden Rater eine Tabelle erstellt, in der festgehalten ist, ob eine Intonationsphrasengrenze mit der Grenze eines Wortes übereinstimmt. In den Fällen, in denen sich beide Rater dafür entschieden haben, an einer Wortgrenze eine IP zu beginnen oder zu enden, stimmen sie überein. Ebenso stimmen sie in allen Fällen überein, in denen sie sich beide zu keiner IP-Grenze entschieden haben. In allen anderen Fällen stimmen sie nicht überein.

Die Übereinstimmung beider Annotatoren liegt bei 95%, mit einem κ -Wert nach Cohen (1960) von $\kappa = 0,9$ ($z = 25,68, p < 0,001$, Token = 820), berechnet mit dem R-Paket *irr* (Gamer et al. 2012). Die mit κ gemessene Übereinstimmung kann nach Landis & Koch (1977: 165) als „Almost Perfect“ bezeichnet werden.

4.3.3 Dialogzüge

Zur Prüfung der Annotationsrichtlinien für die Identifizierung und Segmentierung von Füllpartikeln wird ein Inter-Annotator-Agreement-Experiment mit den Richtlinien in Anhang C.3 durchgeführt. Verwendet wird der BeDiaCo-Dialoges *frei_a_f8f9* (ca. 14 min). Die Bearbeitung erfolgt stereo, d. h. beide Tonspuren werden eingeblendet, um die dialogische Struktur besser verstehen zu können. Die Annotation erfolgt in Praat (Boersma 2001), das Abhören des Signals über Kopfhörer (Sennheiser HD 449). Zur Orientierung dienen die aussprachebasierte Transliteration in der Ebene *dipl* sowie die Intonationsphrasenannotation in der Ebene *ip*. Mit der Annotatorin wurden zudem anhand einer Übungssitzung von 50 Sekunden eines anderen Dialoges die Annotationswerte für die Dialogzüge exemplifiziert. Zur Berechnung von κ mussten bei Überlappungen von Werten eine synthetische Trennung in Einzelwerte vorgenommen werden.

Die Übereinstimmung beider Annotatoren liegt bei 81,65%, mit einem κ -Wert nach Cohen (1960) von $\kappa = 0,75$ ($z = 21,9, p < 0,001$, Token = 218), berechnet mit dem R-Paket *irr* (Gamer et al. 2012). Die mit κ gemessene Übereinstimmung ist nach Landis & Koch (1977: 165) „substanziell“.

Zusammenfassend ergeben die Richtlinien für die Annotationen moderate (Füllpartikeln), substanzielle (Dialogzüge) und fast perfekte (Intonationsphrasen) Übereinstimmungen. Bezüglich der Füllpartikeln besteht somit durchaus die Gefahr, trotz sorgfältigen Abhörens der Dialoge doch einige potenzielle Kandidaten zu übersehen. Für Dialogzüge besteht die Schwierigkeit eher in der Zuordnung der Äußerungen zu einem spezifischen Zug. Das Erkennen und die Annotation von Intonationphrasen erzielte die höchste Reliabilität.

4.4 Abfrage und phonetische Analyse

Die Extraktion aller Annotationen sowie der akustischen Parameter erfolgt aus den vorhandenen Korpora, indem diese zuvor mithilfe von *emuR* (Winkelmann et al. 2016) in *R* (R Core Team 2018) in eine EMU-Datenbank umgewandelt werden (Casidy & Harrington 2001; Harrington 2010; Winkelmann et al. 2017).

4.4.1 Dauer und Sprechgeschwindigkeit

Die Dauer von Segmenten wird über den Zeitstempel der linken und rechten Intervallgrenzen ermittelt. Die Sprechgeschwindigkeit wird in Silben pro Sekunde (σ/s) gemessen, exklusive stiller Pausen, Atmungspausen und extralinguistischer Geräusche wie Lachen, Husten, etc. Silben werden mit dem R-Paket *silly* gezählt (Michalke 2017).

Um die Dauer von Füllpartikeln über verschiedenen Versuchspersonen hinweg vergleichen zu können, bedarf es einer Normalisierung. Da die Sprechgeschwindigkeit (in σ/s) berechnet wird, kann diese als normalisierende Größe mit der Füllpartikeldauer (in s) multipliziert werden. Je kleiner der auf diese Weise errechnete Wert (gemessen in Anzahl Silben) ist, desto langsamer wurde die Füllpartikel artikuliert, beziehungsweise desto länger hat sie gedauert, und umgekehrt.

4.4.2 Grundfrequenz

Die Grundfrequenz (f_0) wird mit dem R-Paket *wrassp* (Bombien et al. 2018) und der Funktion *ksvF0()* mit geschlechtsspezifischen Einstellungen in R berechnet (Bereich von 80–640 Hz für Frauen, 50–400 Hz für Männer)¹⁰² und als Track in der EMU-Datenbank gespeichert. Anschließend kann die Grundfrequenz an einer bestimmten Stelle mit *emuR* ausgelesen werden. NA-Werte werden mit dem R-Paket *zoo* (Zeileis & Grothendieck 2005) interpoliert.

4.4.3 Formanten

Da in der Annotation aufgrund der Ähnlichkeit von zentral produzierten Vokalen keine ohrenphonetische Transkription verwendet wird, messe ich Formanten, um die Vokalqualität zu ermitteln. Hierzu können entweder alle Messpunkte der Trajektorie innerhalb eines annotierten Intervalls gemittelt werden oder der Formant im „quasi-stationären ‚steady state‘ Vokalteil“ (Pompino-Marschall 2009: 125) gemessen werden – beide Methoden verwende ich zu unterschiedlichen Zeitpunkten der Analyse.

Formanten werden mit dem Formantentracker von Praat (Boersma 2001) über das R-Paket *PraatR* (Albin 2014) mithilfe des Skripts *PraatToFormants2AsspDataObj*

¹⁰²<https://www.rdocumentation.org/packages/wrassp/versions/0.1.8/topics/ksvF0>, besucht am 23.08.2019.

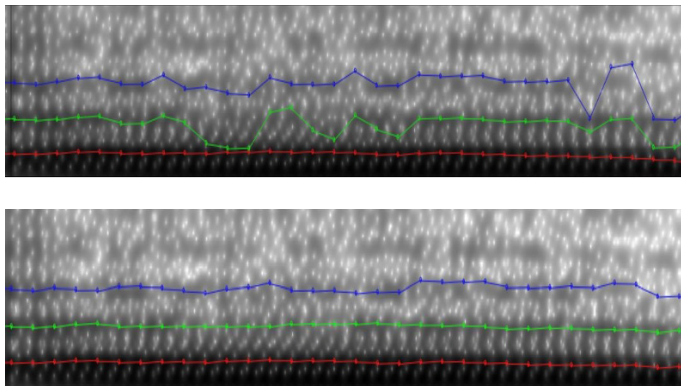


Abbildung 4.10: Beispiel für einen Vokal von 200 ms in *ähm* mit F_1 (rot), F_2 (grün) und F_3 (blau) nach der Formantenermittlung durch Praat (oberes Bild) und nach der manuellen Korrektur (unteres Bild), in GECO-FP, Sprecherin D, Dialog multi_C-D_right, Sek. 4,86–5,86.

(Winkelmann 2015) mit geschlechtsspezifischen Einstellungen ermittelt und als Track in der EMU-Datenbank gespeichert (nach Burg, maximale Anzahl von fünf Formanten, obere Schwelle für den maximalen Formanten bei 5500 Hz für Frauen und 5000 Hz für Männer, Fensterlänge von 0,025 s, Vorempfase von 50 Hz). Für alle in der Analyse relevanten Vokalsegmente werden die ersten drei Formanten manuell in der EMU-Datenbank korrigiert (vgl. Abbildung 4.10). Dabei werden klare Ausreißer an die umgebende Trajektorie angepasst.

Nach der Bereinigung können die Formanten an den für die Forschungsfrage relevanten Stellen in der Datenbank gefunden und extrahiert werden. Zur Vergleichbarkeit der Sprecherinnen und Sprecher untereinander werden immer auch ihre ermittelten Referenzvokale (vgl. Abschnitt 4.4.3.1 und 4.4.3.2) auf der `vowel1`-Ebene extrahiert und alle Werte einer Versuchsperson sodann `sprecher-intrinsisch`¹⁰³, `vokal-extrinsisch`¹⁰⁴ und `formant-intrinsisch`¹⁰⁵ lobanovnormalisiert (Lobanov 1971). Hierbei wird die häufig in der Literatur verwendete und leicht vom Original abweichende

¹⁰³Jeder Sprecher und jede Sprecherin wird für sich normalisiert.

¹⁰⁴Es wird über Formantdaten mehrerer von einem Sprecher oder einer Sprecherin geäußelter Vokal normalisiert.

¹⁰⁵Es wird nur innerhalb eines Formanten normalisiert.

Formel verwendet, welche mit der Standardabweichung statt dem quadratischen Mittel arbeitet, um den normalisierten Formantwert F_i^N zu ermitteln (vgl. Flynn 2011): $F_i^N = \frac{F_i - \mu_i}{\sigma_i} - 1$ ($i = \{1, 2, 3\}$), wobei N für normalisierte Werte steht und i für den ersten, zweiten oder dritten Formanten.

4.4.3.1 Referenzvokale in GECO-FP

Auf der Ebene `vowel` annotiere ich zur Verortung der Füllpartikeln im Vokalraum einer Versuchsperson und somit einer besseren interindividuellen Vergleichbarkeit alle 15 Monophthonge des Deutschen (falls vorhanden) und zusätzlich die beiden Reduktionsvokale [ə] und [ɐ] in möglichst klar und eindeutig artikulierten Kontexten mit mindestens zwei Vorkommen. Eine vokal-extrinsische Normalisierung, wie ich sie verwenden möchte, benötigt idealerweise alle Vokale des Vokalsystems eines Sprechers (Adank et al. 2004: 3105). F_1 , F_2 und F_3 aller Referenzvokale wurden manuell in der EMU-Datenbank überprüft und gegebenenfalls korrigiert.

4.4.3.2 Referenzvokale in BeDiaCo

BeDiaCo_c beinhaltet eine je Versuchsperson randomisierte Wortliste mit allen 15 Monophthongen des Deutschen in betonter Position im Silbennukleus sowie den beiden Reduktionsvokalen in wortfinaler unbetonter Position in offener Silbe, die vor und nach dem Experiment einzeln vorgelesen wird (vgl. Tabelle 4.4). Zusätzlich enthält die Liste folgende Wörter, deren erste Silbe eine orthographische Ähnlichkeit zu *äh* und *ähm* aufweist: *ähnlich*, *Äther* und *emsig*, *Ämter*.

4.4.4 Phonationsart und Intensität

Die Distribution der auf der Annotationsebene `phon` in BeDiaCo_c und `coart` in GECO-FP perzeptiv kategorisierten Werte wird in Abhängigkeit von den sequenziellen und hierarchischen Kontextkategorien ausgewertet. Zur Annotation habe ich mich am visuellen Signal orientiert (vgl. die Annotationsrichtlinien in Tabelle A.4 und Tabelle B.4). Phonationsart kann zwar perzeptiv erfasst werden, was aber auch hochgradig subjektiv ist. Eine objektivere Annäherung sind Maße wie der akustische Öffnungsquotient für Phonationsart oder die spektrale Neigung für Intensität. Aufgrund der erheblichen Herausforderungen bei der Aggregation und Evaluierung

Tabelle 4.4: Wortliste in BeDiaCo v. 1. Alle Wörter sind in den Kontext „Sage X bitte“ eingebettet.

Onsetartikulationsort der Ultima	bilabial		palatal		velar	
Final mit [ə]	piepe	[i:]	Güte	[y:]	bücke	[ɥ]
	Bube	[u:]	Beete	[e:]	Böcke	[œ]
	Kippe	[ɪ]	böte	[ø:]	Pocke	[ɔ]
	Puppe	[ʊ]	Bote	[o:]	Tage	[a:]
			bäte	[e:]	packe	[a]
		Kette	[ɛ]			
Final mit [ɐ]	Pieper	[i:]	Güter	[y:]	Bäcker	[ɛ]
	Tupper	[ʊ]	Puder	[u:]	Höcker	[œ]
	Kaper	[a:]	Mütter	[ɥ]	Packer	[a]
	Geber	[e:]	Köder	[ø:]	Kicker	[ɪ]
			Toter	[o:]		
			Täter	[e:]		
		Dotter	[ɔ]			

phonetischer Maße aus nichtkontrolliertem linguistischen Material, wie es hier als Spontansprache vorliegt, kann solch eine signalbasierte Auswertung der Phonationsart und der Intensität jedoch nicht mehr in diese Arbeit einfließen. Im Folgenden werde ich in einem Exkurs eine solche Auswertung an einem Beispiel problematisieren.

Der akustische Öffnungsquotient ist die Differenz zwischen den Amplitudenmaxima der ersten und zweiten Harmonischen in dem Energiespektrum eines Lautes (H_1-H_2). Hierzu vergleicht man die Amplitude der Grundfrequenz (die gleichzeitig die erste Harmonische H_1 ist) mit der Amplitude höherfrequenter Harmonischer (Gordon & Ladefoged 2001: 397), indem man ihre Werte subtrahiert. Die Differenz ist für glottale Laute eher positiv und für behauchte Laute eher negativ, während sich modale Laute dazwischen befinden (ebd.: 398). Die spektrale Neigung (Engl. ‚spectral tilt‘) gibt an, wie stark die Intensität in höheren Frequenzen abfällt (ebd.: 397). So kann die Energiedifferenz zwischen H_1 und A_3 als Maß für die Lautstärke herangezogen werden (Mooshammer 2010). Je größer H_1-A_3 , desto leiser ist das Signal. Dieses Maß ist reliabler als am Mikrofon gemessene Dezibelwerte, da diese stark von der Nähe der Versuchsperson zum Mikrofon abhängen.

Die Dezibelwerte der Harmonischen sind in einem DFT-Spektrum (Diskrete Fouriertransformation) als Maxima ablesbar, die Dezibelwerte der Formanten werden

als Maxima einer spektralen Hüllkurve mittels LPC-Analyse (Linear Predictive Coding) ermittelt (beides über das R-Paket *emuR*). Zur Ermittlung der Maxima der Harmonischen verwende ich den Algorithmus *findPeaks*.¹⁰⁶ Dieser arbeitet mit der Variablen m , die bestimmt, wie lokal die Maxima ausgewählt werden; diese Justierung ist aber oft nicht für alle Daten gleich anwendbar. Abbildung 4.11 zeigt die DFT-Spektren und die spektralen Hüllkurven für eine vokalische und eine vokalischnasale Füllpartikel mit $m = 2, 7$ und einer unteren Schwelle von 130 Hz, was unter dem normalen Grundfrequenzbereich von Frauen liegt. Die vom Algorithmus errechneten Maxima sind rot markiert. Im oberen Bild (*äh129*) wird aufgrund der Schwelle von 130 Hz jedoch H_1 nicht richtig erkannt, da H_1 hier der Grundfrequenz von ca. 110 Hz entspricht. Folglich entspräche eine aus den ersten beiden Maxima des Algorithmus errechnete Differenz eben nicht H_1-H_2 . Gleiches gilt für die Berechnung der Intensität (H_1-F_3). Die untere Schwelle für die Grundfrequenz müsste also nach einer visuellen Vorkontrolle der Spektren in mehreren Schritten angepasst werden, um die gerade für glottalisierte Sequenzen häufig tiefere Grundfrequenz abzufangen.

Im unteren Bild von Abbildung 4.11 (*äh163*) findet der Algorithmus nicht nur die Maxima von H_1 und H_2 , sondern auch ein Maximum dazwischen. Ohne manuelle Kontrolle bestünde somit die Gefahr, dass dieser Wert als H_2 gezählt würde und somit falsche Differenzen berechnet werden. Obwohl die Berechnung akustischer Korrelate von Phonationsart und Intensität also technisch machbar wäre, birgt sie bei der Verarbeitung spontaner gesprochener Sprache doch noch einige Schwierigkeiten, deren ausstehende Lösung hiermit als Forschungsdesiderat identifiziert wurde. Bis dato müsste zur reliablen Auswertung von Phonations- und Intensitätsmaßen in dieser Arbeit jede einzelne Instanz visuell kontrolliert werden, da der Algorithmus mit einer bestimmten Voreinstellung nicht über alle anwendbar ist.¹⁰⁷

¹⁰⁶<https://github.com/stas-g/findPeaks>, besucht am 28.08.2019.

¹⁰⁷Nach einer solchen manuellen Korrektur müsste im Anschluss zudem der Effekt der Formanten auf die Harmonischen korrigiert werden ($H_1^*-H_2^*$ und $H_1^*-F_3^*$, vgl. bspw. Iseli et al. 2007).

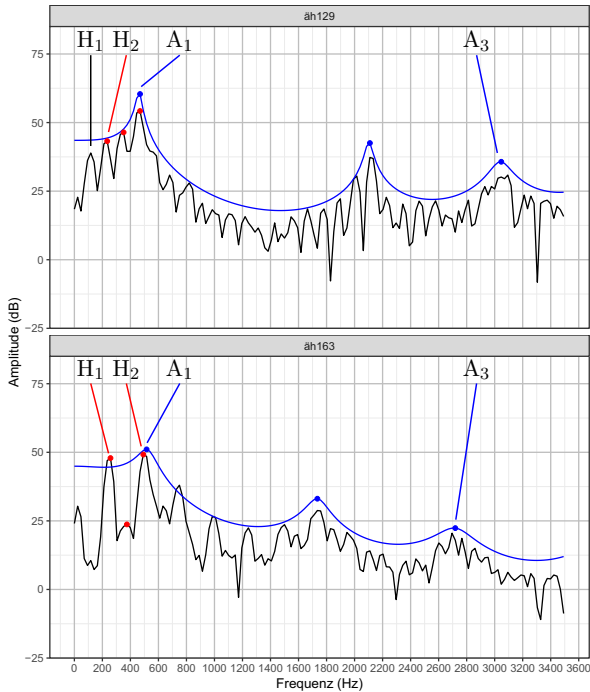


Abbildung 4.11: DFT-Spektra (schwarz) und spektrale Hüllkurve (blau) zweier Füllpartikeln in GECO-FP, gemessen in der Vokalmitte. Rote und blaue Punkte stellen die ersten drei von *findPeaks* berechneten Maxima dar (untere Schwelle 130 Hz, $m = 2, 7$). Im oberen Bild wird aufgrund der Schwelle bei 130 Hz H_1 nicht erkannt; im unteren Bild wird fälschlicherweise ein Maximum bei ca. 375 Hz erkannt.

4.5 Statistik

Die Visualisierung der Daten erfolgt mithilfe des R-Pakets *ggplot2* (Wickham 2009), der Export von Tabellen nach L^AT_EX mithilfe von *xtable* (D. B. Dahl 2016). Die L^AT_EX-Dateien werden dynamisch mit dem R-Paket *knitr* (Xie 2019) erzeugt.

Kategoriale Daten werden mit χ^2 -Tests getestet. Zur *Post-hoc*-Analyse der χ^2 -Tests wird Fishers exakte 2x2-Annäherung nach Shan & Gerstenberger (2017) mit folgendem R-Code implementiert, wobei *obs* für die Tabelle mit den beobachteten Werten steht.¹⁰⁸

```

1 fish<-matrix(nrow(obs)*ncol(obs), ncol=ncol(obs),nrow = nrow(obs))
2 for (i in 1:nrow(obs)) {
3   for (j in 1:ncol(obs)) {
4     small <- matrix(c(obs[i,j],
5                       sum(obs[-i,j]),
6                       sum(obs[i,-j]),
7                       sum(obs)-
8                       sum(obs[-i,j])-
9                       sum(obs[i,-j])-
10                      obs[i,j]), ncol=2)
11     sm<-fisher.test(small)
12     fish[i,j]<-sm$p.value
13   }
14 }

```

Zur Evaluierung der Überlappung zweier Vokalverteilungen (zur Normalisierung vgl. Abschnitt 4.4.3) verwende ich den Pillai-Wert (vgl. Hay et al. 2006). Der Pillai-Wert ist eine Test-Statistik aus einer multivariaten Varianzanalyse (MANOVA), rangiert zwischen 0 und 1 und bildet die gemeinsame Variation von F_1 und F_2 ab (Nycz & Hall-Lew 2013: 5). Je höher der Pillai-Wert, desto größer ist die Distanz von F_1 und F_2 zwischen zwei Vokalen (beziehungsweise zwischen zwei Verteilungen). Umgekehrt gilt für niedrige Pillai-Werte, dass sich zwei Vokale überlappen (Hay et al. 2006: 467). Auch der dritte Formant F_3 kann in die MANOVA einbezogen werden.

Lineare gemischte Modelle (*linear mixed models* oder LMM) werden mithilfe der R-Pakete *lme4* (Bates et al. 2015, 2014) und *lmerTest* (Kuznetsova et al. 2017)

¹⁰⁸Zeile 9 korrigiert das positive Vorzeichen der letzten Variable von Spalte 3, Zeile 3 in Tabelle 2 in Shan & Gerstenberger (2017) ($N - m_i - n_j + x_{ij}$) zu einem negativen ($N - m_i - n_j - x_{ij}$), vgl. <https://doi.org/10.1371/journal.pone.0188709.t002> (besucht am 04.10.2019).

gerechnet. Beginnend mit einem Nullmodell mit den Versuchspersonen als *random intercept* werden sukzessive unabhängige Faktoren hinzugefügt, bis ein maximales Modell erreicht ist. Hierbei wird darauf geachtet, nur solche Prädiktoren (*fixed effects*) zu inkludieren, die konzeptuell zur Forschungsfrage passen (vgl. Harrison et al. 2018). Ein Modell wird mit zusätzlichen Prädiktoren verbessert, wenn das Modell mit dem neuen Faktor ein niedrigeres Informationskriterium nach Akaike (AIC) zeigt (Akaike 1974) und der Likelihood-Quotienten-Test mittels ANOVA-Vergleich beider Modelle mindestens ein α -Niveau von 0,05 erreicht. Nach der vollständigen Identifizierung der Prädiktoren in dem Modell mit *random intercepts* teste ich ebenfalls mit Likelihood-Quotienten-Tests (diesmal jedoch ohne *reduced maximum likelihood*), ob *random slopes* für die Versuchspersonen das Modell abermals verbessern. Schließlich wird die Normalverteilung der Residuen als Voraussetzung für die Anwendung gemischter Modelle getestet, indem eine visuelle Überprüfung mithilfe von Quantil-Quantil-Diagrammen mithilfe des R-Paket *car* (J. Fox & Weisberg 2011) stattfindet. Die Residuen werden ferner visuell auf Varianzhomogenität (Homoskedastizität) geprüft. Für jedes Modell wird der marginale (R_m^2) und konditionale (R_c^2) Determinationskoeffizient als Gütekriterium der Regression angegeben (R-Paket *MuMIn*, Barton 2018). Dies sind Maße für die Varianz, die allein von den unabhängigen Faktoren (R_m^2) erklärt wird sowie für die Varianz, die von den unabhängigen und zufälligen Faktoren gemeinsam (R_c^2) erklärt wird (Nakagawa & Schielzeth 2013). *Post-hoc*-Vergleiche verschiedener Faktorstufen werden mit dem R-Paket *emmeans* durchgeführt (Lenth 2019). Der Export der LMM-Ergebnisse nach L^AT_EX erfolgt mithilfe von *texreg* (Leifeld 2013).

Nicht-lineare (,kurvige‘) Effekte können von linearen gemischten Modellen nicht erfasst werden, finden sich aber beispielsweise in Grundfrequenz- oder Formantentrajektorien. Diese nichtlinearen Kurven werden mithilfe von generalisierten additiven gemischten Modellen (*generalized additive mixed models* oder GAMs) mit dem R-Paket *mgcv* (Wood 2011, 2017, 2019) geschätzt und mit dem R-Paket *itsadug* (van Rij et al. 2017) visualisiert. Zur Ermittlung, ob eine unabhängige Variable das Modell verbessert, werden die AIC-Werte (Akaike 1974) der Modelle verglichen (Funktion *compareML*). Zuvor wird ihre *maximum likelihood* neu geschätzt (vgl. Wieling 2018). Da die Residuen aufeinanderfolgender Punkte in Zeitreihen korreliert sind, werden die Modelle für diese sogenannte Autokorrelation mithilfe des Parameters

Tabelle 4.5: Statistische Tests je abhängiger Variable.

Variable	Test	Post-hoc
Anzahl	χ^2 , Exakter Fisher-Test Binomialtest	Fisher nach Shan & Gerstenberger (2017)
Dauer	t-Test LMM	Bonferroni-korrigierte paarweise Vergleiche Multiple Vergleiche mit Tukey-Korrektur
Grundfrequenz	t-Test LME	Multiple Vergleiche mit Tukey-Korrektur
Formanten	MANOVA LMM GAM	Multiple Vergleiche mit Tukey-Korrektur Multiple Vergleiche mit Tukey-Korrektur

rho korrigiert (vgl. Wieling 2018). Die Residuen werden auf Normalverteilung und Homoskedastizität visuell überprüft. Formanten werden vor der Modellierung normalisiert. Tabelle 4.5 gibt einen Überblick über die verwendeten Tests je abhängiger Variable.

4.6 Zugang und Wiederverwendung

musst dir halt einfach vorstellen dass jetzt (Pause, 1,3s) irgendwann in zwei Jahren
oder so irgendwann jemand sitzt und sich unser Gebrabbel hier reinzieht und dann
(Einatmung, 560 ms) was auch immer damit macht so
(BeDiaCo:frei_m4m5_ch2, 571–582 s)

Um obigen Ausschnitt linguistisch analysieren zu können, muss das Korpus, in welchem er vorkommt, zugänglich sein: „Das Teilen von Korpora ist im gleichem Maße Voraussetzung für deren Wiederverwendung wie die Korpusdokumentation selbst.“ (Odebrecht 2018: 183). Die im Rahmen meiner Dissertation aufgenommenen anonymisierten Audiodaten und die erstellten Transkriptionen und Annotationen werden daher von mir in verschiedener Weise dokumentiert und unter Berücksichtigung geltender Datenschutzbestimmungen veröffentlicht.

4.6.1 GECO v.1.1

Das GECO-Korpus ist über die Universität Stuttgart verfügbar und für akademische Nutzung lizenziert.¹⁰⁹ Daher können von mir keine Audiodaten bereitgestellt werden. Meine Annotationsebenen und korrigierten Grundfrequenz- und Formanttrajektorien sind als TextGrid-Dateien und als EMU-Datenbank unter dem Namen GECO-FP veröffentlicht: <https://doi.org/10.18452/20794> (Belz 2019b).

4.6.2 BeDiaCo v.1

Von den in BeDiaCo_c analysierten 12 Dialogen und 24 Versuchspersonen haben insgesamt 16 Versuchspersonen (10 männliche, 6 weibliche) in 8 Dialogen (3m/m-, 1 f/f-, 4 f/m-Dialoge) zugestimmt, dass die von ihnen erhobenen und anonymisierten/pseudonymisierten Daten an ein Datenzentrum zur Archivierung und weiteren wissenschaftlichen Nutzung für zukünftige themenverwandte Forschungsprojekte übermittelt und Wissenschaftlerinnen und Wissenschaftlern für ausschließlich wissenschaftliche Zwecke im Bereich linguistischer Forschung unter Beachtung der jeweils geltenden Datenschutzbestimmungen zur Verfügung gestellt werden dürfen. Die Daten dieser Versuchspersonen sind als BeDiaCo v.1 auf dem Medienrepositorium der Humboldt-Universität zu Berlin veröffentlicht (Belz & Mooshammer 2020) und enthalten Audio- und Annotationsdaten der Dialoge und Wortlisten. Die Dokumentation des Korpus ist auf dem edoc-Server der Humboldt-Universität zu Berlin veröffentlicht (Belz et al. 2020).

¹⁰⁹<https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/IMS-GECO.html>, besucht am 02.05.2019. Informationen zur Zugänglichkeit unter <http://hdl.handle.net/11022/1007-0000-0000-8E5C-A>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





5 Ergebnisse der explorativen Studie

Zwei zentrale Hypothesen werden im Literaturteil dieser Arbeit aufgestellt, nämlich die Hypothese der Formvariabilität (Kapitel 2: Welche phonetische Varianten treten auf?) und die Form-Kontext-Hypothese (Kapitel 3: Kann ein Teil der vorkommenden Formvariabilität durch den jeweiligen Kontext erklärt werden?). Im Folgenden werde ich mich diesen Fragen explorativ annähern.

5.1 Füllpartikeln

5.1.1 Allgemeine und sprecherspezifische Häufigkeit

Tabelle 5.1 gibt einen Überblick über die Anzahl der Füllpartikeln und ihre sprecher-spezifische Auftretenshäufigkeit in den Dialogen aus GERman COversations-Füllpartikeln v. 1 (GECO-FP).¹¹⁰ Die Sprechgeschwindigkeiten der Versuchspersonen rangieren zwischen 4,9 und 6,6 σ/s , die Anzahl der Füllpartikeln je Minute zwischen 1,4 und 4 FP/min, mit einem Mittelwert von 2,9 FP/min. Es besteht weder eine signifikante Korrelation (nach Pearson) zwischen Sprechgeschwindigkeit und Füllpartikeln je Minute ($r = 0,21, t = 0,52, df = 6, p = 0,6, KI_{95\%} = [-0,58; 0,8]$) noch zwischen Tokenanzahl und Füllpartikelanzahl ($r = 0,68, t = 2,28, df = 6, p = 0,1, KI_{95\%} = [-0,04; 0,94]$) und auch nicht zwischen Silbenanzahl und Füllpartikelanzahl ($r = 0,67, t = 2,22, df = 6, p = 0,1, KI_{95\%} = [-0,06; 0,93]$). Bei der geringen Sprecherinnenanzahl kann nicht mit Sicherheit angegeben werden, welcher Verteilung

¹¹⁰Die Tokenanzahl wurde zur Berechnung der Sprechgeschwindigkeit um 29 leere Token sowie um 1 463 extralinguistische Phänomene wie Husten, Lachen, etc., um 12 480 stille Pausen sowie um 111 Clicks reduziert. Enthalten sind jedoch Füllpartikeln, insbesondere auch glottale Füllpartikeln (annotiert mit *fg*), die orthographisch nicht wiederzugeben waren und auf der Ebene *wordscor* nur impressionistisch und *ad hoc* beschrieben werden. Ihre Unterscheidung in dieser Ebene ist für diese Studie nicht relevant. Dies sind die 101 Werte ⟨brummen⟩ (2), ⟨G⟩ (15), ⟨knack⟩ (11), ⟨knacken⟩ (33), ⟨knarr⟩ (1), ⟨knarren⟩ (20), ⟨knarrquietsch⟩ (1), ⟨knarzen⟩ (1), ⟨knattern⟩ (1), ⟨schnurren⟩ (8), ⟨stimmlippen⟩ (1), ⟨stimmtön⟩ (6) und ⟨Stimmtön⟩ (1).

Zusatzmaterial online

Zusätzliche Informationen sind in der Online-Version dieses Kapitel (https://doi.org/10.1007/978-3-662-62812-6_5) enthalten.

© Der/die Autor(en) 2021

M. Belz, *Die Phonetik von äh und ähm*,

https://doi.org/10.1007/978-3-662-62812-6_5

Tabelle 5.1: Artikulationsdauer je Versuchsperson (VP), Anzahl der Token, Silben (σ) und Füllpartikeln (FP) sowie Sprechgeschwindigkeit (σ/s), Prozent FP je Token und Silbe sowie FP je Minute für alle Versuchspersonen (VP) in GECO-FP. Zusätzlich sind die addierten Gesamtsummen (Σ) sowie die Mittelwerte (\bar{x}) und Standardabweichung (s) über alle VP angegeben.

VP	Σ min	Token ^a	Silben	σ/s	FP ^b	FP/Token (%)	FP/ σ (%)	FP/min
A	19,03	4 204	5 821	5,10	30	0,71	0,52	1,58
C	23,69	4 731	6 904	4,86	80	1,69	1,16	3,38
D	21,95	4 974	6 796	5,16	86	1,73	1,27	3,92
F	18,69	4 058	5 678	5,06	27	0,67	0,48	1,44
H	19,63	4 956	7 104	6,03	79	1,59	1,11	4,02
J	26,45	7 197	10 427	6,57	61	0,85	0,59	2,31
K	32,10	8 189	11 600	6,02	120	1,47	1,03	3,74
M	25,69	5 900	8 834	5,73	75	1,27	0,85	2,92
Σ	187,2	44 209	63 164	–	558	–	–	–
\bar{x}	23,4	5 526,1	7 895,5	5,6	69,8	1,2	0,9	2,9
s	4,6	1 473	2 172,9	0,6	30,5	0,4	0,3	1

^a Ohne leere Token, extralinguistische Phänomene, stille Pausen und Clicks.

^b Alle mit *fv* und *fg* annotierte Füllpartikeln.

die Daten folgen. Tests auf Normalverteilung werden erst ab ca. 50 Datenpunkten robust. Somit kann hier nicht ausgeschlossen werden, dass die zugrundeliegende Verteilung der Datenpunkte von der Normalverteilung abweicht.¹¹¹ Tabelle 5.2 enthält die glottalen und nicht-glottalen Füllpartikelformen je Versuchsperson.

Tabelle 5.2: Anzahl der glottalen (*fg*) und nicht-glottalen (*fv*) Füllpartikeln je Versuchsperson in GECO-FP.

	A	C	D	F	H	J	K	M	Σ	%
fg	13	24	17	3	6	1	38	9	111	19,89
fv	17	56	69	24	73	60	82	66	447	80,11
Σ	30	80	86	27	79	61	120	75	558	

Auch hier kann aufgrund der Stichprobengröße nicht mit Sicherheit angegeben werden, welcher Verteilung die Daten folgen.¹¹² Fast 20 % aller Füllpartikelformen

¹¹¹Ein Shapiro-Wilk-Test auf Normalverteilung ergibt keine Anzeichen, dass die Anzahl der Füllpartikeln über alle Versuchspersonen nicht normalverteilt ist ($W = 0,93, p = 0,5$). Selbiges gilt für die Anzahl an Füllpartikeln je Minute ($W = 0,89, p = 0,2$) und den prozentualen Anteil in Füllpartikeln ($W = 0,87, p = 0,1$).

¹¹²Über alle Versuchspersonen zeigen auch diese Verteilungen weder für die glottalen Füllpartikeln ($W = 0,91, p = 0,4$) noch für die nicht-glottalen Füllpartikeln ($W = 0,87, p = 0,1$) Abweichungen

in GECO-FP sind glottale Füllpartikeln.

5.1.2 Segmentale Formen

5.1.2.1 Nicht-glottale Formen

Auf der Ebene *segm* sind 67 verschiedene Formen von Segmentabfolgen vorhanden. Durch Zusammenfassung konzeptuell ähnlicher Segmentabfolgen kann die Anzahl der Formen auf 55 segmental-abstrahierte Formen reduziert werden. Tabelle 5.3 zeigt die zehn häufigsten segmental-abstrahierten Formen (für alle segmental-abstrahierten Formen s. Tabelle D.2 im Anhang). Am häufigsten kommen ?VN, ?V, VN und V vor, also vokalische und vokalisch-nasale Formen mit und ohne vorangehendem glottalen Plosiv. Weniger häufig sind Formen, in denen dem Vokal eine Sequenz glottaler Plosive vorausgeht (GVN, GV).

Tabelle 5.3: Anzahl der zehn häufigsten segmental-abstrahierten Formen aller als *fv* annotierten Füllpartikeln in GECO-FP (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal).

Segmente	Anzahl
?VN	105
?V	76
VN	51
V	50
GVN	33
GV	23
N	15
?VNG	7
?VC	5
?VN?	5

Abbildung 5.1 zeigt die Verteilung der abstrahierten Segmentformen und ihre Darstellung auf einer Log-log-Skala. Die logarithmisierten Häufigkeitswerte und ihre Ränge nähern sich einer Geraden mit negativer Steigung ($y = -x$) an, was auf eine Zipfverteilung der Daten deutet (Zipf 1949). Segmentale Füllpartikelformen verhalten sich somit ähnlich wie viele andere sprachliche Phänomene.

Die orthographische Repräsentation ist eine weitere Abstraktion der phonetisch vorhandenen Segmente. Abbildung 5.2 zeigt die Relation der segmental-abstrahierten von der Normalverteilung (Shapiro-Wilk-Test auf Normalverteilung).

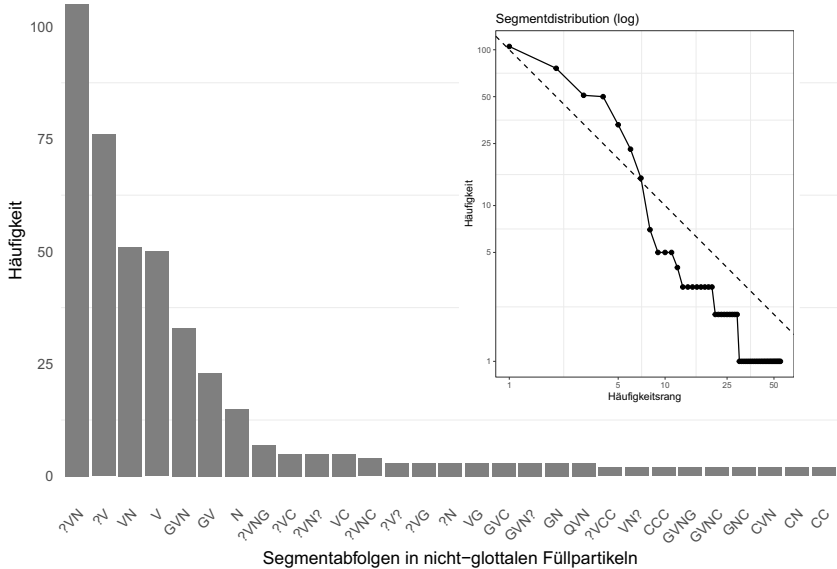


Abbildung 5.1: Anzahl ($n > 1$) der abstrahierten Segmentabfolgen in Füllpartikeln (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal, C = weitere Konsonanten und Clicks, Q = hohe gepresste Stimme). Der innere Graph enthält alle Rangfrequenzen der Füllpartikelformen (logarithmisch) und eine ideale Zipf-Verteilung (gestrichelt).

Formen zu den orthographischen Repräsentationen.¹¹³ In Abbildung 5.2 lässt sich gut erkennen, dass die als *äh*, *ähm* und *hm* wiedergegebenen orthographischen Formen phonetisch divers sind. Besonders zu Anfang und Ende einer Füllpartikel treten ein oder mehrere glottale Plosive auf. Die häufigsten *hm*-Formen in dieser Datenstichprobe enthalten keinen glottalen Frikativ [h]. Die drei häufigsten orthographischen Vorkommen sind *ähm* (224 mal), *äh* (179 mal) und *hm* (20 mal). *ähm* kommt somit 1,3 mal häufiger vor als *äh*, was signifikant häufiger ist (zweiseitiger exakter Binomialtest, $p < 0,05$). Dies deckt sich mit früheren Studien, in denen Frauen

¹¹³Für alle Formen, die mehr als fünf Vorkommen haben (was 98 abstrahierte Formen beziehungsweise 21,9% der 447 nicht-glottalen Füllpartikeln ausschließt). Orthographisch ist *m* und *mh* zur ‚gängigen‘ Form *hm* zusammengefasst worden.

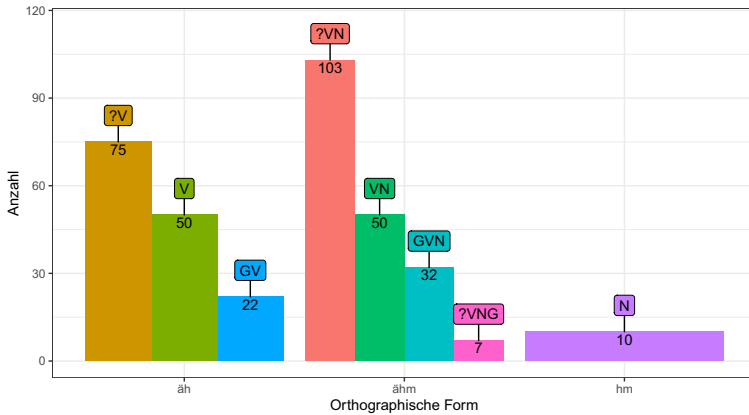


Abbildung 5.2: Zusammenfassung verschiedener segmentaler Realisierungen ($n > 5$) unter zwei orthographischen Formen (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal, C = Konsonant).

mehr VN-Formen als V-Formen produzieren (vgl. Wieling et al. 2016). Da alle Versuchspersonen in GECO-FP weiblich sind, kann dieses Verhältnis allerdings erst in der nächsten Studie mit Männern verglichen werden. Offen bleibt die Frage, wieso Frauen häufiger VN-Formen als V-Formen verwenden.

5.1.2.2 Glottale Formen

Für die glottalen Formen abstrahiere ich Segmentabfolgen mit mehr als einem glottalen Plosiv und perceptiv deutlich abgesetzten (im akustischen Signal mit Stille > 50 ms markierten) Sequenzen zu ?+ (beispielsweise[???]). Laut Tabelle D.3 im Anhang bestehen die häufigsten glottalen Füllpartikelformen aus einer engen Abfolge glottaler Plosive (G), aus singulären glottalen Plosive (?) und aus einer losen, perceptiv vereinzelt wahrnehmbaren Abfolge glottaler Plosive (?+). Wie schon bei den nicht-glottalen Formen zeigt auch die visuelle Inspektion der segmental-abstrahierten glottalen Füllpartikelformen eine Distribution, die eine Zipf-Verteilung vermuten lässt (vgl. Abbildung 5.3).

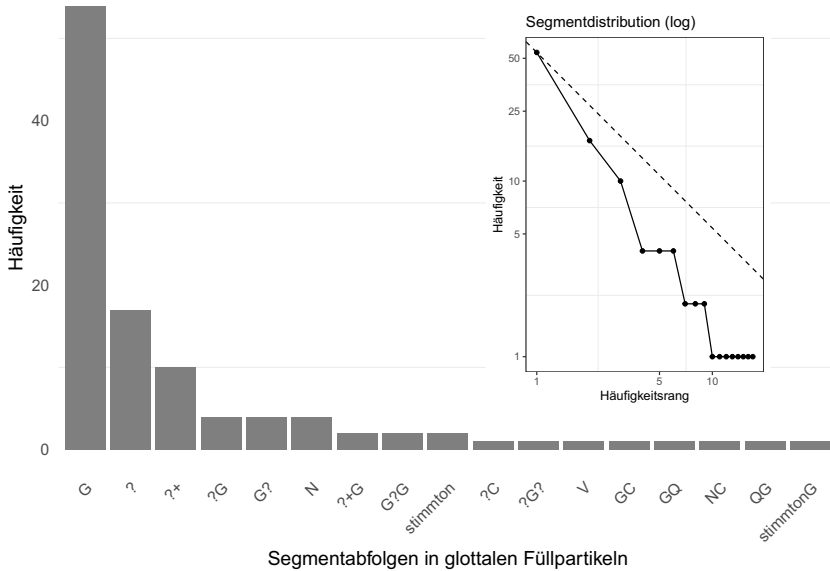


Abbildung 5.3: Anzahl der abstrahierten glottalen Segmentabfolgen in Füllpartikeln (? = ein bis drei glottale Plosive, ?+ = perzeptiv vereinzelt wahrnehmbare Abfolge glottaler Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal, C = Konsonant, Q = hohe gepresste Stimme, stimmton = Stimmton). Der innere Graph enthält alle Rangfrequenzen der Füllpartikelformen (logarithmisch) und eine ideale Zipf-Verteilung (gestrichelt).

5.1.3 Dauer

5.1.3.1 Nicht-glottale Formen

Die geringste Dauer einer nicht-glottalen Füllpartikel in GECO-FP beträgt 34,6 ms (?→E beziehungsweise abstrahiert ?V), die längste 1 239,6 ms (?→E→p→f beziehungsweise abstrahiert ?VCC). Abbildung 5.4 zeigt die Distribution der auf Silben normalisierten nicht-glottalen Füllpartikeln für die häufigsten Formen ($n > 5$). Zur Berechnung der normalisierten Dauer vgl. Abschnitt 4.4.1.

Aus Abbildung 5.4 lässt sich eine positive Korrelation (nach Pearson) zwischen Segmentanteilen in der phonetischen Realisierung und der Länge von Füllpartikeln

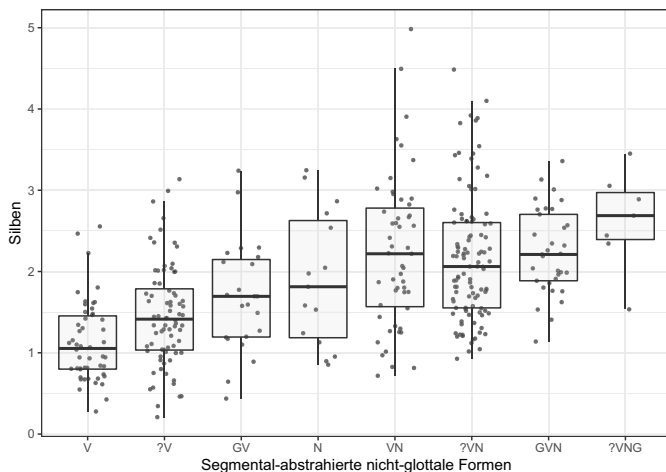


Abbildung 5.4: Dauer (in Silben) der häufigsten ($n > 5$) segmental-abstrahierten nicht-glottalen Füllpartikelformen in GECO-FP (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal).

ableiten ($r = 0,42, t = 9,83, df = 445, p < 0,001, KI_{95\%} = [0,34; 0,5]$). Je mehr Segmente phonetisch realisiert werden, desto länger ist ihre Dauer. Mithilfe von bonferronikorregierten *Post-hoc*-Tests werden die Formen ermittelt, die sich signifikant voneinander unterscheiden (vgl. Tabelle 5.4).

Tabelle 5.4: Bonferronikorregierte p-Werte für paarweise t-Tests der Mittelwerte der acht segmental-abstrahierten nicht-glottalen Füllpartikelformen in Abbildung 5.4.

	V	?V	GV	N	VN	?VN	GVN
?V	0,353	–	–	–	–	–	–
GV	0,057	1	–	–	–	–	–
N	<0,01	0,757	1	–	–	–	–
VN	<0,001	<0,001	0,063	1	–	–	–
?VN	<0,001	<0,001	0,129	1	1	–	–
GVN	<0,001	<0,001	0,12	1	1	1	–
?VNG	<0,001	<0,01	0,066	0,735	1	1	1

Die V-Form ist signifikant kürzer als die vokalischnasalen Formen VN, ?VN, GVN, ?VNG und die nasale Form N. Die Form mit vorangehendem glottalen Plosiv

?V ist immer noch signifikant kürzer als alle vokalisch-nasalen Formen. Die GV-Form (mit vorangehenden Sequenzen glottaler Plosive) ist in ihrer Dauer nicht von den anderen Formen zu unterscheiden. Während also prinzipiell denkbar wäre, dass eine V-Form mit gleicher Dauer wie eine VN-Form geäußert wird, zeigen die Daten doch eine systematische Unterscheidung beider Formen bezüglich ihrer Dauer. Der Vollständigkeit halber und zum Vergleich mit anderen Studien gibt Tabelle 5.5 die durchschnittliche Dauer der acht häufigsten Formen in der auf Sprechgeschwindigkeit normalisierten Dauer in Silben und der nicht-normalisierten Dauer in Millisekunden an. Insgesamt haben V-Formen eine durchschnittliche Dauer von $\bar{x} = 1,5 \sigma$ ($s = 0,7 \sigma$), VN-Formen von $\bar{x} = 2,3 \sigma$ ($s = 0,8 \sigma$) und N-Formen von $\bar{x} = 2,5 \sigma$ ($s = 1,1 \sigma$).

Tabelle 5.5: Mittelwerte und Standardabweichungen der normalisierten Dauer in Silben (σ) pro Sekunde und der nicht-normalisierten Dauer (in Millisekunden) der acht häufigsten segmental-abstrahierten nicht-glottalen Füllpartikelformen in Abbildung 5.4.

Form	\bar{x} (in σ)	s (in σ)	\bar{x} (in ms)	s (in ms)
V	1,1	0,5	205,6	93,0
?V	1,5	0,6	256,7	110,9
GV	1,7	0,7	291,4	108,1
N	1,9	0,8	334,4	150,4
?VN	2,2	0,8	371,5	149,1
VN	2,2	0,9	381,9	127,2
GVN	2,2	0,5	393,4	94,1
?VNG	2,6	0,6	473,5	124,0

Die nicht-normalisierten Mittelwerte und Standardabweichungen für die orthographischen Formen (*äh*: $\bar{x} = 262,2$, $s = 121$; *ähm*: $\bar{x} = 395,8$, $s = 139,5$; *hm*: $\bar{x} = 450,4$, $s = 182,6$; in Millisekunden) in GECO-FP können dank Angabe von Stichprobengröße, Mittelwerten und Standardabweichungen mit denen von Leeuw (2007) verglichen werden (vgl. Tabelle 2.5). Keine der drei Formen weicht signifikant voneinander ab (t-Test mit ungleichen Stichprobengrößen und ungleichen Varianzen nach Welch-Satterthwaite), womit die durchschnittliche unnormalisierte Dauer von Füllpartikeln in den Interviews von Leeuw repliziert wird.

In der Literatur wird ein möglicher Effekt der Vokaldauer in Abhängigkeit von der Füllpartikelform angesprochen (Jessen 2012). Zur Überprüfung rechne ich ein lineares gemischtes Modell mit Dauer als abhängiger Variable und Form als unabhängiger Variable. Versuchspersonen gehen als Random Intercepts ein, mit Random

Slopes für die Form je Versuchsperson. Abbildung 5.5 visualisiert die Vorhersagen des Modells, Tabelle 5.6 enthält die Werte. Das Modell wird folgendermaßen in R dargestellt.

```
1 | lmer(Vokaldauer ~ Form + (1+Form|VP), data ...
```

Die Vokaldauer in vokalisch-nasalen Füllpartikeln ist mit -68,6 ms tatsächlich signifikant kürzer als in vokalischen Füllpartikeln.

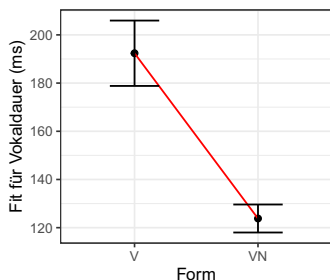


Abbildung 5.5: Effekt für die Vokaldauer in Abhängigkeit von der Füllpartikelform in GECO-FP.

Tabelle 5.6: Linear gemischtes Modell für die Vokaldauer in Abhängigkeit von *Form* + $(1+Form/VP)$, mit erklärter Varianz R^2 .

	Vokaldauer
(Intercept)	192,4 (13,6)***
FormVN	-68,6 (12,8)***
AIC	4381,0
Num. obs.	391
Num. groups: vp	8
R_m^2/R_c^2	0,2/0,29

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

5.1.3.2 Glottale Formen

Die häufigsten glottalen Formen mit mehr als fünf Vorkommen sind ?, G und ?+. Die Form ? ist signifikant kürzer als die Form G und ?+ (paarweiser bonferronikorrektierter t-Test, jeweils $p < 0,001$). Abbildung 5.6 zeigt ihre Distribution, während Ta-

belle 5.7 die auf Silben normalisierten sowie die nicht-normalisierten Werte auflistet. Der geringste Wert einer glottalen Füllpartikel liegt bei 27 ms (?→G beziehungsweise abstrahiert ?G), der größte Wert bei 1 068,4 ms (G).

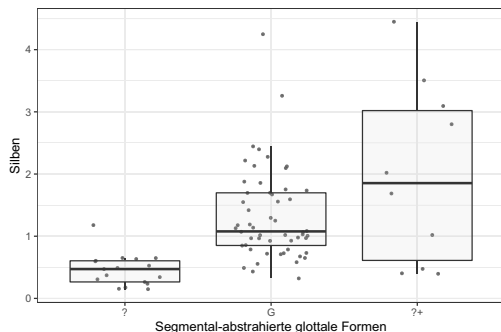


Abbildung 5.6: Dauer (in Silben) der häufigsten ($n > 5$) segmental-abstrahierten glottalen Füllpartikelformen in GECCO-FP (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, ?+ = Sequenz vereinzelt perzipierbarer glottaler Plosive mit intersegmentalen Pausen > 50 ms).

Tabelle 5.7: Mittelwerte und Standardabweichungen der normalisierten Dauer (in Silben σ) und der nicht-normalisierten Dauer (in Millisekunden) der drei häufigsten segmental-abstrahierten glottalen Füllpartikelformen in Abbildung 5.6.

Form	\bar{x} (in σ)	s (in σ)	\bar{x} (in ms)	s (in ms)
?	0,5	0,3	80,7	50,2
G	1,3	0,7	239,6	133,8
?+	2,0	1,4	347,2	236,5

5.1.4 Grundfrequenz

Für die Analyse der Grundfrequenz (f_0) werden nur die nicht-glottalen Formen betrachtet, da in den glottalen Formen die Stimmlippenschwingung zu unregelmäßig für die reliable Ermittlung der Grundfrequenz ist. Die Grundfrequenz messe ich in der Mitte der Füllpartikel, wobei vorangehende und nachfolgende mögliche andere Lautsegmente nicht berücksichtigt werden. Dennoch kann es zu Fällen kommen, in denen die Grundfrequenz in Teilen nicht ermittelt werden kann. Zur ersten Übersicht

zeigt das Histogramm in Abbildung 5.7 die Häufigkeit der Grundfrequenz über die Anzahl der vokalischen und vokalisch-nasalen Füllpartikeln. In 28 Fällen misst der Algorithmus 0 Hz. Die restlichen Daten zerfallen in zwei Gruppen; eine erste zwischen 0 und 150 Hz (der Mittelwert dieser Häufung beträgt 95,2) und eine zweite ab 150 Hz (der Mittelwert beträgt 202,9). Möglicherweise sind die Daten in der ersten Gruppe auf glottalisierte Sequenzen zurückzuführen.

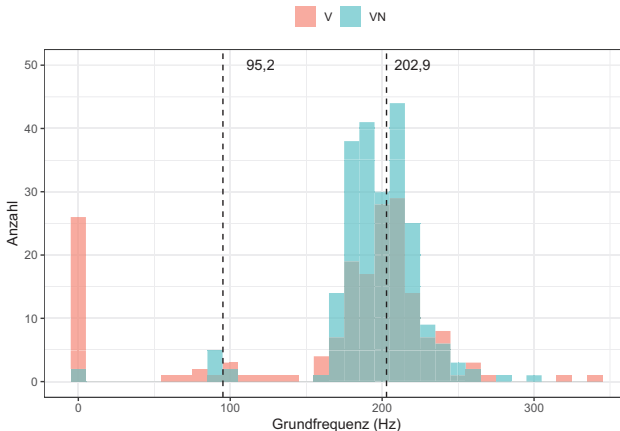


Abbildung 5.7: Histogramm der im Mittelpunkt aller V- (rot) und VN-Formen (blau) gemessenen Grundfrequenz mit einer Klassenbreite von 10 Hz und den Mittelwerten der beiden Häufungen (gestrichelte Linien) in GECO-FP.

Die Heuristik, in der Mitte eines Wortes beziehungsweise einer Füllpartikel zu messen, kann bei solchen glottalisierten oder laryngalisierten Sequenzen dazu führen, dass zufällig ein solcher Abschnitt in den Messpunkt fällt, der dann als Grundfrequenz der gesamten Füllpartikel interpretiert wird. Abbildung 5.8 veranschaulicht dies exemplarisch an allen Füllpartikelgrundfrequenzverläufen von Sprecherin C. Somit kann die aus dem Histogramm in Abbildung 5.7 inferierte Dichotomie zweier Häufungen nicht aufrechterhalten werden, da es sich bei den Fällen, welche bei Sprecherin C die erste Gruppe zwischen 0 und 150 Hz konstituieren, fast ausschließlich um eine zufälligerweise am Punkt der Messung vorhandene tiefe Grundfrequenz (Abbildung 5.7 *id=20*) oder um Interpolierungen (Abbildung 5.7 *id=51, 58*) handelt und nicht um komplette Verläufe. Zwar könnte für jede Versuchsperson ein idealisierter

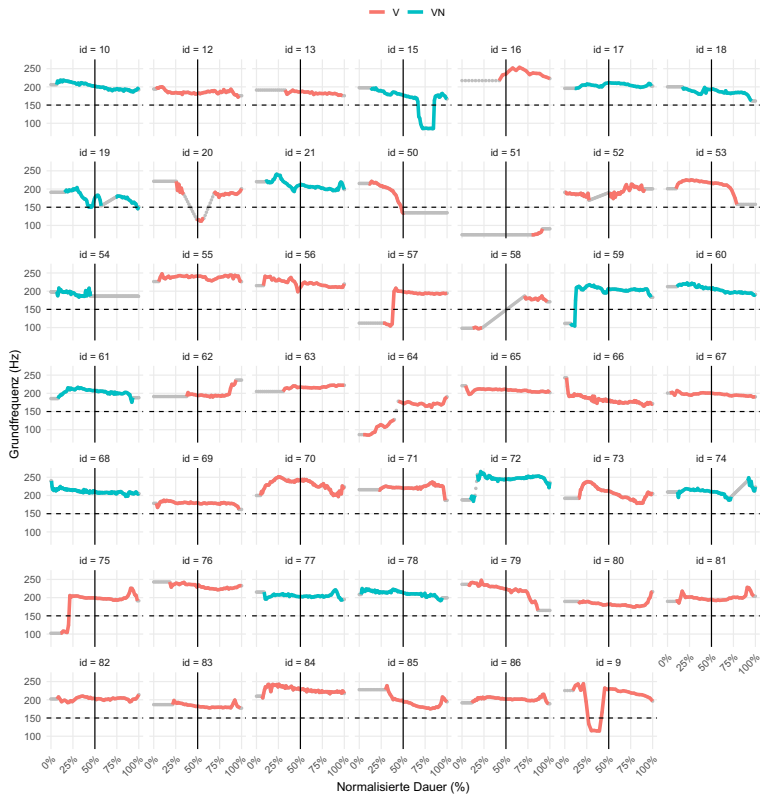


Abbildung 5.8: Grundfrequenzverläufe aller Segmente in V- und VN-Formen von Sprecherin C in GECO-FP mit normalisierter Dauer. Graue Punkte sind interpolierte Daten, die ursprünglich fehlen. Die vertikale Linie kennzeichnet den Messpunkt für die Grundfrequenz, die horizontale gestrichelte Linie markiert 150 Hz.

Grundfrequenzverlauf mithilfe der Interpolierung fehlender Daten erzeugt werden, die Aussagekraft dieser Verläufe wäre jedoch nur schwer interpretierbar, da auch fallende, progrediente, steigende oder polynom verlaufende Konturen in eine Kurve gefittet würden. Daher nehme ich im Folgenden auch Abstand von der Analyse der Grundfrequenztrajektorien und beschränke mich auf im Mittelpunkt gemessenen Werte, jedoch ohne Interpolierung.

Zur Einordnung der Grundfrequenzdaten wird die Grundfrequenz aller Füllpartikeln je Versuchsperson (\bar{f}_{0vp}^{FP}) mit der mittleren Grundfrequenz je Versuchsperson (\bar{f}_{0vp}) verglichen, i. e. die über alle Äußerungen einer Versuchsperson gemittelte Grundfrequenz (vgl. Braun & Rosin 2015). Anders als in bisherigen Studien ist es nach der Sichtung des Histogramms in Abbildung 5.7 jedoch geboten, nur Grundfrequenzen über 150 Hz zu betrachten, um mögliche Messfehler oder möglicherweise starke Glottalisierungen auszuschließen. Ein genauer Wert zur Trennung zwischen solcher modaler und ‚anderer‘ Phonation ist jedoch immer arbiträr und erfordert die Betrachtung aller mittleren Messpunkte aller Wörter aller Versuchspersonen, wie sie in Abbildung 5.9 dargestellt ist. Nach dieser Visualisierung kann die Schwelle aufgrund der eher niedrigeren Grundfrequenz von Sprecherin K auf 140 Hz abgesenkt werden. Tabelle 5.8 zeigt die Anzahl der Füllpartikeln und Wörter über und unter dieser Schwelle für alle Sprecherinnen. Die Schwelle von 140 Hz schließt 11,7% aller V- und VN-Füllpartikeln aus.

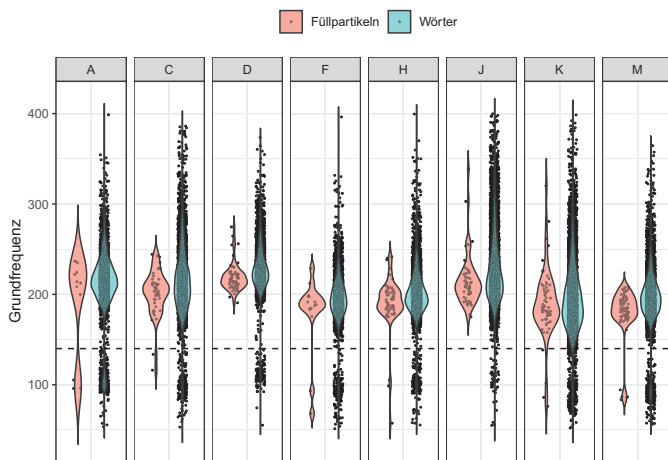


Abbildung 5.9: Mittlere Grundfrequenz für Füllpartikeln und Wörter je Versuchsperson in GECO-FP (im Mittelpunkt gemessen). Die horizontale gestrichelte Linie markiert 140 Hz. Die Violinen zeigen die Häufigkeitsverteilung der Messpunkte (je breiter, desto mehr Messpunkte sind enthalten).

Tabelle 5.9 stellt die Mittelwerte und Standardabweichungen der Einzelmesswer-

Tabelle 5.8: Anzahl der in der Mitte gemessenen V- und VN-Füllpartikeln und Wörter je Grundfrequenzschwelle in GECCO-FP.

Grundfrequenz	Füllpartikeln	Wörter
0 Hz	28	12078
> 0 und < 140 Hz	355	30647
≥ 140 Hz	19	1382
Σ	402	44107

te der beiden Füllpartikelgrundfrequenzgruppen und die über alle Äußerungen einer Versuchsperson gemittelte \bar{f}_{0vp} mit dem gleichen Wertebereich dar. Eine Evaluierung erfolgt mithilfe paarweiser Vergleiche eines gerichteten t-Tests.¹¹⁴ Ein gerichteter Test wird durchgeführt, da die in der Literatur gefundenen Werte bisher häufig eine niedrigere \bar{f}_{0vp}^{FP} ergaben als die mittlere Grundfrequenz einer Versuchsperson (vgl. Braun & Rosin 2015; Klug 2013). Füllpartikeln ($n = 402$) zeigen für den Bereich oberhalb von 140 Hz tatsächlich signifikant niedrigere Grundfrequenzen als die restlichen Äußerungen ($n = 44\,107$) nach Versuchsperson oberhalb von 140 Hz ($t = -4,3$, $p < 0,01$, $KI_{95\%} = [-Inf; -5,32]$), bis auf Sprecherin A.

Tabelle 5.9: Mittelwerte und Standardabweichung für die in der Mitte von Füllpartikeln (\bar{f}_{0vp}^{FP}) und von allen Wörtern (\bar{f}_{0vp}) einer Versuchsperson gemessene Grundfrequenz in Hertz in GECCO-FP.

VP	Füllpartikeln		Wörter	
	\bar{x}	s	\bar{x}	s
A	221,1	13,3	220,1	26,2
C	205,3	18,8	220,2	37,3
D	218,9	15,0	230,7	23,7
F	194,0	13,9	199,1	22,6
H	196,2	15,8	204,3	25,7
J	215,1	28,4	232,0	42,9
K	195,8	30,7	201,0	39,8
M	186,5	11,7	201,2	26,2

¹¹⁴Die Wahl des t-Tests statt des nicht-parametrischen Wilcoxon-Rangsummentest wird durch die Normalverteilung der Differenzen der zu vergleichenden Werte motiviert. Somit kann der mächtigere t-Test verwendet werden.

5.1.5 Formanten

Zur Ermittlung der Formanten und damit der Vokalqualität der Füllpartikeln extrahiere ich aus jeder Füllpartikel einer Versuchsperson F_1 und F_2 in der Mitte des Vokals; ebenso für die Referenzvokale. Alle Formanten wurden zuvor manuell in der EMU-Datenbank korrigiert. Alle Referenzvokale einer Qualität werden je Versuchsperson zu einem einzigen Vertreter dieser Klasse gemittelt. Anschließend wird mit den Werten der Referenzvokale und der Füllpartikelvokale eine formant-intrinsische, sprecher-intrinsische und vokal-extrinsische Lobanovnormalisierung durchgeführt (vgl. Abschnitt 4.4.3). Die so normalisierten Vokalwerte werden für F_1 und F_2 in Abbildung 5.10 in einem Vokalraum dargestellt, der die Streuung der im quasi-stationären Vokalteil (50%) gemessenen Referenz und Füllpartikelvokale als Datenellipsen zeigt.

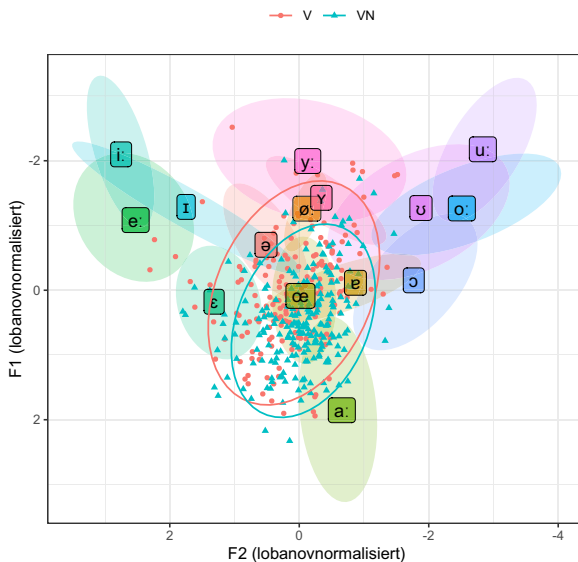


Abbildung 5.10: Mittig gemessene Vokale in V- und VN-Formen (Datenellipsen beschreiben 95% aller Messwerte) im lobanovnormalisierten Vokalraum aller Versuchspersonen in GECO-FP mit schattierten Datenellipsen für die Referenzvokale (beschreiben zur besseren Sichtbarkeit 80% aller Messwerte).

Die Vokalqualität für vokalische Füllpartikeln bewegt sich hauptsächlich im Bereich von [œ ɐ ə] und [ɛ]. Die Vokale vokalisch-nasaler Füllpartikeln überlappen mit den Bereichen von [œ ɐ ə] und [a]. Im untermittelhohen Bereich zwischen [ɛ] und [a] könnte die Vokalqualität des im Deutschen nicht phonologisch vorhandenen [æ] getroffen werden. Als Maß für die Vokalüberlappung berechne ich den Pillai-Wert (vgl. Abschnitt 4.4.3). Je höher der Pillai-Wert, desto größer ist die Distanz von F_1 , F_2 und F_3 zwischen zwei Vokalen. Der in einer MANOVA ermittelte Pillai-Wert mit F_1 , F_2 und F_3 als abhängiger Variable für die Distribution der Vokale in V- und VN-Formen in Abbildung 5.10 beträgt 0,1 (Df = 398, F = 14,23, $p < 0,001$, $\eta^2 = 0,1$). Die Vokale beider Formen überlappen sich also zum größten Teil (F_3 als Maß für die Lippenrundung kann in der zweidimensionalen Graphik nicht dargestellt werden, wird allerdings weiter unten nochmals diskutiert, vgl. Abbildung 5.13). Abbildung 5.11 zeigt, dass bei der Interpretation des errechneten Pillai-Wertes für alle Füllpartikeln in Abbildung 5.10 Zurückhaltung geboten ist, da sich die einzelnen Versuchspersonen sehr unterschiedlich verhalten, was in einer MANOVA nicht modelliert werden kann. Die Effektstärke η^2 liegt jedoch durchaus im Bereich eines mittelgroßen Effektes.¹¹⁵

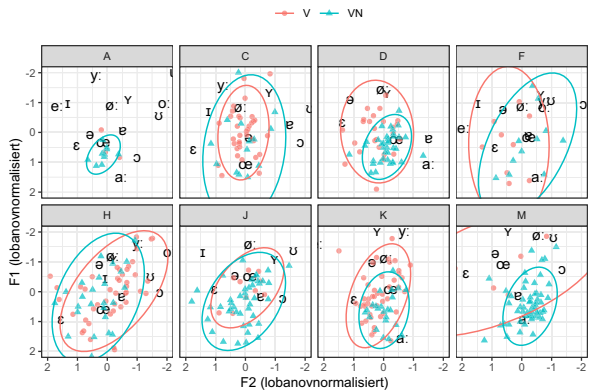


Abbildung 5.11: Vokalische (rot) und vokalisch-nasale (blau) Füllpartikeln in GECO-FP je Sprecherin in lobanovnormalisierten Vokalräumen (mittige Messung im Vokal, gezoomt). Datenellipsen umspannen 95 % der Messpunkte.

¹¹⁵Zur Orientierung kann man von einem kleinen Effekt bei $\eta^2 = 0,01$, von einem mittleren Effekt bei $\eta^2 = 0,06$ und von einem großen Effekt bei $\eta^2 = 0,14$ sprechen (Cohen 1988: 188 f.).

Tabelle 5.10 enthält die Pillai-Werte der V- und VN-Füllpartikeln aller acht Sprecherinnen in GECO-FP zum 50%-Messpunkt. Höhere Pillai-Werte beschreiben eine größere Distanz zwischen zwei Vokalverteilungen. Demnach unterscheiden Sprecherin M und F die Vokale in beiden Formen deutlicher als D und K. Für J, C, H und A lässt sich keine Aussage treffen, da ihre Modelle das Signifikanzniveau nicht erreichen.

Tabelle 5.10: Freiheitsgrade (Df), Pillai-, F- und p-Werte für die Vergleiche der Vokaldistribution in V- und VN-Formen je Versuchsperson (VP) in GECO-FP mittels einer multivariaten Varianzanalyse mit F_1 , F_2 und F_3 als abhängiger Variable, sortiert nach Pillai-Werten.

VP	Df	Pillai	F	p
F	11	0,68	6,3	<0,05
M	39	0,59	18,1	<0,001
A	9	0,29	1,0	0,456
D	34	0,29	4,4	<0,05
K	51	0,16	3,1	<0,05
J	39	0,16	2,4	0,087
C	43	0,09	1,4	0,255
H	57	0,05	1,0	0,417

Die tiefere Realisierung des ersten Formanten in VN-Formen bei Versuchsperson M in Abbildung 5.11 und auch der Gesamteindruck in Abbildung 5.10 wirft die Frage auf, ob F_1 möglicherweise durch den nachfolgenden Nasal in VN-Formen abgesenkt wird. Allerdings gibt es für eine Absenkung von F_1 in der Umgebung von Nasalen in einer Studie zum Englischen und über alle Vokale betrachtet keine Evidenz (Styler 2017: 2476). Betrachtet man die Formantkonturen aller Versuchspersonen für V- und VN-Formen in Abbildung 5.12, so ist auch hier zum Zeitpunkt des Messpunkts in der Mitte des Vokals (durchgezogene vertikale Linie) keine auffällige Koartikulation zu beobachten.

Zur Absicherung werden nochmals Messpunkte am Zeitpunkt des ersten Drittels des Vokals entnommen (vertikale gestrichelte Linie). Die Pillai-Werte für beide Zeitpunkte je Versuchsperson weichen jedoch nicht signifikant voneinander ab (ungerichteter Paired t-test, $t = 0,3$, $df = 7$, $p = 0,8$).

Der dritte Formant F_3 ist ein Maß für die Lippenrundung; je tiefer F_3 , desto stärker ist die Lippenrundung (Stevens 1997: 474). Abbildung 5.13 gibt die F_3 -Werte der Füllpartikeln wieder und vergleicht sie mit ausgewählten Referenzvokalen aller acht Sprecherinnen in GECO-FP. Die Auswahl erfolgte anhand der Überschneidung

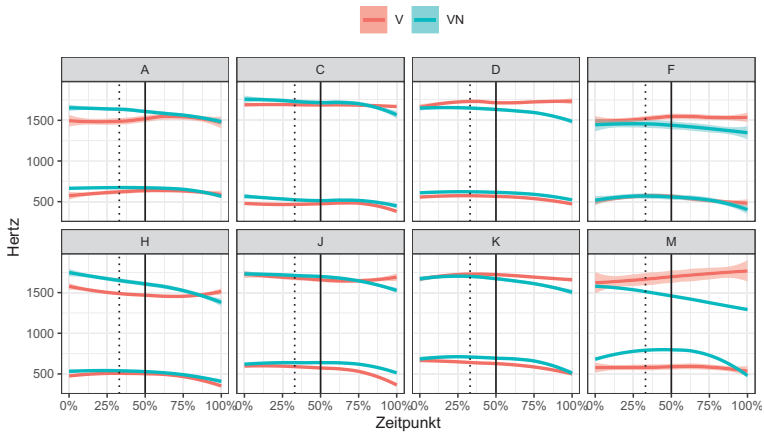


Abbildung 5.12: Gemittelte Formanttrajektorien von F_1 (unter 1000 Hz) und F_2 (über 1000 Hz) für den Vokal in V- (rot) und VN-Füllpartikeln (blau) je Versuchsperson in GECO-FP mit Zeitpunkten für mittige Messung der Formanten zum 50%-Zeitpunkt (vertikale durchgezogene Linie) und zum 33%-Zeitpunkt (vertikale gepunktete Linie).

der Füllpartikel- mit den Referenzvokalen in Abbildung 5.10. F_3 zeigt besondere Ähnlichkeit mit [ɐ] und [a:], wobei außerhalb der ersten Standardabweichung auch gerundete Werte (Ähnlichkeit mit [œ]) und ungerundete Werte (Ähnlichkeit mit [ə] und [ɛ]) auftreten.

Die Vokaldauer für vokalische Füllpartikeln ist signifikant länger als für vokalisch-nasale Füllpartikeln (vgl. Abschnitt 5.2.1.1 sowie Abschnitt 6.1.3 in der konfirmatorischen Studie zur Erklärung mittels Ausgleichskürzung). Es ist daher möglich, dass sich die Vokaldauer auch auf die Formantwerte auswirkt. Abbildung 5.14 inspiert nun den Zusammenhang zwischen Vokaldauer und den ersten beiden Formanten. Für F_1 in Abbildung 5.14 a visualisieren zwei getrennte lineare Regressionslinien die verschiedenen Korrelationen zwischen dem ersten Formanten und Vokallängen unter und über 100 ms. Für F_1 steigt die Regressionslinie unter 100 ms stark an, ist allerdings für vokalische Füllpartikelvokale nicht signifikant (gerichtete Pearson's product-moment correlation, $r = 0,27, t = 1,08, Df = 15, p = 0,1$), für Vokale in vokalisch-nasalen Füllpartikeln hingegen durchaus (gerichtete Spe-

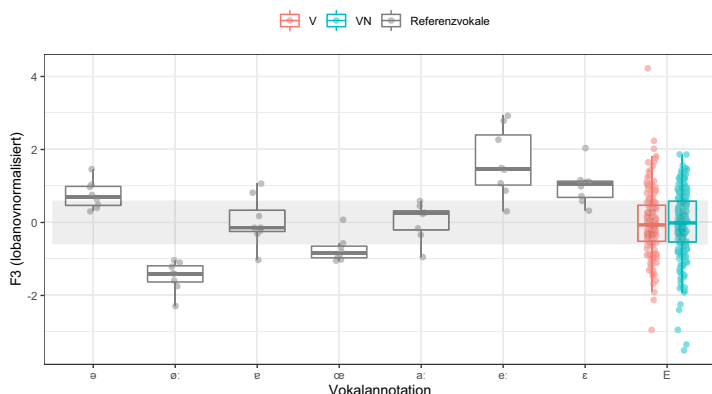


Abbildung 5.13: F₃-Werte der Vokale in vokalischen (rot) und vokalisch-nasalen (grün) Füllpartikeln in GECO-FP für ausgewählte Referenzvokale (mittige Messung im Vokal). Der graumarkierte Bereich kennzeichnet eine Standardabweichung um den F₃-Mittelwert der Füllpartikelvokale.

arman's rank correlation rho, $\rho = 0,4, S = 59047, p < 0,001$).¹¹⁶ Über 100 ms ist die Korrelation zwischen F₁ und Vokaldauer sehr klein, erreicht jedoch kein Signifikanzniveau (für Vokale in V: gerichtete Spearman's rank correlation rho, $\rho = -0,03, S = 715605, p = 0,7$; für Vokale in VN: gerichtete Spearman's rank correlation rho, $\rho = 0,05, S = 413954, p = 0,5$).¹¹⁷ Möglicherweise erreichen die F₁-Werte der Vokale in VN-Formen unter 100 ms aufgrund ihrer Kürze nicht ihr eigentliches Ziel, was einen sogenannten *target undershoot* darstellt.

Für F₂ ergeben sich keine visuellen Anzeichen für eine Zweiteilung der Regressionsgeraden. Für V-Formen ergibt sich eine negative Korrelation zwischen der Vokaldauer und der F₂-Höhe, d. h. je länger der Vokal ist, desto niedriger (und damit posteriorer) wird er (gerichtete Spearman's rank correlation rho, $\rho = -0,31, S = 1228567, p < 0,001$). Zwischen F₂ und Vokaldauer in VN-Formen besteht nur ein niedriger positiver, nicht-signifikanter Zusammenhang (ungerichtete Spearman's

¹¹⁶Aufgrund der nicht normalverteilten vokalisch-nasalen Vokaldauern unter 100 ms (Shapiro-Wilk normality test, $W = 0,95, p < 0,01$) wird eine Spearman-Korrelation verwendet.

¹¹⁷Aufgrund der nicht normalverteilten V- Vokaldauern über 100 ms (Shapiro-Wilk normality test, $W = 0,94, p < 0,001$) sowie der nicht normalverteilten VN-Vokaldauern über 100 ms (Shapiro-Wilk normality test, $W = 0,96, p < 0,001$) wird in beiden Fällen eine Spearman-Korrelation verwendet.

rank correlation rho, $\rho = 0,09$, $S = 1\,653\,958$, $p = 0,2$).

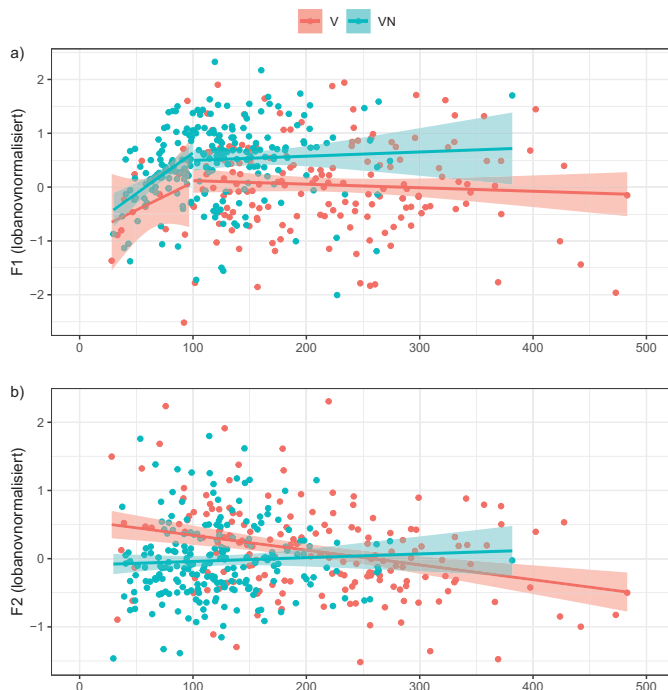


Abbildung 5.14: Zusammenhang zwischen F_1 in (a) beziehungsweise F_2 in (b) und Vokaldauer für vokalische (rot) und vokalischnasale (blau) Füllpartikel. Jeweils drei Datenpunkte über 500 ms sind nicht abgebildet, aber in der Rechnung enthalten. Lineare Regressionen in a) sind separat berechnet für Werte unter und über 100 ms.

Nachdem sich also Füllpartikelvokallängen unter 100 ms noch stark auf die Position des Formanten auswirken und möglicherweise ein *target undershoot* vorliegt, liegt es nahe, den Formantraum über alle Füllpartikel nochmal zweigeteilt darzustellen – jeweils für Vokale unter und über 100 ms Länge. Abbildung 5.15 a veranschaulicht, dass sich die Variabilität für Dauern unter 100 ms besonders für V-Füllpartikel stark erhöht, mit einem erhöhten Pillai-Wert als Indikator für die abweichenden Distributionen im Vokalraum und einer großen Effektstärke (Pillai

= 0,25, $df = 99$, $F = 11$, $p < 0,001$, $\eta^2 = 0,25$). Dies mag daran liegen, dass ihr direkter segmentaler Kontext variabler als der von VN-Formen ist, die im rechten Segment immer den Nasal enthalten und somit ein anschließendes akustisches Ziel vorgeben. Für die Vokaldauern über 100 ms in Abbildung 5.15 b ändert sich kaum etwas im Vergleich zur Gesamtschau in Abbildung 5.10 – sowohl der Pillai-Wert als auch die Effektstärke η^2 bleiben auf dem gleichen Niveau (Pillai = 0,11, $df = 297$, $F = 12$, $p < 0,001$, $\eta^2 = 0,11$).

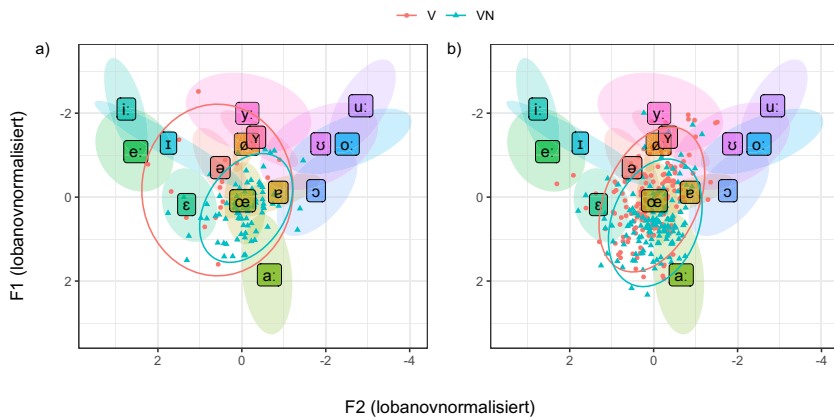


Abbildung 5.15: Lobanovnormalisierter Vokalraum aller Versuchspersonen in GECO-FP mit Datenellipsen für die Distributionen der Referenzvokale (Ellipsen enthalten zur besseren Sichtbarkeit 80% aller Referenzvokalmittelwerte) für Vokaldauern unter 100 ms (a) und über 100 ms (b). Vokalische Füllpartikeln sind mit roten Punkten, vokalisch-nasale Füllpartikeln mit blauen Dreiecken gekennzeichnet (mittige Messung im Vokal). Die entsprechenden Datenellipsen kennzeichnen 95% aller Messwerte.

Zur Quantifizierung der Überlappung zwischen V- und VN-Formantwerten wird für jeden Formanten ein lineares gemischtes Modell in Abhängigkeit von der Formausprägung der Füllpartikeln und der logarithmierten Dauer gerechnet. Versuchspersonen gehen als Random Intercept, die unkorrelierte Vokaldauer der Füllpartikel als Random Slopes in das Modell ein. Interaktionen verbessern das Modell nicht. Das Modell ist in R wie folgt spezifiziert:

```
1 | lmer(F ~ Form + log(Dauer) + (1|VP) + (0+log(Dauer)|vp), data = ...
```

Abbildung 5.16 zeigt die jeweiligen Effekte für jeden Formanten. Dort und in Tabel-

Tabelle 5.11: Linear gemischte Modelle für F_1 , F_2 und F_3 in Abhängigkeit von *Form* + $\log(\text{Dauer}) + (1/VP) + (0+\log(\text{Dauer})/VP)$, mit erklärter Varianz R^2 .

	F1	F2	F3
(Intercept)	353,8 (41,3)***	1767,1 (81,4)***	2688,1 (87,9)***
FormVN	65,6 (9,2)***	-28,9 (17,9)	8,2 (19,3)
$\log(\text{Dauer})$	45,7 (10,0)***	-28,8 (17,0)	-2,0 (17,9)
AIC	4253,2	5025,8	4994,1
Num. obs.	372	393	386
Num. groups: vp	8	8	8
R_m^2/R_c^2	0,07/0,62	0,01/0,32	0/0,24

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

le 5.11 wird deutlich, dass nur der erste Formant signifikante Unterschiede zwischen V- und VN-Formen und ihrer Dauer zeigt: In VN-Formen ist er ca. 66 Hz höher als in V-Formen (d. h. er liegt tiefer im Vokalraum) und steigt für längere Vokaldauern weiter an (bzw. bewegt sich im Vokalraum nach unten).

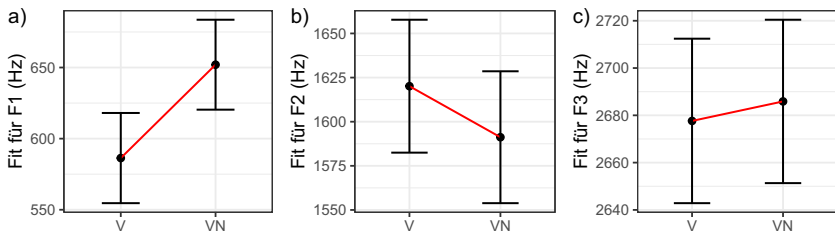


Abbildung 5.16: Effekte für a) F_1 , b) F_2 und c) F_3 des mittig gemessenen Vokals in V- und VN-Formen mit Form als Prädiktor.

5.1.6 Phonationsart

Abbildung 5.17 stellt die Verteilung der Vokale in V- und VN-Formen in GECO-FP nach Phonationsart dar. Nasale wurden nicht annotiert. VN-Formen zeigen keine laryngalisierten Vokale (von Laryngalisierung spreche ich bei phrasenfinaler Glottalisierung, vgl. Abschnitt 2.2.5). Insgesamt sind 13,9% ($n = 55$) aller V- und VN-Füllpartikeln glottalisiert. Von den V-Formen sind es 20,3%, von den VN-Formen 8,6%.

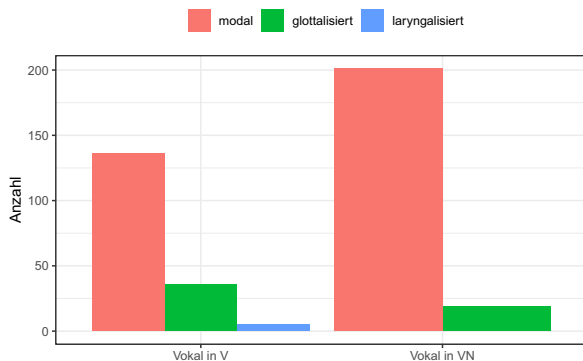


Abbildung 5.17: Anzahl der Vokale in V- und VN-Formen nach Phonationsart in GECO-FP.

Die Vokaldauern je Füllpartikelform sind in Abbildung 5.18 dargestellt, mit den genauen Werten für normalisierte und nichtnormalisierte Dauer in Tabelle 5.12. Modale Vokale in V-Formen sind nach paarweisen Vergleichen mithilfe von t-Tests mit gepoolter Standardabweichung und bonferronikorregiertem p-Wert signifikant länger als glottalisierte ($p < 0,05$) und laryngalisierte Vokale ($p < 0,05$); glottalisierte und laryngalisierte Vokaldauern in V-Formen sind nicht signifikant unterschiedlich, was wohl daran liegt, dass insgesamt nur 5 laryngalisierte Vokale vorkommen. Auch modale Vokale in VN-Formen sind signifikant länger als glottalisierte Vokale (ungerichteter Welch Two Sample t-test, $t = 4$, $df = 31, 3$, $p < 0,001$, $KI_{95\%} = [0, 11; 0, 33]$). Zusammenfassend sind modale Vokale in Füllpartikeln also länger als glottalisierte Vokale.

Tabelle 5.12: Mittelwert und Standardabweichung für nichtnormalisierte Dauer (in ms) und normalisierte Dauer (in Silben σ) für verschiedene Vokalphonationen je Füllpartikelform in GECO-FP.

Form	Phonationsart	\bar{x} (ms)	s	\bar{x} (σ)	s
Vokal in V	modal	219,0	96,3	1,2	0,6
Vokal in V	glottalisiert	174,4	83,1	1,0	0,5
Vokal in V	laryngalisiert	94,8	29,8	0,5	0,2
Vokal in VN	modal	126,6	66,7	0,7	0,4
Vokal in VN	glottalisiert	92,5	37,8	0,5	0,2

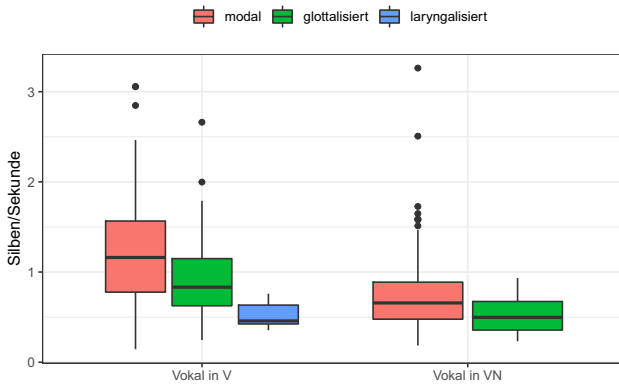


Abbildung 5.18: Normalisierte Dauer der Vokale in V- und VN-Formen nach Phonationsart in GECO-FP.

Das Histogramm in Abbildung 5.19 zeigt die Häufigkeit der Grundfrequenz über die Anzahl der modalen und glottalisierten Füllpartikeln. In 32 Fällen misst der Algorithmus 0 Hz. Die restlichen Daten zerfallen in zwei Gruppen (was aus Abschnitt 5.1.4 schon bekannt ist, vgl. Abbildung 5.7), eine erste zwischen 0 und 140 Hz (der Mittelwert dieser Häufung beträgt 98,1 Hz) und eine zweite ab 140 Hz (der Mittelwert beträgt 203,5 Hz). In Abschnitt 5.1.4 wurde gemutmaßt, dass die Daten in der ersten Gruppe möglicherweise auf glottalisierte Formen zurückzuführen seien – dieser Eindruck wird nun durch die Betrachtung der Phonationsartannotation bestätigt.

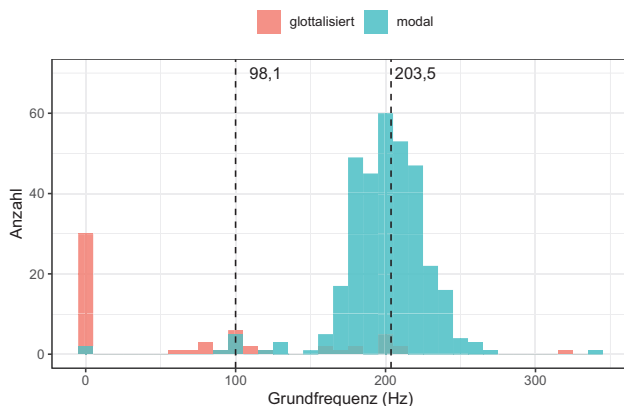


Abbildung 5.19: Phonationsarthistogramm der im Mittelpunkt aller nicht-glottalen Füllpartikeln gemessenen Grundfrequenz mit einer Klassenbreite von 10 Hz und den Mittelwerten der beiden Häufungen (gestrichelte Linien) über alle Formen je Häufung in GECO-FP.

5.2 Form und Kontext

5.2.1 Sequenzielle Kontexte

Sequenzielle Kontexte können auf Wort- oder Segmentebene analysiert werden. Zur Betrachtung und dem Vergleich der Wort-Pause-Kontexte rekatégorisiere ich die sequenziellen Kontexte der *hes*-Ebene in Anlehnung an Leeuw (2007) in Wort- und Pausenkontexte (unabhängig von ihrer Art). Tabelle 5.13 enthält die Distribution der nicht-glottalen Füllpartikelformen in Wort-Pause-Kontexten für V-, VN- und N-Formen (orthographisch repräsentiert als *äh*, *ähm* und *hm*) im Vergleich zu den deutschsprachigen Interviewdaten von Leeuw (vgl. Abschnitt 3.2.2.2).¹¹⁸

Obwohl die Interviewdaten von Leeuw (ebd.) nur unter Vorbehalt mit den hier analysierten spontansprachlichen Dialogen vergleichbar sind, ist doch auffällig, dass der Wort-FP-Wort-Kontext hier ähnliche proportionale Verteilungen für Füllpartikelformen zeigt. In den spontansprachlichen Dialogen scheint es in Wort-FP-Wort- und Pause-FP-Wort-Kontexten eine leichte Formpräferenz für *äh* zu geben, während

¹¹⁸Diese Einteilung enthält die auf der Ebene *hes* annotierten Kontexte (vgl. im Anhang Tabelle A.2) und ordnet sie den Kategorien ‚Wort‘ (*as/ps*) oder ‚Pause‘ (*ap/pp*, *ah/ph*, *at/pt*) zu (ohne Clicks, extralinguistische Entitäten und ohne Zweifelsfälle).

Tabelle 5.13: Prozentuale Anteile vokalischer, vokalisch-nasaler und nasaler Füllpartikeln an Wort-Pause-Kontexten in GECO-FP mit Anzahl (n) und bei Leeuw (2007) mit p-Wert und Teststatistik für ihre Unterschiede je Zeile (p für χ^2 -Test basierend auf 2000 Simulationen).

Kontext	GECO-FP						Leeuw (2007)			p	Test
	V	(n)	VN	(n)	N	(n)	V	VN	N		
Wort-FP-Wort	54	(73)	46	(61)	0	(0)	42	58	0	= 0,12	Fisher
Wort-FP-Pause	33	(28)	66	(56)	1	(1)	17	82	1	< 0,05	$\chi^2 = 6,9$
Pause-FP-Wort	53	(40)	41	(31)	5	(4)	6	75	19	< 0,001	$\chi^2 = 55,2$
Pause-FP-Pause	12	(5)	84	(36)	5	(2)	27	58	15	< 0,001	$\chi^2 = 16,2$

nach einem Wort und vor einer Pause sowie interpausal eher *ähm* präferiert wird.

Die Annotation der füllpartikeladjazenten Kontexte in dieser Arbeit ermöglicht nun einen detaillierten Einblick in die Verteilung der Formen je Kontext. Zu diesem Zweck betrachte ich die prominenten glottalen Formen ? und G und vergleiche sie mit nicht-glottalen vokalischen und vokalisch-nasalen Formen in den häufigsten direkten Kontexten ($n > 5$) (vgl. Abbildung 5.20). Der häufigste Kontext *as FP ps* enthält

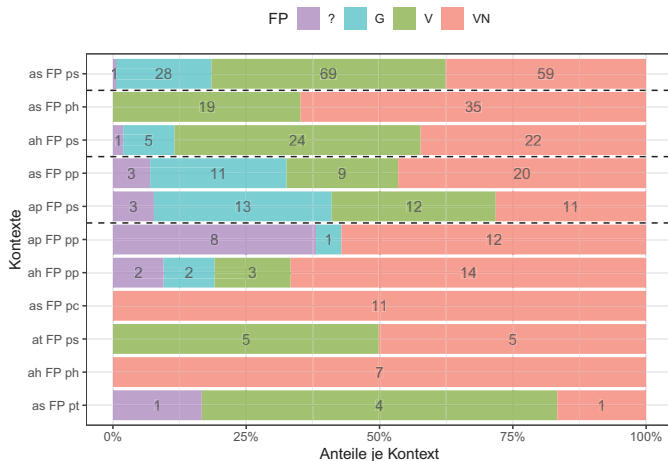


Abbildung 5.20: Anteil der V- und VN-Füllpartikelformen je Kontexte ($n > 5$) in GECO-FP (FP = Füllpartikel, Antezedentia: at = Turn, as = Segment, ap = Pause, ah = Atmungspause; Postzedentia: ps = Segment, pp = Pause, ph = Atmungspause, pc = Click). Gestrichelte Linien gliedern die Graphik in ähnliche Kontexte.

vor und nach der Füllpartikel ein Lautsegment; 15 % der Füllpartikelformen entfallen auf die glottalen Formen ? und G, 36 % auf die V und 28 % auf die VN-Formen. Der zweit- und dritthäufigste Kontext ist komplementär zueinander; im zweiten (*as FP ph*) folgt die Füllpartikel auf einen Laut und steht vor Atmung (31 % V, 57 % VN), im dritten (*ah FP ps*) erfolgt erst die Atmung, dann die Füllpartikel und anschließend wird ein Lautsegment geäußert (38 % V, 36 % VN). Der vierte und fünfte Kontext beschreibt Segmente und Pausen vor und nach Füllpartikeln. Glottale Füllpartikeln sind im Kontext *as FP pp* proportional weniger häufig als in *ap FP ps*. In den ersten fünf Kontexten kommt die V-Form prozentual am häufigsten in *as FP ps* und *ah FP ps* vor, die VN-Form hingegen am häufigsten in *as FP ph* und *as FP pp*. In *ap FP ps* sind V- und VN-Formen gleich häufig, während glottale Füllpartikeln etwa ein Drittel dieses Kontextes ausmachen. Ab dem sechsten Kontext nehmen die enthaltenen absoluten Instanzen stark ab, weswegen ich hier von einer Interpretation absehe. Allerdings ist noch erwähnenswert, dass die VN-Form zwischen Pause, stillen Pausen oder Kombinationen davon immer mindestens 50 % erreicht (in *ap FP pp*, *ah FP pp* und *ah FP ph*).

Hängt die Wahl einer Formvariante von ihrem sequenziellen Lautkontext ab? Abbildung 5.21 zeigt die Verteilung adjazenter Laute (in SAMPA) für den intersegmentalen Kontext *as FP ps* für V- und VN-Formen. 11 Kontexte kommen sowohl bei V- als auch bei VN-Formen vor, einige Treffer häufiger als einmal. Um zu ermitteln, ob die beiden Tableaus unabhängig voneinander sind, werden 10 000 Permutationen des V-Tableaus simuliert und für jede Permutation die jeweils mit dem VN-Tableau übereinstimmenden Treffer gezählt. Anschließend ermittle ich mithilfe eines Binomialtests die Wahrscheinlichkeit, 11 oder mehr Übereinstimmungen in den 10 000 Vergleichen beider Tableaus zu erhalten. Da dies in 284 von 10 000 Fällen geglückt ist, ist $p < 0,01$ (der p-Wert entspricht hier der oberen Grenze des 95 %-Konfidenzintervalls). Die Hypothese, dass die beiden Tableaus unabhängig sind, muss somit abgelehnt werden. Dies bedeutet, dass es keine Evidenz dafür gibt, dass V-Formen einen anderen segmentalen Kontext haben als VN-Formen (sonst hätte es weniger gemeinsame Treffer geben müssen, nämlich maximal acht). Die Hypothese, dass die Wahl einer V- oder VN-Form von ihren direkt vorangehenden und nachfolgenden Lauten abhängt, muss demnach abgelehnt werden.

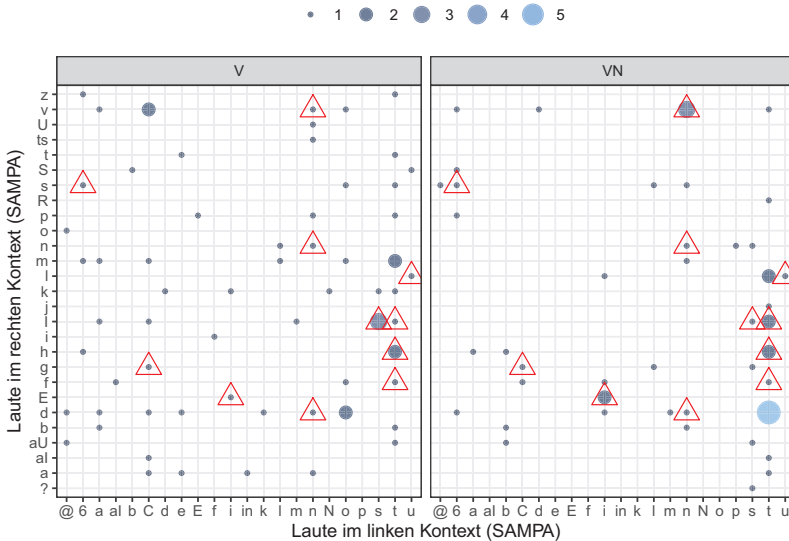


Abbildung 5.21: Anzahl und Verteilung der V- und VN-Füllpartikelformen je linkem und rechtem Kontext für alle Kontexte mit Segmenten ($n > 5$) in GECO-FP (für SAMPA vgl. Tabelle D.1). Rote Dreiecke kennzeichnen identische Kontexte für V und VN.

5.2.1.1 Dauer

Über den Zusammenhang von Füllpartikeldauer und sequenzielle Kontexte gibt es bisher hauptsächlich Studien, welche die nachfolgenden oder vorangehenden Pausen untersuchen. Hier dehne ich den Skopus der Untersuchung weiter aus und frage zusätzlich, ob sequenzielle Kontexte einen Einfluss auf die Füllpartikeldauer haben. Die Ergebnisse eines linearen gemischten Modells mit Dauer als abhängiger Variable werden in Abbildung 5.22 und Tabelle 5.14 dargestellt. Als unabhängige Variablen gehen die Kontexte mit $n \geq 10$ Füllpartikeln in die Analyse ein (dabei muss jede Form mindestens einmal vorkommen), sowie die beiden Form V und VN ein, als zufällige Schnittpunkte die Versuchspersonen. Random slopes und Interaktionen verbessern das Modell nicht.

```
1 | lmer(FP-Dauer ~ Kontext + Form + (1|VP), data = ...)
```

Füllpartikeln zwischen Segment und Pause sind mit ca. 66,7 ms signifikant länger

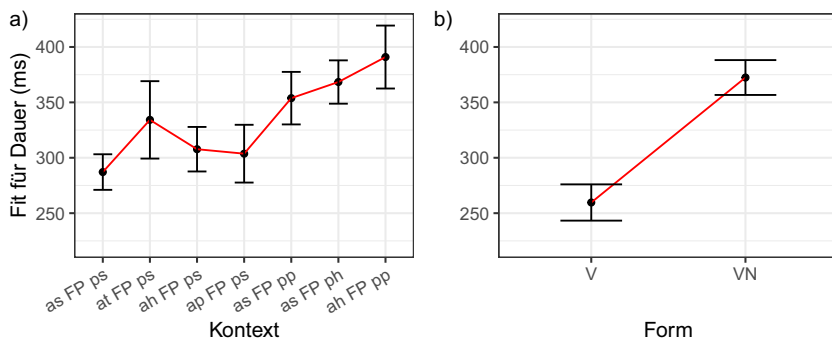


Abbildung 5.22: Effekte der Füllpartikeldauern a) für sequenzielle Kontexte wie Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) sowie Atmung-FP-Pause-Kontexten (*ah FP pp*) und b) für Form in GECO-FP.

als intersegmentale Füllpartikeln; ebenso sind Füllpartikeln zwischen einem Segment im linken und Atmung im rechten Kontext mit ca. 81,3 ms und Füllpartikeln zwischen Atmung und Pause mit 103,8 ms signifikant länger als intersegmentale Füllpartikeln. Zudem sind, wie schon bekannt, VN-Füllpartikeln mit ca. 112,8 ms signifikant länger als V-Füllpartikeln. *Post-hoc*-Tests mit Tukey-Korrektur ergeben keine weiteren signifikanten Unterschiede. Es stellt sich jedoch die Frage, ob die hier vorliegenden Effekte längerer Füllpartikeln vor Pausen und Atmungspausen mit Phrasengrenzen zusammenfallen.

Zu Pausenkontexten besteht in der Literatur mindestens seit der Studie von Clark & Fox Tree (2002) die Annahme, dass die Länge der nachfolgenden stillen Pause von der Dauer der Füllpartikel abhängt. Abbildung 5.23 zeigt die Abhängigkeit der Länge der nachfolgenden Pause von der Länge der Füllpartikelform. Die normalisierte mittlere Dauer der nachfolgenden Pause nach V-Formen beträgt $\bar{x} = 3\sigma$ ($sd = 2,68$) und nach VN-Formen $\bar{x} = 2,45\sigma$ ($sd = 1,62$). Die linearen Geraden zeigen für V-Formen einen schwachen negativen Zusammenhang ($\rho = -0,2$), für VN-Formen einen schwachen positiven Zusammenhang ($\rho = 0,1$), beide jedoch nicht signifikant. Ein lineares gemischtes Modell mit der logarithmischen Füllpartikeldauer als abhängiger Variable, Form und Dauer der Füllpartikel als unabhängiger Variable und Versuchspersonen als Random Intercepts zeigt keine signifikanten Effekte.

Tabelle 5.14: Linear gemischtes Modell für die Füllpartikeldauer in Abhängigkeit von *Kontext + Form + (1/VP)*, mit erklärter Varianz R^2 .

Dauer	
(Intercept)	227,6 (17,1)***
Kontextat FP ps	47,1 (33,7)
Kontextah FP ps	20,6 (17,7)
Kontextap FP ps	16,6 (23,9)
Kontextas FP pp	66,7 (21,8)**
Kontextas FP ph	81,3 (16,9)***
Kontextah FP pp	103,8 (27,0)***
FormVN	112,8 (13,4)***
AIC	3715,6
Num. obs.	311
Num. groups: vp	8
R_m^2/R_c^2	0,32/0,4

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

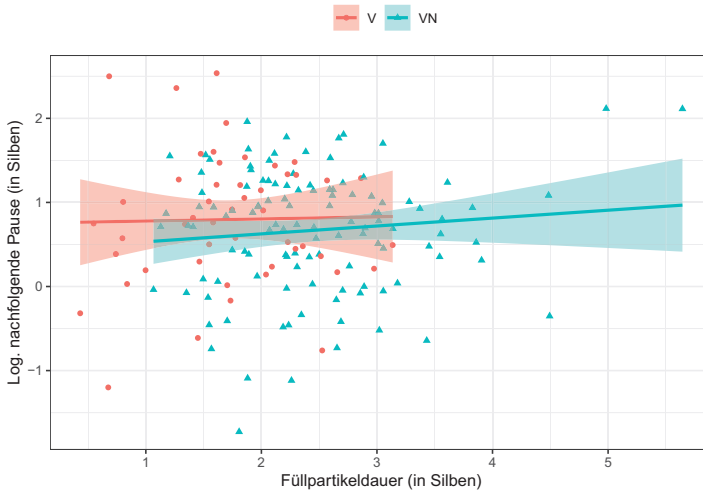


Abbildung 5.23: Füllpartikeldauer und Dauer der nachfolgenden Pause für V- und VN-Füllpartikelformen ($n_V = 49$, $n_{VN} = 116$) in GECO-FP.

```
1 | lmer(log(Postpositionsdauer) ~ Form + Formdauer + (1|VP), data = ...)
```

Damit können die Ergebnisse von Clark & Fox Tree (2002) nicht repliziert werden. Auch das Maß der ‚totalen Verzögerung‘ in Anlehnung an Rose (2015) (vgl. Abschnitt 3.2.2.1) führt zu keinem anderen Ergebnis. Es beträgt für die Dauer von V-Formen mit nachfolgender Pause $\bar{x} = 4,73 \sigma/s$ ($sd = 2,69$), für VN-Formen mit nachfolgender Pause $\bar{x} = 4,9 \sigma/s$ ($sd = 1,92$). Der Unterschied ist nicht signifikant (ungerichteter Welch Two Sample t-test, $t = -0,42$, $df = 69,54$, $p = 0,68$, $KI_{95\%} = [-1,02; 0,67]$).

Die hier vorliegenden Ergebnisse zeigen also keine Abhängigkeit der nachfolgenden Pausenlänge von der Form oder Dauer einer Füllpartikel. Füllpartikeln sind jedoch sowohl interpausal als auch präpausal signifikant länger als intersegmental (vgl. Abbildung 5.22 und Tabelle 5.14).

5.2.1.2 Grundfrequenz

Die f_0 einer Füllpartikel niedriger als die mittlere f_0 einer Versuchsperson. Aber ist die Grundfrequenz nicht-glottaler Füllpartikeln auch von ihrem Kontext abhängig? Hierfür ziehe ich die im vorigen Abschnitt 5.2.1.1 betrachteten Kategorien heran. Für jede Füllpartikel ($n_V = 179$, $n_{VN} = 179$) wird der Mittelwert gebildet. Werte unter 140 Hz (Messfehler und glottalisierte Phasen) und über 350 Hz sowie nichtnormalverteilte Residuen werden ausgeschlossen (insgesamt 1 %). Die Abhängigkeit der f_0 -Mittelwerte wird in einem linearen gemischten Modell berechnet, mit Form und Position als unabhängigen Faktoren sowie Random Intercepts für Versuchspersonen. Abbildung 5.24 visualisiert die Modellvorhersagen in Tabelle 5.15.

```
1 | lmer(F0 ~ Form + Kontext + (1|VP), data = ...)
```

V- und VN-Füllpartikeln unterscheiden sich selbst nicht in ihrer Grundfrequenz. Ausgehend von Füllpartikeln in intersegmentaler Position (*as FP ps*) sind Füllpartikeln, denen eine Turnpause (die andere Person spricht) vorangeht und ein Segment nachfolgt (*at FP ps*) signifikant höher (20,3 Hz). Selbiges gilt für Füllpartikeln, denen eine Atmungspause vorangeht und ein Segment nachfolgt (12,3 Hz höher). Der Aufbau dieser beiden Kontexte könnte auch der Beginn einer prosodischen Phrase sein, was die relativ hoch einsetzende f_0 erklären würde – sicher kann das jedoch nur in Kombination mit einer Intonationsphrasenannotation festgestellt werden (und wird dort auch bestätigt, vgl. Abschnitt 6.2.3.2). Umgekehrt gilt für die Kontexte *as FP*

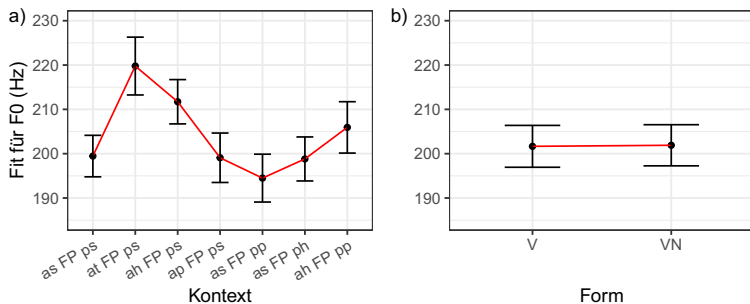


Abbildung 5.24: Effekte der f_0 -Mittelwerte für a) sequenzielle Kontexten wie Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) sowie Atmung-FP-Pause-Kontexten (*ah FP pp*) und b) Form in GECO-FP.

pp und *as FP ph*, die mit signifikant niedrigerer f_0 geäußert werden als *at FP ps* und *ah FP ps* (vgl. Tabelle 5.16), dass hier einer prosodischen Phrase vorliegen könnte (auch dies ist später für die konfirmatorische Studie in Abbildung 6.33 zu sehen).

Tabelle 5.15: Modellvorhersagen für $\bar{x}_{f_0} \sim \text{Form} + \text{Kontext} + (1|\text{VP})$, mit erklärter Varianz R^2 .

	Mittlere f_0
(Intercept)	199,3 (4,8)***
FormVN	0,2 (2,0)
Kontextat FP ps	20,3 (5,0)***
Kontextah FP ps	12,3 (2,6)***
Kontextap FP ps	-0,4 (3,6)
Kontextas FP pp	-5,0 (3,3)
Kontextas FP ph	-0,6 (2,5)
Kontextah FP pp	6,5 (4,0)
AIC	2471,9
Num. obs.	300
Num. groups: vp	8
R_m^2/R_c^2	0,07/0,8

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Dass Prosodie durchaus eine konfundierende Rolle spielen könnte, zeigt sich auch daran, dass Füllpartikeln in *at FP ps*- und *ah FP ps*-Kontexten signifikant höher produziert werden als in *ap FP ps*-Kontexten (vgl. Tabelle 5.16) – obwohl in allen

drei Kontexten eine neue prosodische Phrase beginnen könnte, ist dies für solche Kontexte mit stiller Pause vor der Füllpartikel (*ap FP ps*) weniger wahrscheinlich, da sie eine niedrigere Grundfrequenz zeigen.

Tabelle 5.16: Signifikante Kontraste der Position für f_0 (ohne Referenzniveau *as FP ps*), mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
at FP ps - ap FP ps	20,7	5,9	286,6	3,27	38,11	3,53	<0,01
at FP ps - as FP pp	25,3	5,6	286,0	8,60	41,97	4,50	<0,001
at FP ps - as FP ph	21,0	5,2	285,8	5,64	36,28	4,06	<0,01
ah FP ps - ap FP ps	12,6	3,9	286,3	0,93	24,34	3,21	<0,05
ah FP ps - as FP pp	17,2	3,7	286,1	6,17	28,29	4,63	<0,001
ah FP ps - as FP ph	12,9	3,1	286,6	3,82	22,00	4,22	<0,001

Im Rahmen dieser Analyse ist es demnach möglich, ein besseres Bild intersegmentaler Füllpartikeln zu erhalten. Hierzu extrahiere ich nochmals alle Füllpartikeln, die links und rechts jeweils zwei Wörter aufweisen (Fünfgramme). Hierbei ist natürlich nicht ausgeschlossen, dass die Wörter vor oder nach der Füllpartikel zu einer anderen Intonationsphrase zugeordnet sind. Da dies in GECO-FP nicht annotiert ist, kann dies für die hiesige Analyse nicht gelöst werden. Aus demselben Grund ist auch eine Fünfgram-Analyse von Pause-Füllpartikel-Wort oder Wort-Füllpartikel-Pause-Sequenzen nicht interpretierbar – dies gilt für solche Einheiten sogar noch eindringlicher, da Pausen oftmals Grenzen einer prosodischen Phrase markieren. Letztere Kontexte bleiben also vorerst außen vor.

Für jedes Wort in dem Fünfgram Wort-Wort-FP-Wort-Wort ($n_V = 5 \cdot 73$, $n_{VN} = 5 \cdot 61$) wird der Mittelwert gebildet. Werte unter 140 Hz (Messfehler und glottalisierte Phasen) und über 350 Hz sowie nichtnormalverteilte Residuen werden ausgeschlossen (insgesamt 19,3%) Die Abhängigkeit der f_0 -Mittelwerte wird in einem linearen gemischten Modell berechnet, mit Form, Position und ihrer Interaktion als unabhängigen Faktoren sowie *random intercepts* für Versuchspersonen und einem Index, der für jedes Fünfgram gleich ist. Abbildung 5.25 visualisiert die Modellvorhersagen in Tabelle 5.17.

```

1 | lmer(F0 ~ Form + Position + Form*Position + (1|VP) + (1|Index),
2 | data = ...)

```

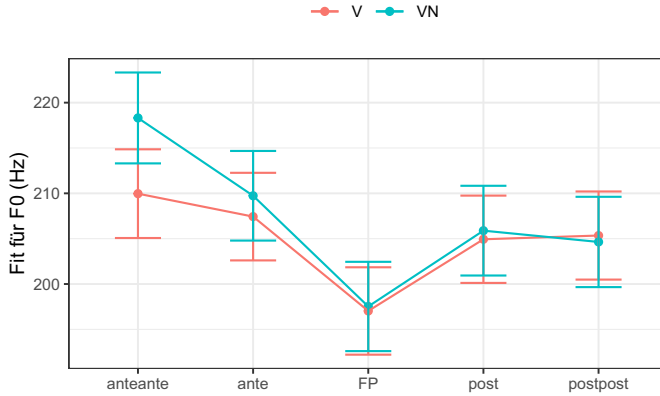



Abbildung 5.25: Modelleffekte für die f_0 -Mittelwerte zwei Wörter (*anteante*) und ein Wort (*ante*) vor sowie ein Wort (*post*) und zwei Wörter (*postpost*) nach nicht-glottalen Füllpartikeln (FP) in GECO-FP.

Tabelle 5.17: Modellvorhersagen für $\bar{x}_{f_0} \sim \text{Form} + \text{Position} + \text{Form} * \text{Position} + (1|VP) + (1|Id)$, mit erklärter Varianz R^2 .

	Mittlere f_0
(Intercept)	210,0 (4,9)***
FormVN	8,3 (3,6)*
Positionante	-2,5 (2,0)
PositionFP	-12,9 (1,9)***
Positionpost	-5,0 (1,9)**
Positionpostpost	-4,6 (2,0)*
FormVN:Positionante	-6,0 (2,8)*
FormVN:PositionFP	-7,9 (2,8)**
FormVN:Positionpost	-7,4 (2,8)**
FormVN:Positionpostpost	-9,1 (2,9)**
AIC	4277,5
Num. obs.	541
Num. groups: id	134
Num. groups: vp	8
R_m^2/R_c^2	0,07/0,8

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Das Modell zeigt für VN-Formen eine signifikante Abnahme von 8,6 Hz von *anteante* zu *ante* sowie von 12,2 Hz von *ante* zur Füllpartikel, wonach die Grundfrequenz zu den nächsten beiden Wörtern wieder zwischen 8,4 Hz (von FP zu *post*) und 7,1 Hz (von FP zu *postpost*) ansteigt, aber auf einem um 13,7 Hz niedrigeren Niveau als zum Zeitpunkt *anteante* verbleibt. Dies passt zur allgemein sinkenden f_0 -Deklination innerhalb einer prosodischen Phrase, was in der Diskussion weiter betrachtet wird. V-Formen zeigen eine signifikante Interaktion mit VN-Formen, nämlich ein um 8,3 Hz niedrigeres Ausgangsniveau zum Zeitpunkt *anteante* (vgl. Tabelle 5.17). In den anderen Positionen weicht die V- f_0 allerdings nicht signifikant von der VN- f_0 ab. Auch die f_0 zu den Zeitpunkten *anteante* und *postpost* ist für V-Formen nicht signifikant verschieden, d. h. für V-Formen kann nicht zwingend von einer generell absinkenden f_0 über das Fünfgram hinweg ausgegangen werden, sondern es wird wieder ungefähr das Ausgangsniveau erreicht. Somit wurde gezeigt, dass Füllpartikeln in Fünfgrammen eine signifikant niedrigere f_0 zeigen als ihr Kontext. Ohne Prosodieannotation bleibt die Analyse jedoch in Teilen Stipulation, da nicht auszuschließen ist, dass nach einer V-Füllpartikel zum Zeitpunkt *ante* oder *anteante* eine neue Intonationphrase beginnt und die f_0 somit wieder ansteigt.

5.2.1.3 Formanten

Die Auswertung in Abschnitt 5.1.5 hat gezeigt, dass die Messpunkte in der Mitte der Füllpartikelvokale reliabel für die Bewertung der Vokalqualität sind. Hier überprüfe ich, ob die Vokalqualität zu diesem Messpunkt von den abstrahierten adjazenten Kontexten abhängig ist. Als erklärenden Faktor verwende ich wieder die sieben häufigsten Kontexte (vgl. Abschnitt 5.2.1) mit 315 Instanzen ($n_V = 148, n_{VN} = 167$). Für F_1 und F_2 werden zwei separate gemischte lineare Modelle gerechnet, mit Kontext, Form und der logarithmisierten Dauer als Prädiktoren sowie Versuchspersonen als Random Intercepts und die unkorrelierte logarithmische Länge des Vokals als Random Slope je Versuchsperson.¹¹⁹ Interaktionen verbessern das Modell nicht. 3,8 % der F_1 - und 3,2 % der F_2 -Daten mussten aufgrund ihrer nichtnormalverteilten Residuen ausgeschlossen werden. Abbildung 5.26 zeigt die Effekte für F_1 und F_2 , die zugehörigen Werte befinden sich in Tabelle 5.18.

¹¹⁹ F_3 hängt nicht von den direkten Kontexten ab, da ein Likelihood-Vergleich eines Modells mit Kontext ($F_3 \sim \text{Kontext} + (0 + \log(\text{Dauer})/VP)$) gegen ein Nullmodell ($F_3 \sim 1 + (0 + \log(\text{Dauer})/VP)$) nicht signifikant besser ist ($p = 0,55$).

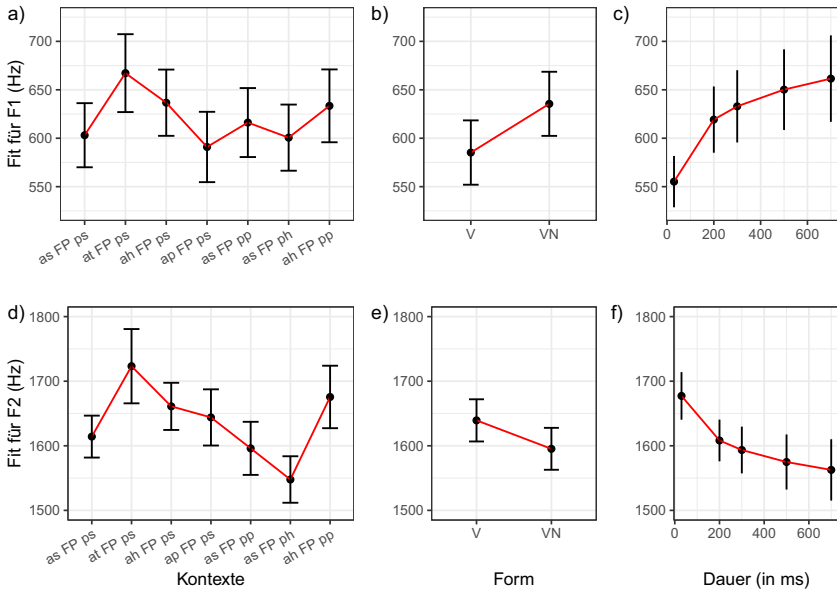


Abbildung 5.26: Modelleffekte für F_1 - und F_2 -Werte in a)/d) adjazenten sequenziellen Kontexten für Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) und Atmung-FP-Pause-Kontexte (*ah FP pp*), b)/e) für Form und c)/f) für Dauer in GECO-FP.

Ausgehend von Füllpartikeln in intersegmentaler Position (*as FP ps*) liegen die Füllpartikelvokale, denen einen Turnpause (die andere Person spricht) vorangeht und ein Segment nachfolgt (*at FP ps*) signifikant tiefer (64,1 Hz) und anteriorer (109 Hz) im Vokalraum. Füllpartikelvokale, denen eine Atmungspause vorangeht und ein Segment nachfolgt, sind auch signifikant tiefer (33,6 Hz) im Vokalraum als intersegmentale Füllpartikelvokale. Füllpartikelvokale zwischen Segment und Atmung (*as FP ph*) sind signifikant anteriorer (-66,5 Hz) als intersegmentale Füllpartikelvokale. *Post-hoc*-Tests nach Tukey für Familien mit sieben Schätzungen (vgl. Tabelle 5.19) zeigen außerdem signifikante Unterschiede für F_2 zwischen den gegensätzlichen Kontexten *at FP ps* und *as FP ph* (175,5 Hz) sowie zwischen *ah FP ps* und *as FP ph* (113,3 Hz).

Tabelle 5.18: Linear gemischte Modelle für F_1 und F_2 in Abhängigkeit von *Kontext + Form + log(Dauer) + (1/VP) + (0+log(Dauer)/VP)*, mit erklärter Varianz R^2 .

	F1	F2
(Intercept)	405,2 (48,6)***	1822,4 (95,0)***
Kontextat FP ps	64,1 (25,1)*	109,0 (51,3)*
Kontextah FP ps	33,6 (13,2)*	46,8 (25,8)
Kontextap FP ps	-12,1 (17,5)	29,7 (33,9)
Kontextas FP pp	13,1 (16,6)	-18,2 (32,1)
Kontextas FP ph	-2,5 (13,0)	-66,5 (24,9)**
Kontextah FP pp	30,3 (20,8)	61,4 (41,7)
FormVN	50,3 (11,3)***	-44,0 (21,4)*
log(Dauer)	33,8 (11,4)**	-36,4 (19,0)
AIC	3450,9	3858,8
Num. obs.	303	305
Num. groups: vp	8	8
R_m^2/R_c^2	0,07/0,62	0,08/0,3

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Tabelle 5.19: Signifikante Kontraste für multiple paarweise Vergleiche der positionalen Faktorausprägungen, mit Schätzer β und Standardabweichung (s) für F_2 , Freiheitsgraden (df), 95%-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiples Testen nach Tukey (Turnpause-FP-Segment = *at FP ps*, Atmung-FP-Segment = *ah FP ps*, Segment-FP-Atmung = *as FP ph*).

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
at FP ps - as FP ph	175,5	52,8	289,7	18,79	332,18	3,33	<0,05
ah FP ps - as FP ph	113,3	30,2	293,2	23,59	202,94	3,75	<0,01

Ähnlich wie bei den Ergebnissen für die Grundfrequenz wirken sich also Kontexte, die potenziell prosodische Phrasen beginnen können (linkerhand eine Turnpause oder Atmung, rechterhand ein Laut), signifikant auf die Lage und Höhe der Vokale in Füllpartikeln aus. Der Kontext erklärt somit einen Teil der Formvariation von Füllpartikeln.

Vokale in V-Formen liegen im Allgemeinen signifikant höher und anteriorer als VN-Formen. Die Dauer hat einen signifikant positiven Einfluss auf F_1 (je länger, desto tiefer) und einen signifikant negativen Einfluss auf F_2 (je länger, desto posteriorer).

Für diese Unterschiede der Vokalqualitäten je Form gibt es keine zufriedenstellende Erklärung. Da die Messung in der Vokalmitte erfolgte, sollte der Einfluss des nachfolgenden Nasals gering sein – zudem wirken sich zumindest im Englischen Na-

sale nach Vokalen nicht auf den ersten Formanten des Vokals aus (vgl. Styler 2017: 2476). Möglich ist, dass aufgrund der generell kürzen Vokaldauer in VN-Formen (vgl. Abbildung 5.5 und Tabelle 5.6) die Vokale in VN-Formen nicht ihr eigentliches Target erreichen können (wenn man davon ausgeht, dass die Vokale in beiden Formen dasselbe zugrundeliegende Target haben). Dies spiegelt sich in der Abhängigkeit von F_1 und F_2 von der Dauer. Neben diesen Erkenntnissen ist besonders die Erklärungskraft der Kontexte für die Vokalvariabilität aufschlussreich.

5.2.1.4 Phonationsart

Für die direkten sequenziellen Kontexte mit der absoluten Häufigkeit von $n > 5$ zeigt Abbildung 5.27 die absoluten und relativen Werte je Phonationsart. Glottalisierte Vokale stehen nur in solchen Kontexten, in denen im Ante- oder Postzedens ein Segment oder Click vorkommt (außer in *at FP ps*), nicht jedoch in interpausaler Position (*ah FP pp*, *ap FP pp*, *ah FP ph*). Somit ist die Vokalphonation nicht unabhängig von den sequenziellen Kontexte (beidseitiger exakter Fisher-Test basierend auf 2000 Simulationen, $p < 0,01$).

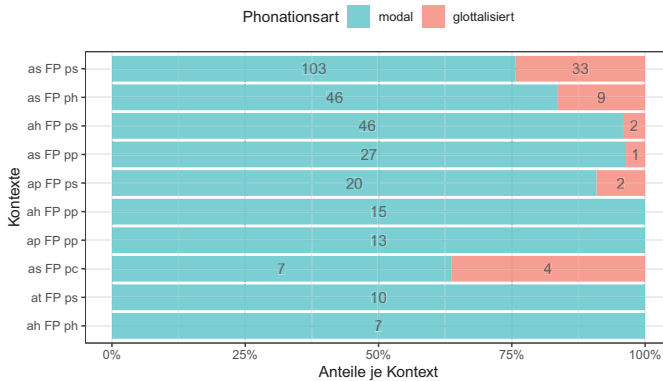


Abbildung 5.27: Absolute und relative Werte der Phonationsart in adjazenten sequenziellen Kontexten ($n > 5$) in GECO-FP (FP = nicht-glottale Füllpartikel, Antezedentia: at = Turn, as = Segment, ap = Pause, ah = Atmungspause; Postzedentia: ps = Segment, pp = Pause, ph = Atmungspause, pc = Click).

5.2.2 Dialogzüge

Tabelle 5.20 zeigt die Häufigkeiten der Annotationswerte je Dialogzugkategorie, die über die Schwelle von 25 Vorkommen hinauskommen.¹²⁰ Die Anzahl der verschiedenen Dialogzüge und die Anzahl der Wörter je Dialogzug hängt stark von den Annotationsrichtlinien ab. Zur Vergleichbarkeit mit anderen Studien gebe ich daher die prozentuale Verteilung der Wörter über die Dialogzüge an. Die in den Dialogzügen enthaltenen Anteile glottaler (FG) und nicht-glottaler Füllpartikeln (FV) unterscheiden sich signifikant von den erwarteten Häufigkeiten¹²¹ ($\chi^2 = 22,6$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0,05$). Nach *Post-hoc*-Vergleichen (Fisher-Test nach Shan & Gerstenberger (2017)) sind dafür KOMPLEXE ANTWORTEN, ERZÄHLUNGEN und TRUNKIERTE ERZÄHLUNGEN verantwortlich. In KOMPLEXEN ANTWORTEN treten glottale Füllpartikeln signifikant seltener auf, als erwartet (4 mal statt 10 mal, $p < 0,05$). Auch in ERZÄHLUNGEN treten glottale Füllpartikeln signifikant seltener auf, als erwartet (83 mal statt 71 mal, $p < 0,01$). In TRUNKIERTEN ERZÄHLUNGEN jedoch treten sie überzufällig häufig auf (6 mal statt 3 mal, $p < 0,05$).

Die relative Häufigkeit nicht-glottaler Füllpartikeln über die Wörter in einer Dialogzugkategorie (Spalte FV/W) ist für KOMPLEXE ANTWORTEN (3,1%), UNSICHERE ANTWORTEN (2,9%) und NEGATIVE ANTWORTEN (2,6%) am höchsten. Die Anzahl der Wörter und die Anzahl der nicht-glottalen Füllpartikeln sind nicht unabhängig voneinander ($\chi^2 = 144,1$, $p < 0,001$). NEGATIVE ANTWORTEN enthalten im Vergleich zur ihren enthaltenen Wörtern signifikant häufiger nicht-glottale Füllpartikeln, als erwartet (17 mal statt 7 mal, $p < 0,001$); ebenso verhält es sich für UNKLARE ANTWORTEN (28 mal statt 10 mal, $p < 0,001$) und für KOMPLEXE ANTWORTEN (48 mal statt 16 mal, $p < 0,001$). BACKCHANNELZÜGE (4 mal statt 28 mal, $p < 0,001$) und ERZÄHLUNGEN (284 mal statt 317 mal, $p < 0,001$) enthalten signifikant weniger FV-Formen als erwartet.

Dasselbe Vorgehen bestätigt zwar auch einen signifikanten Unterschied zwischen der beobachteten und erwarteten Häufigkeit von Wörtern und glottalen Formen ($\chi^2 = 23,7$, $df =$, $p < 0,05$), allerdings geht aus den *Post-hoc*-Tests hervor, dass dies

¹²⁰Die Schwelle wurde so gewählt, dass die Summe der ausgeschlossenen Werte ca. 1% aller in GECO-FP annotierten Werte beträgt. Ausgeschlossen wurden av-n (11), aw-u (10), fe-t (3), fw-t (3), i (3), ae-t (2), av-u (1), aw-fe (1), aw-t (1), und v (1).

¹²¹Ohne die Zeile mit dem Dialogzug *r* (REDEBEREITSCHAFT).

Tabelle 5.20: Häufigkeit der Annotationswerte (n), Wörter (W), Anteil der Wörter (W/n), Anteil der Wörter je Dialogzug am Datensample, Anzahl glottaler (FG) und nicht-glottaler (FV) Füllpartikeln, relative Häufigkeit nicht-glottaler Füllpartikeln (FV/W), sowie für FV die Anzahl der vokalischen (V), vokalisch-nasalen (VN) und nasalen (N) Formen und ihre Summe je Dialogzug ($n > 25$) in GECO-FP (ae-p/ae-n/ae-u/aw = POS./NEG./UNKLARE/KOMPLEXE ANTWORT, av-p = POS. ANTWORT AUF VERGEWISSERUNGSFRAGE, b = BACKCHANNELING, e = ERZÄHLUNG, e-t = TRUNKIERTE ERZÄHLUNG, fe = ENTSCHEIDUNGSFRAGE, fv = VERGEWISSERUNGSFRAGE, fw = W-FRAGE, r = REDEBEREITSCHAFT), sortiert nach FV/W.

DZ	n	Wörter (W)	W/n	W/Σ(W)	FG	FV	FV/W	V	VN	N	Σ
aw	105	1 542	14,7	3,5%	4	48	3,1%	14	31	2	47
ae-u	28	973	34,8	2,2%	3	28	2,9%	5	16	4	25
ae-n	53	666	12,6	1,5%	1	17	2,6%	8	6	2	16
fe	220	1 718	7,8	3,9%	1	21	1,2%	10	10	0	20
e-t	44	738	16,8	1,7%	6	8	1,1%	4	3	0	7
ae-p	123	1 764	14,3	4,1%	6	18	1%	11	6	1	18
e	835	31 604	37,8	72,6%	83	284	0,9%	115	144	9	268
fw	143	793	5,5	1,8%	0	6	0,8%	3	2	1	6
av-p	41	449	11	1%	1	2	0,4%	1	1	0	2
fv	70	442	6,3	1%	0	1	0,2%	0	1	0	1
b	1 863	2 791	1,5	6,4%	0	4	0,1%	3	0	0	3
r	47	66	1,4	0,2%	0	0	0%	0	0	0	0
Σ	3 572	43 546		100%	105	437		174	220	19	413

hauptsächlich auf die signifikant höhere FG-Häufigkeit in der Kategorie ABGEBROCHENE ERZÄHLUNGEN (6 mal statt 2 mal, $p < 0,01$) und die signifikant niedrigere Häufigkeit in der Kategorie BACKCHANNELING zurückzuführen ist (0 mal statt 7 mal, $p < 0,01$). Kritisch anzumerken ist, ob im Fall der ABGEBROCHENEN ERZÄHLUNG nicht eher Abbrüche denn Füllpartikeln annotiert sind, da Abbrüche im Deutschen mit glottalem Plosiv markiert werden können (Schwitalla 2012). Diese Kategorisierung ist jedoch auch eine Frage der funktionalen Perspektive und wird in dieser Arbeit ausgeblendet.

Die nicht-glottalen Füllpartikeln in Tabelle 5.20 sind nochmals detaillierter für ihre drei häufigsten abstrahierten Formen V, VN und N genauer aufgelistet.¹²² Ihre beobachteten Häufigkeiten unterscheiden sich jedoch nicht signifikant von den erwarteten Häufigkeiten, d. h. ihre Verteilung scheint unabhängig von den Dialogzügen

¹²²Die Summe für V, VN und N differiert von FV aufgrund von anderen Formen wie beispielsweise Sequenzen von Vokalen und Konsonanten, die hier jedoch aufgrund ihrer kleinen Zahl nicht berücksichtigt werden.

gen zu sein (χ^2 -Text mit simuliertem p-Wert, basierend auf 2000 Wiederholungen; $\chi^2 = 17,78, p = 0,53$).

Welche Positionen nehmen Füllpartikeln innerhalb der Dialogzüge ein? Für V und VN-Füllpartikeln sind die Dialogzüge mit $n > 7$ Vorkommen in Tabelle 5.21 aufgeführt. Füllpartikelpositionen können in der EMU-Datenbank für Anfang, Mitte und Ende eines Dialogzuges ermittelt werden, wobei ‚Anfang‘ als Gleichheit der linken Intervallgrenze von Dialogzug und Füllpartikel und ‚Ende‘ als Gleichheit der rechten Intervallgrenze definiert ist. ‚Mitte‘ bezeichnet alle sonstigen Positionen. V- und VN-Formen sind für sich genommen über die drei Positionen signifikant unterschiedlich verteilt (für V: $\chi^2 = 29,87$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0,01$; für VN: $\chi^2 = 55,53$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0,001$).

Tabelle 5.21: Häufigkeit vokalischer und vokalisch-nasaler Füllpartikeln ($\Sigma(V + VN) > 7$) nach Position im Dialogzug (DZ) und relativem Anteil am absoluten Vorkommen eines DZ in GECO-FP (ae-n/ae-p/ae-u/aw = NEG./POS./UNKLARE/KOMPLEXE ANTWORT, e = ERZÄHLUNG, fe = ENTSCHEIDUNGSFRAGE).

DZ	äh			ähm			Σ
	Anfang (%)	Mitte (%)	Ende (%)	Anfang (%)	Mitte (%)	Ende (%)	
ae-n	2 (3,8)	6 (11,3)	0 (0)	2 (3,8)	4 (7,5)	0 (0)	14
ae-p	3 (2,4)	8 (6,5)	0 (0)	1 (0,8)	5 (4,1)	0 (0)	17
ae-u	1 (3,6)	4 (14,3)	1 (3,6)	5 (17,9)	11 (39,3)	1 (3,6)	23
aw	5 (4,8)	10 (9,5)	0 (0)	13 (12,4)	18 (17,1)	1 (1)	47
e	3 (0,4)	114 (13,7)	5 (0,6)	3 (0,4)	140 (16,8)	2 (0,2)	267
fe	2 (0,9)	7 (3,2)	1 (0,5)	0 (0)	9 (4,1)	1 (0,5)	20
Σ	16	149	7	24	187	5	388

Nach *Post-hoc*-Tests für alle drei Positionen und sechs Dialogzüge sind V-Formen am Anfang KOMPLEXER ANTWORTEN signifikant häufiger als erwartet (5 mal statt 1 mal, $p < 0,01$), in ihrer Mitte jedoch seltener (10 mal statt 13 mal, $p < 0,05$). In ERZÄHLUNGEN sind V-Formen anfangs signifikant seltener (3 mal statt 11 mal, $p < 0,001$) und mittig signifikant häufiger als erwartet (114 mal statt 106 mal, $p < 0,001$).

VN-Formen zeigen für Erzählungen das gleiche Verhalten. In Antworten auf W-Fragen sind auch sie zu Beginn signifikant häufiger als erwartet (13 mal statt 4 mal, $p < 0,001$); zudem sind sie gleichzeitig mittig signifikant seltener als erwartet (18 mal statt 28 mal, $p < 0,001$). Wie auch die V-Formen sind die VN-Formen in ERZÄH-

LUNGEN anfangs signifikant seltener (3 mal statt 16 mal, $p < 0,001$) und mittig signifikant häufiger als erwartet (140 mal statt 126 mal, $p < 0,001$). VN-Formen sind zudem in UNKLAREN ANTWORTEN anfangs signifikant häufiger (5 mal statt 2 mal, $p < 0,05$) und mittig signifikant seltener als erwartet (11 mal statt 15 mal, $p < 0,05$).

Vergleicht man V- und VN-Formen jeweils bezüglich ihrer Vorkommen für jede Position einzeln miteinander, finden sich keine signifikanten Unterschiede (V vs. VN am Anfang: $\chi^2 = 7,94$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p = 0,18$; V vs. VN in der Mitte: $\chi^2 = 5,33$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p = 0,4$; V vs. VN am Ende: Exakter Test nach Fisher, $p = 0,83$). Offen bleibt, inwiefern der Beginn eines Dialogzuges beziehungsweise seine ‚Mitte‘ konfundierende Variablen wie beispielsweise Intonationsphrasen enthält, weswegen die Aussagekraft dieser Daten mit Vorsicht zu behandeln ist.

Tabelle 5.22 zeigt die Häufigkeitsverteilung der häufigsten nicht-glottalen Füllpartikelkontexte mit mindestens einem Kontext, der öfter als fünfmal in einem der Dialogzüge vorkommt.

Tabelle 5.22: Häufigkeiten nicht-glottaler Kontexte $n \geq 5$, in den Dialogzügen ae-n/ae-p/ae-u/(NEG./POS./UNKLARE/KOMPLEXE ANTWORT), e (ERZÄHLUNG) und fe (ENTSCHEIDUNGSFRAGE). Ante- und Postzedentia in den Kontexten beinhalten Atmung (ah/ph), Pausen (ap/pp), Segmente (as/ps) und Clicks (ps). Pfeile kennzeichnen Zeilen, die in die gemischten Modelle für Dauer, f_0 und Formanten eingehen.

Kontexte	Dialogzüge						Σ
	ae-n	ae-p	ae-u	aw	e	fe	
ah FP ph	0	0	0	1	6	0	7
as FP pc	1	0	0	3	7	0	11
ap FP pp	0	1	0	3	8	1	14
ah FP pp	0	1	1	1	11	0	17
→ ap FP ps	3	0	1	4	14	2	24
as FP pp	0	0	2	0	22	3	29
→ ah FP ps	3	1	5	8	24	1	47
→ as FP ph	1	1	1	3	46	0	55
→ as FP ps	4	8	5	9	88	11	134
Σ	12	12	15	32	226	18	338

Je Dialogzug können die Kontexte recht unterschiedlich verteilt sein; so kommt beispielsweise der Kontext Segment-FP-Pause nur in Erzählungen vor, jedoch nicht in Antworten auf W-Fragen. Problematisch sind die geringen oder fehlenden Werte in den einzelnen Zellen besonders für die Auswertung kontinuierlicher Parameter wie

Dauer, Grundfrequenz und Formanten. Daher betrachte ich für die weitere Auswertung nur die mit Pfeilen markierten Kontexte und die Dialogzüge *aw* und *e*.

5.2.2.1 Dauer

Die Auswahl der vier häufigsten adjazenten Kontexte und die Auswahl von Dialogzügen mit mindestens fünf Vorkommen führt zur Analyse der Daten in Tabelle 5.23 in einem linearen gemischten Modell.

Tabelle 5.23: Häufigkeit vokalischer und vokalisch-nasaler Füllpartikeln für Dialogzüge in ERZÄHLUNGEN (*e*) und KOMPLEXEN ANTWORTEN (*aw*) in den vier ausgewählten Kontexten (vgl. Tabelle 5.22) in GECO-FP.

	V	VN
aw	10	14
e	88	84

Füllpartikeldauer wird als abhängige Variable, Füllpartikelform, Kontext und Dialogzug als unabhängige Variable modelliert. Versuchspersonen gehen als zufällige Intercepts in das Modell ein. Interaktionen und Random Slopes verbessern das Modell nicht. Abbildung 5.28 und Tabelle 5.24 zeigen die Effekte des Modells.

```
lmer(FP-Dauer ~ Form + Kontext + Dialogzug + (1|VP), data = ...)
```

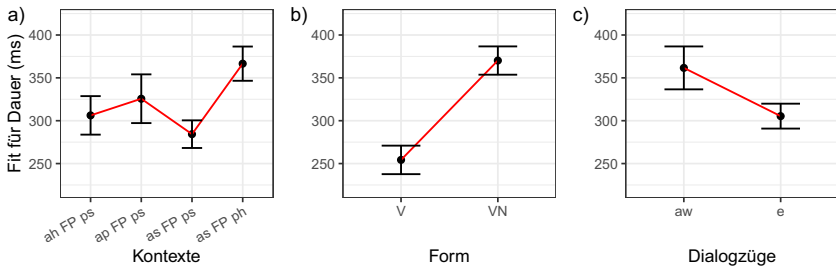


Abbildung 5.28: Modelleffekte für Füllpartikeldauern in a) sequenziellen Kontexten für Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*) und Segment-FP-Atmung-Kontexte (*as FP ph*), b) für Form und c) für Dialogstruktur (*aw* = KOMPLEXE ANTWORT, *e* = ERZÄHLUNG) in GECO-FP.

Wie in Abschnitt 5.2.1.1 sind vokalisch-nasale Formen signifikant länger als vokalische ($\beta = 115,9$ ms), während Füllpartikeln in Kontexten, in denen die Füll-

partikel zwischen einem Segment und Atmungspause steht, ca. 60 ms länger sind als Füllpartikeln in dem umgekehrten Kontext Atmung-FP-Segment (aufgrund der Verlangsamung nach Segmenten und vor Atmung wieder ein Indiz für prosodische Phrasen, vgl. Abschnitt 6.2.3.1). Der Kontext Segment-FP-Segment ist zudem *post hoc* signifikant kürzer als der Kontext Segment-FP-Atmung (Vergleich nach Turkey, Werte für *as FP ps - as FP ph*, $\beta = -82,3$, $s = 19,9$, $df = 188,2$, $KI_{95\%} = [-133,7, -30,8]$, $t = -4,1$, $p < 0,001$). Die Füllpartikellänge in ERZÄHLUNGEN ist signifikant kürzer (-56 ms) als in KOMPLEXEN ANTWORTEN (Abbildung 5.28 c). Dies könnte ein Indiz für eine erhöhte Planung bei Antworten sein.

Tabelle 5.24: Lineares gemischtes Modell für die Füllpartikeldauer in Abhängigkeit von *Dialogzug + Kontext + Form + (1|VP)*, mit erklärter Varianz R^2 .

	Dauer
(Intercept)	297,7 (29,7)***
FormVN	115,9 (17,4)***
Kontextap FP ps	19,4 (31,3)
Kontextas FP ps	-21,9 (22,3)
Kontextas FP ph	60,3 (25,6)*
Dialogzuge	-56,3 (24,1)*
AIC	2350,0
Num. obs.	196
Num. groups: vp	8
R_m^2/R_c^2	0,32/0,38

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

5.2.2.2 Grundfrequenz

Da aus Abschnitt 5.2.1.2 hervorgeht, dass sich V- und VN-Formen bezüglich der Grundfrequenz nicht signifikant unterscheiden und ein Modellvergleich auch keinen Unterschied belegt, werden diese Formen nicht separat im Modell betrachtet. Das beste Modell für die Grundfrequenz beinhaltet Kontext, Dialogzüge, und ihre Interaktion als unabhängige Variable. Versuchspersonen gehen als zufällige Intercepts in das Modell ein. Random Slopes verbessern das Modell nicht.

1 | lmer(FP-Dauer ~ Kontext * Dialogzug + (1|VP), data = ...)

Abbildung 5.29 zeigt den Interaktionseffekt von Kontexten und Dialogzügen auf die Grundfrequenz, Tabelle 5.25 ihre Werte. In KOMPLEXEN ANTWORTEN werden Füll-

partikeln in den Kontexten Pause-FP-Segment (*ap FP ps*) und Segment-FP-Atmung (*as FP ph*) signifikant tiefer geäußert als im Kontext Atmung-FP-Segment (*ah FP ps*). Dies deutet zumindest für Atmung-FP-Segment vs. Segment-FP-Atmung wieder auf einen Effekt der Position zu Beginn beziehungsweise zum Ende einer Intonationsphrase hin.

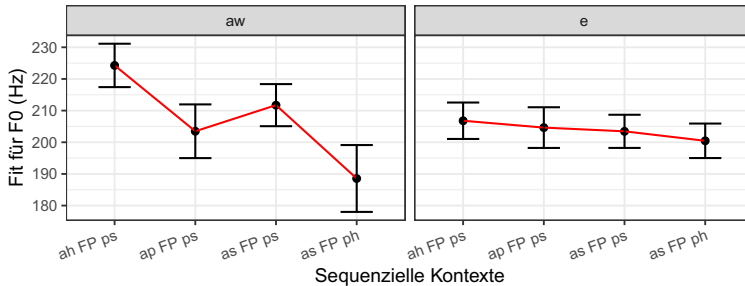


Abbildung 5.29: V- und VN-Füllpartikel- f_0 für sequenzielle Kontexte Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*) und Segment-FP-Atmung (*as FP ph*) und Dialogstruktur (aw = KOMPLEXE ANTWORT, e = ERZÄHLUNG GECCO-FP).

Tabelle 5.25: Linear gemischte Modelle für die Füllpartikeldauer in Abhängigkeit von *Dialogzug + Kontext + (1/VP)*, mit erklärter Varianz R^2 .

	Grundfrequenz
(Intercept)	224,3 (6,9)***
Kontextap FP ps	-20,8 (8,3)*
Kontextas FP ps	-12,5 (6,5)
Kontextas FP ph	-35,7 (10,5)***
Dialogzuge	-17,5 (5,5)**
Kontextap FP ps:Dialogzuge	18,6 (9,6)
Kontextas FP ps:Dialogzuge	9,2 (7,2)
Kontextas FP ph:Dialogzuge	29,4 (11,0)**
AIC	1436,7
Num. obs.	181
Num. groups: vp	8
R_m^2/R_c^2	0,07/0,58

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Tabelle 5.26 zeigt die signifikanten Kontraste aus *Post-hoc*-Tests. Füllpartikeln

im Kontext Atmung-FP-Segment zeigen eine signifikante höhere f_0 in KOMPLEXEN ANTWORTEN als in allen vier Kontexten des Dialogzugs ERZÄHLUNG. Offen bleibt, woher dieser Unterschied rührt und ob er in ungesehenen Daten replizierbar ist (da das Signifikanzniveau von $\alpha = 0,05$ nur knapp unterschritten wird).

Tabelle 5.26: Signifikante Kontraste für multiple paarweise Vergleiche der positionalen Faktorausprägungen, mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95%-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiples Testen nach Tukey (Atmung-FP-Segment = *ah FP ps*, Pause-FP-Segment = *ap FP ps*, Segment-FP-Segment = *as FP ps*, Segment-FP-Atmung = *as FP ph*, KOMPLEXE ANTWORT = *aw*, ERZÄHLUNG = *e*) in GECO-FP.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
ah FP ps,aw - as FP ph,aw	35,7	10,5	167,7	3,34	68,10	3,39	<0,05
ah FP ps,aw - ah FP ps,e	17,5	5,5	167,6	0,53	34,42	3,17	<0,05
ah FP ps,aw - ap FP ps,e	19,6	6,3	168,6	0,19	39,10	3,10	<0,05
ah FP ps,aw - as FP ps,e	20,8	5,1	169,1	5,29	36,36	4,11	<0,01
ah FP ps,aw - as FP ph,e	23,8	5,2	168,8	7,95	39,68	4,61	<0,001

5.2.2.3 Formanten

Wirkt sich die Dialogstruktur auf die Vokalqualität einer Füllpartikeln aus? In Tabelle 5.22 sind die Kontexte, die für die kontinuierlichen abhängigen phonetischen Parameter genauer analysiert werden, mit Pfeilen markiert, nämlich *ah FP ps*, *ap FP ps*, *as FP ps* und *as FP ph*. Im Folgenden vergleiche ich das Modell aus Tabelle 5.22 mit einem neuen Modell, das zusätzlich den Prädiktor Dialogzug enthält, für alle Faktorstufen mit $n > 6$: POSITIVE, NEGATIVE, UNKLARE UND KOMPLEXE ANTWORTEN sowie ERZÄHLUNGEN und ENTSCHEIDUNGSFRAGEN.¹²³ Abhängige Variable ist F_1 (in der Mitte des Vokals gemessen). Folgender Beispielcode veranschaulicht das neue Modell in R.

```
1 lmer(F1 ~ Form + Kontext + Dialogzug + log(Dauer) + (1|VP) +
  (0+log(Dauer)|VP), data = ...)
```

Der zusätzliche Prädiktor Dialogzug verbessert das alte Modell aus Abschnitt 5.2.1.3 in Likelihood-Vergleichen jedoch nicht signifikant ($p = 0,8$). Auch für F_2 verbessert sich das Modell nicht ($p = 0,1$). Für F_3 wird ein Modell mit Dialogzügen gegen

¹²³Hierzu müssen allerdings alle sieben sequenziellen Kontexte aus Tabelle 5.22 berücksichtigt werden, also auch *as FP pp*, *at FP ps* und *ah FP pp*.

ein Nullmodell verglichen. Das Nullmodell ist signifikant besser ($p < 0,05$). Somit besteht keine Evidenz für einen Einfluss der Dialogstruktur auf die Vokalqualität.

Aus Tabelle 5.20 geht hervor, dass ERZÄHLUNGEN und KOMPLEXE ANTWORTEN absolut die größte Anzahl an Füllpartikeln zeigen. Zusätzlich zeigt Tabelle 5.22, dass intersegmentale Füllpartikeln absolut in den neun häufigsten Kontexten jeweils am häufigsten sind. Daher rechne ich ein neues Modell mit den reduzierten Daten in Tabelle 5.27, in das als Dialogzugausprägungen nur ERZÄHLUNGEN und KOMPLEXE ANTWORTEN eingehen und Füllpartikeln in intersegmentalen Kontexten mit allen anderen Kontexten in einem binären Faktor *bin.Kontext* kontrastiert werden, was die Datenabdeckung deutlich erhöht.

Tabelle 5.27: Anzahl der für Formanten analysierten Füllpartikelformen in KOMPLEXEN ANTWORTEN und ERZÄHLUNGEN in intersegmentalen (*as FP ps*) und nicht-intersegmentalen (anderen) Kontexten in GECO-FP.

Dialogzug	bin.Kontext	V	VN
aw	as FP ps	3	6
aw	andere	11	25
e	as FP ps	46	41
e	andere	68	101

Für F_1 ist ein Modell mit den Daten in Tabelle 5.27 signifikant besser als ein Modell, welches nur sequenzielle Kontexte enthält ($p < 0,01$). Für F_2 und F_3 verbessert der Faktor Dialogzug das Modell wiederum nicht (für F_2 $p = 0,7$; für F_1 $p = 1$). Abbildung 5.30 und Tabelle 5.28 zeigen die Anpassungen für das beste Modell für F_1 mit Füllpartikelform und logarithmisierte Dauer sowie der Interaktion zwischen Dialogzügen und sequenziellen Kontexten in Abbildung 5.30a als unabhängigen Variablen, mit Versuchspersonen als Random Intercepts und der logarithmisierten Füllpartikeldauer je Versuchsperson als Random Slopes.

```
1 | lmer(F1 ~ Form + bin.Kontext * Dialogzug + log(Dauer) + (1|VP) +
  (0+log(Dauer)|VP), data = ...)
```

Die Höhe des Füllpartikelvokals liegt in nicht-intersegmentalen Kontexten für KOMPLEXE ANTWORTEN signifikant tiefer im Vokalraum (F_1 ist um 67 Hz höher) als in nicht-intersegmentalen Kontexten in ERZÄHLUNGEN und auch signifikant tiefer im Vokalraum (F_1 ist um 60 Hz höher) als in intersegmentalen Kontexten in ERZÄHLUNGEN (vgl. die *Post-hoc*-Tests in Tabelle 5.29). Innerhalb KOMPLEXER ANT-

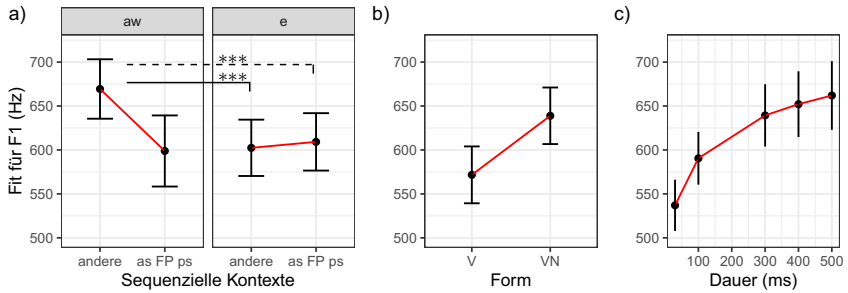


Abbildung 5.30: Modelleffekte für F_1 in Abhängigkeit von a) sequenziellen Kontexten (intersegmental = *as FP ps*; nicht-intersegmental = *andere*) je Dialogzug (*aw* = KOMPLEXE ANTWORT, *e* = ERZÄHLUNG), b) Form (V = vokalisch, VN = vokalisch-nasal) und c) Dauer in GECO-FP. Signifikante Kontraste in (a) ergeben sich aus Tabelle 5.29 (** $p < 0,001$).

Tabelle 5.28: Lineares gemischtes Modell für F_1 in Abhängigkeit von *Form + bin.Kontext* * *Dialogzug + log(Dauer) + (1/VP) + (0+log(Dauer))/VP*, mit erklärter Varianz R^2 .

	F_1
(Intercept)	334,6 (60,1)***
FormVN	67,2 (11,1)***
bin.Kontextandere	70,5 (28,6)*
Dialogzuge	10,4 (27,1)
log(Dauer)	44,4 (11,2)***
bin.Kontextandere:Dialogzuge	-77,3 (30,1)*
AIC	3174,6
Num. obs.	279
Num. groups: vp	8
R_m^2/R_c^2	0,11/0,64

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

WORTEN ist der Unterschied zwischen intersegmentalen und nicht-intersegmentalen Kontexten nicht signifikant. Nicht-intersegmentale Kontexte innerhalb von KOMPLEXEN ANTWORTEN sind demnach verschieden von allen Kontexten in ERZÄHLUNGEN. Dies würde bedeuten, dass Füllpartikeln in nicht-intersegmentaler Position von KOMPLEXEN ANTWORTEN beeinflusst werden.

Tabelle 5.29: Signifikante Kontraste für multiple paarweise Vergleiche der Faktorausprägungen für Dialogzüge (aw = KOMPLEXEN ANTWORT, e = ERZÄHLUNG) und sequenziellen Kontexten (intersegmental = *as FP ps*, nicht-intersegmental = *andere*), mit Schätzer β für F_1 , Standardabweichung (s), Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiples Testen nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
andere,aw - as FP ps,e	60,16	15,57	267,09	19,92	100,41	3,86	<0,001
andere,aw - andere,e	66,95	14,29	267,86	30,00	103,91	4,68	<0,001

Im Folgenden untersuche ich, ob sich die beiden Kontext- und Dialogzuggruppen nicht nur für Mittelpunktmessungen voneinander unterscheiden, sondern ob möglicherweise auch die Formanttrajektorien der Dialogzüge ERZÄHLUNG und KOMPLEXE ANTWORT zwischen intersegmentalen nicht-intersegmentalen Kontexten changieren. Hierzu nutze ich generalisierte additive Modelle (GAM), die die zeitnormalisierten und lobanovnormalisierten F_1 -Werte in Abhängigkeit von Füllpartikelform, Kontext und Dialogzug in dem Interaktionsprädiktor *DZ.Form.Kontext* berücksichtigen und mit Likelihood-Tests (ML-gefittet) verglichen werden. Das beste Modell (vgl. R-Code unten) enthält die Interaktion aller drei Prädiktoren (Z. 1), Kurvenglättungen für diese Prädiktoren (Z. 2) sowie zufällige Kurvenglättungen über die Prädiktoren je Versuchsperson (Z. 3) und eine Korrektur für Autokorrelation (Z. 4).

```

1 bam(lob.f1 ~ DZ.Form.Kontext +
2     s(Zeit, by = DZ.Form.Kontext, k = 10) +
3     s(Zeit, VP, by = DZ.Form.Kontext, bs = "fs", m = 1),
4     rho = rho.acf, AR.start = dat$start.event,
5     data = dat)

```

Abbildung 5.31 stellt die im Modell angepassten Trajektorien für F_1 und die Differenzen ihrer Kurven dar. Signifikante Unterschiede der Trajektorien sind vorhanden, wenn die Differenzkurve von der Nulllinie abweicht. In Abbildung 5.31 a–d ist dies nicht der Fall. In 5.31 f weichen die Trajektorien der V-Füllpartikeln in nicht-intersegmentalen Kontexten (5.31 e) zwischen 8 % und 100 % der normalisierten Zeit signifikant voneinander ab; in 5.31 h ist dies für die F_1 -Trajektorie der beiden Kurven in 5.31 g zwischen 21 % und 94 % der normalisierten Zeit der Fall. Tabelle 5.30 zeigt die Werte für die im Modell enthaltenen Glättungsdrücke. In nicht-intersegmentalen Positionen ist F_1 für V- und VN-Füllpartikeln also signifikant niedriger, wenn die Füllpartikeln in KOMPLEXEN ANTWORTEN stehen, als wenn sie

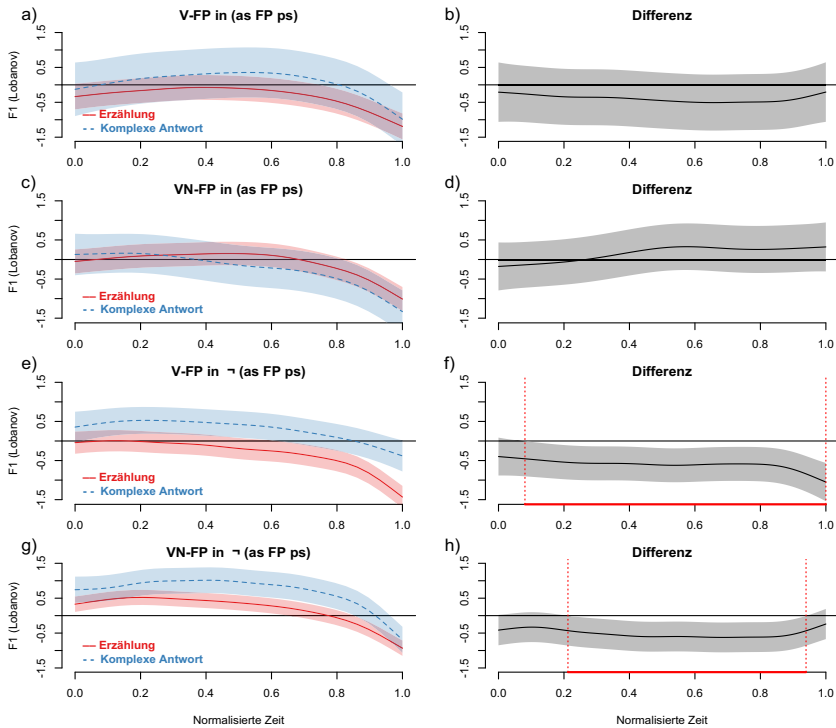


Abbildung 5.31: Angepasste F_1 -Trajektorien der Vokale in V- und VN-Füllpartikel in intersegmentalen Kontexten (as FP ps) für die Dialogzüge ERZÄHLUNG (rot) und KOMPLEXE ANTWORT (blau) in a–d) in nicht-intersegmentalen Kontexten (\neg as FP ps) in e)–h) in GECO-FP. Die rechten Bilder zeigen die Differenz der jeweils links abgebildeten Trajektorien; rote Balken kennzeichnen signifikante Abweichungen von der Nulllinie.

in ERZÄHLUNGEN vorkommen. Dieser positional bedingte Effekt könnte auf die stillen Pausen und Atmungspausen zurückgehen, die links, rechts, oder beidseits der Füllpartikel auftreten. Zusätzlich könnte hier auch ein prosodisch bedingter Effekt vorliegen, wenn die Füllpartikel am Anfang oder Ende einer Intonationsphrase steht, so dass sie sich in genau diesen nicht-intersegmentalen Kontexten befindet. Daher ist die Einbeziehung der Position innerhalb von Intonationsphrasen für weitere Aussagen über diesen Effekt unbedingt notwendig.

Tabelle 5.30: Effektive Freiheitsgrade (edf), F- und p-Werte für die unabhängigen (DZ.Form.Kontext) und zufälligen (DZ.Form.Kontext je Versuchsperson) Glättungsausdrücke der Trajektorien für F_1 über die normalisierte Zeit.

	edf	F	p
s(Zeit):DZ.Form.Kontextaw.V.intersegmental	4,11	3,90	< 0,01
s(Zeit):DZ.Form.Kontexte.V.intersegmental	5,11	12,27	< 0,001
s(Zeit):DZ.Form.Kontextaw.VN.intersegmental	4,67	7,52	< 0,001
s(Zeit):DZ.Form.Kontexte,VN,intersegmental	6,03	18,74	< 0,001
s(Zeit):DZ.Form.Kontextaw.V.andere	3,57	3,67	< 0,01
s(Zeit):DZ.Form.Kontexte.V.andere	6,40	15,26	< 0,001
s(Zeit):DZ.Form.Kontextaw,VN,andere	7,31	31,61	< 0,001
s(Zeit):DZ.Form.Kontexte.VN.andere	6,96	37,91	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.V.intersegmental	1,33	0,10	< 0,05
s(Zeit,vp):DZ.Form.Kontexte.V.intersegmental	43,79	1,34	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.VN.intersegmental	7,55	0,20	< 0,05
s(Zeit,vp):DZ.Form.Kontexte,VN,intersegmental	30,39	0,68	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.V.andere	8,41	0,33	< 0,01
s(Zeit,vp):DZ.Form.Kontexte.V.andere	52,97	1,28	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw,VN,andere	15,25	0,35	< 0,001
s(Zeit,vp):DZ.Form.Kontexte.VN.andere	47,10	1,26	< 0,001

Für F_2 gibt es wie schon für die Vokalmittelpunkte keine Dialogzugeffekte für V- und VN-Formen in intersegmentalen und nicht-intersegmentalen Kontexten.¹²⁴ Von der Analyse der F_3 -Trajektorien wird abgesehen, da ihre Interpretation unklar bleibt.

5.2.2.4 Phonationsart

Für die Verteilung von Füllpartikeln in Dialogzügen mit einer absoluten Häufigkeit von $n > 5$ je Dialogzug zeigt Abbildung 5.32 die absoluten und relativen Werte je Phonationsart. Die Verteilung weicht nicht signifikant von den erwarteten Werten ab ($\chi = 5, 1, df = 7, p = 0, 65$). Somit kann der Dialogstruktur kein Effekt bezüglich der Phonationsart zugeschrieben werden.

¹²⁴ F_2 in nicht-intersegmentalen Kontexten weichen für Erzählungen und komplexe Antworten in dem Zeitfenster von 10%–21% der normalisierten Zeitdauer zwar signifikant voneinander ab, dieser Unterschied ist aber in Tukey-korrigierten *Post-hoc*-Tests nicht mehr signifikant.

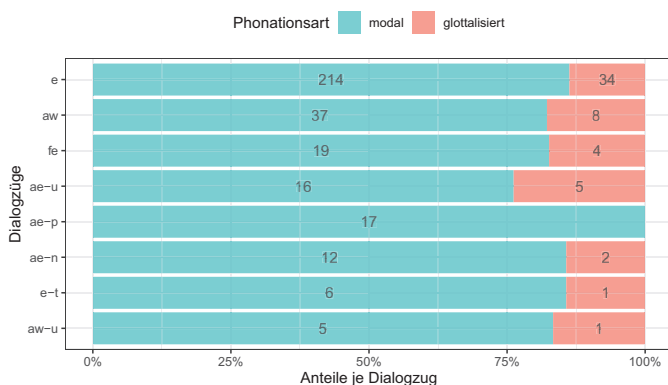


Abbildung 5.32: Absolute und relative Anteile von modalen und glottalisierten Füllpartikeln Dialogzügen ($n > 5$) in GECO-FP (e/e-t = ERZÄHLUNG/TRUNKIERTE ERZÄHLUNG, ae-p/ae-n/ae-u/aw = POSITIVE/NEGATIVE/UNKLARE/KOMPLEXE ANTWORT, fe = ENTSCHEIDUNGSFRAGE).

5.3 Zusammenfassung

In diesem Teil der Arbeit wurden die in GECO-FP annotierten Füllpartikeln bezüglich ihrer kontextfreien Frequenz, Distribution und phonetischen Form sowie ihre kontextbedingte Abhängigkeit von sequenziellen Kontexten und Dialogzügen explorativ ausgewertet. Im ersten Teil (Abschnitt 5.1) wurde die Hypothese der Formvariabilität untersucht. Die Anzahl der Füllpartikeln in GECO-FP ist nicht mit der Anzahl der Wörter oder Silben je Versuchsperson korreliert, und auch die Sprechgeschwindigkeiten aller Sprecherinnen zeigt keine Korrelation mit der Anzahl ihrer Füllpartikeln je Minute (vgl. Abschnitt 5.1.1). Füllpartikeln zeigen sowohl glottale (20 %) als auch nicht-glottale Formen (80 %), wobei sich die Häufigkeiten beider Gruppen an eine Zipfverteilung annähern (Zipf 1949), d. h. es treten einige häufige und viele seltene Segmentabfolgen auf. Zu den häufigsten Formen nicht-glottaler Füllpartikeln zählen vokalische und vokalischnasale Füllpartikeln mit vorangehendem glottalen Plosiv, gefolgt von Formen ohne glottalen Plosiv, gefolgt von GV- und GVN-Formen, die vor dem Vokal längere Sequenzen glottaler Plosive zeigen. Orthographisch zusammengefasst sind in GECO-FP *ähm*-Formen 1,3 mal häufiger als *äh*-Formen (224 vs. 179). Die häufigsten glottalen Füllpartikeln bestehen aus Se-

quenzen glottaler Plosive und vereinzelter glottalen Plosiven (vgl. Abschnitt 5.1.2). Die Füllpartikeldauer ist positiv mit der Anzahl ihrer Segmente korreliert. Vokale in VN-Formen sind signifikant kürzer als in V-Formen (vgl. Abschnitt 5.2.1.1). Die Grundfrequenz von nicht-glottalen Füllpartikeln ist interindividuell verschieden und variiert (bei Frauen in der Mitte der Füllpartikel gemessen) zwischen Durchschnittswerten von ca. 187 bis 221 Hz. Dabei liegt die Füllpartikelgrundfrequenz signifikant niedriger als der Durchschnitt aller anderen Äußerungen einer Versuchsperson (vgl. Abschnitt 5.1.4), was jedoch auch auf viele andere Wörter zutreffen wird und zwar Ergebnisse Braun & Rosin (2015) repliziert, aber nicht sehr aussagekräftig ist. Die akustische Vokalqualität von V-Formen liegt im Bereich von [œ ɐ ə ε], die von VN-Formen liegt im Bereich von [œ ɐ ə a]. Vier der acht Sprecherinnen unterscheiden die Vokalqualität von V- und VN-Formen signifikant voneinander. Insgesamt liegt der Vokal in VN-Formen signifikant tiefer (+65 Hz) im Vokalraum als der Vokal in V-Formen. Zwischen zwischen Vokaldauer und Vokalhöhe gibt es eine positive Korrelation im Bereich unter 100 ms für VN-Formen; in diesem Bereich sind sie noch höher im Vokalraum als über 100 ms. Dies könnte einen *target undershoot* darstellen, wenn man davon ausgeht, dass das Ziel die ab 100 ms erreichte Höhe ist (vgl. Abschnitt 4.4.3). Etwa 20% der V-Füllpartikelvokale werden glottalisiert, aber nur 8,6% der Vokale in VN-Formen. Modale Vokale in V- und VN-Füllpartikeln sind signifikant länger als glottalisierte Vokale. In den meisten Fällen liegt die Grundfrequenz glottalisierter Vokale unter 140 Hz oder sie kann nicht gemessen werden (vgl. Abschnitt 5.1.6).

Bis hierhin wurden alle Füllpartikeln ohne nähere Betrachtung ihrer Kontexte betrachtet. Im zweiten Teil der explorativen Auswertung (vgl. Abschnitt 5.2) wird die Form-Kontext-Hypothese anhand zweier ausgewählter Variablen – sequenzielle Kontexte und Dialogstruktur – untersucht. Füllpartikeln zeigen verschiedene Häufigkeiten je Kontext. V-Formen zeigen Präferenzen für Wort-FP-Wort und Pause-FP-Wort-Kontexte, während VN-Formen proportional am häufigsten in Wort-FP-Pause- und Pause-FP-Pause-Kontexten auftreten. Die Wahl einer Füllpartikelform kann nicht auf den sie umgebenden Lautkontext zurückgeführt werden. Füllpartikeln vor oder zwischen Pausen sind signifikant länger als in anderen Kontexten. Dabei ist ihre Dauer unabhängig von der Länge nachfolgender Pausen.

NEGATIVE, UNKLARE und KOMPLEXE ANTWORTEN enthalten die höchsten An-

teile an Füllpartikeln. Nach Segmenten und vor Atmungspausen sind Füllpartikeln mit 63 ms signifikant länger als nach Atmungspausen und vor Segmenten (vgl. Abschnitt 5.2.1.1), was mit einem Positionseffekt einhergehen könnte (finale Längung). In KOMPLEXEN ANTWORTEN sind die Füllpartikeln in diesen Kontexten nochmals deutlich länger ($60 + 56 = 116$ ms, vgl. Abschnitt 5.2.2.1). Zwar deutet dies *prima facie* auf einen Planungseffekt hin, kann aber in Berlin Dialogue Corpus compiled v. 1 (BeDiaCo_c) nicht repliziert werden.

Die Grundfrequenz von Füllpartikeln ist zwischen Turnpause oder Atmung und einem nachfolgenden Segment signifikant höher als in intersegmentalen Kontexten. Dies könnte ein Effekt einer neubeginnenden prosodischen Phrase sein. Die Grundfrequenz von nicht-glottalen Füllpartikeln ist nicht nur singularär betrachtet tiefer als die mittlere Grundfrequenz einer Sprecherin (vgl. Abschnitt 5.1.4), sondern auch signifikant tiefer als ihr umgebender segmentaler Kontext (vgl. Abschnitt 5.2.1.2). In KOMPLEXEN ANTWORTEN zeigen Füllpartikeln nach Atmung und vor Segmenten wieder eine signifikant höhere Grundfrequenz als nach Segmenten und vor Atmung, während sich jedoch die Kontexte innerhalb des Dialogzug ERZÄHLUNG nicht voneinander unterscheiden (vgl. Abschnitt 5.2.2.2).

Der Füllpartikelvokal liegt nach Turnpausen oder Atmung und vor Segmenten tiefer im Vokalraum als in intersegmentalen Kontexten – in Turnpause-FP-Segment-Kontexten liegt er zudem weiter hinten. Je länger der Vokal ist, desto tiefer und posteriorer wird er artikuliert. Der Vokal liegt in nicht-intersegmentalen Kontexten KOMPLEXER ANTWORTEN signifikant tiefer im Vokaltrakt als in intersegmentalen und nicht-intersegmentalen Kontexten in ERZÄHLUNGEN. Dieser Unterschied gilt für V- und VN-Formen und ist über den größten Teil der F₁-Trajektorie robust (vgl. Abschnitt 5.2.2.3), kann aber in BeDiaCo_c nicht repliziert werden.

Glottalisierte Vokalformen zeigen weder für Kontexte noch für Dialogzüge signifikante Abweichungen von der erwarteten Verteilung je Kontext und Dialogzug (vgl. Abschnitt 5.2.2.4) – sie kommen hauptsächlich intersegmental oder in Turnpause-FP-Segment-Kontexten vor.

Zumindest ein Teil der Formvariabilität von Füllpartikeln ist also tatsächlich kontextbedingt, wobei sequenzielle Kontexte mehr zu Erklärung beitragen als Dialogstruktur. Möglicherweise ist diese zu hierarchisch zu weit entfernt, um sich auf Füllpartikeln auszuwirken. Deutlich wird jedoch für die sequenziellen Mikro- und

die dialogstrukturellen Makrokontexte, dass die prosodischen Realisierung potenziell konfundierenden beziehungsweise erklärenden Einfluss hat. Daher wird dieser Effekt in der konfirmatorischen Studie auf einer Mesoebene untersucht, indem Intonationsphrasen einbezogen werden. Dieser Teil der konfirmatorischen Studie ist somit wiederum als explorativ anzusehen. Zusätzlich wird nun in der konfirmatorischen Studie überprüft, ob die Ergebnisse der hier vorliegenden explorativen Analyse in ungesehenen Daten repliziert werden können.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





6 Ergebnisse der konfirmatorischen Studie

Im Rahmen der Hypothese der Formvariabilität lautet die Nullhypothese, dass in BeDiaCo_c die gleichen Füllpartikelformen zu finden sind wie in GECO-FP. Im Rahmen der Form-Kontext-Hypothese lautet die Nullhypothese, dass die in BeDiaCo_c enthaltenen Füllpartikelformen die gleichen Kontextabhängigkeiten zeigen wie in GECO-FP.

6.1 Füllpartikeln

6.1.1 Allgemeine und sprecherspezifische Häufigkeit

Tabelle 6.1 gibt einen Überblick über die Anzahl der Füllpartikeln und ihre sprecherspezifische Auftretenshäufigkeit in den Dialogen aus BeDiaCo_c.¹²⁵ Die Sprechgeschwindigkeit (in Silben/Sekunde) in BeDiaCo_c rangiert zwischen 4,7 und 7 σ/s , die Anzahl der Füllpartikeln je Minute zwischen 1,9 und 11,3, mit einem Mittelwert von 4,3 FP/min. Wie in GECO-FP besteht keine signifikante Korrelation (nach Pearson) zwischen Sprechgeschwindigkeit und Füllpartikeln je Minute ($r = -0,05, t = -0,21, df = 22, p = 0,8, KI_{95\%} = [-0,44; 0,36]$). Im Gegensatz zu den explorativen Daten besteht hier jedoch eine signifikante Korrelation zwischen Tokenanzahl und Füllpartikelanzahl ($r = 0,5, t = 2,69, df = 22, p < 0,05, KI_{95\%} = [0,12; 0,75]$) und zwischen Silbenanzahl und Füllpartikelanzahl ($r = 0,54, t = 3, df = 22, p < 0,01, KI_{95\%} = [0,17; 0,77]$). Obwohl die Anzahl der Füllpartikeln über alle Versuchsperso-

¹²⁵Die Tokenanzahl wurde zur Berechnung der Sprechgeschwindigkeit um 6245 Pausen, um 21 nicht-sprachlich produzierte Geräusche, um 556 unverständliche Passagen oder extralinguistische Geräusche und 26 Clicks reduziert. Enthalten sind jedoch Füllpartikeln, insbesondere auch glottale Füllpartikeln (annotiert mit *fg*), die orthographisch nicht wiederzugeben waren und auf der Ebene *wordscor* nur impressionistisch beschrieben werden. Dies sind die 20 Werte (glottale) (1), (klicken) (1), (knacken) (8), (knarren) (8), (knarren) (8) und (schnurren) (2).

Zusatzmaterial online

Zusätzliche Informationen sind in der Online-Version dieses Kapitel (https://doi.org/10.1007/978-3-662-62812-6_6) enthalten.

© Der/die Autor(en) 2021

M. Belz, *Die Phonetik von äh und ähm*,

https://doi.org/10.1007/978-3-662-62812-6_6

Tabelle 6.1: Artikulationsdauer je Versuchsperson (VP), Anzahl der Token, Silben (σ) und Füllpartikeln (FP) sowie Sprechgeschwindigkeit (σ/s), Prozent FP je Token und Silbe sowie FP je Minute für alle Versuchspersonen (VP) in Berlin Dialogue Corpus v. 1 (BeDiaCo v. 1). Zusätzlich sind die addierten Gesamtsummen (Σ) sowie die Mittelwerte (\bar{x}) und Standardabweichung (s) über alle VP sowie geschlechtsspezifisch angegeben.

VP	Σ min	Token ^a	Silben	σ/s	FP ^b	FP/Token (%)	FP/ σ (%)	FP/min
f1	6,63	1426	1989	5,00	16	1,12	0,80	2,41
f2	8,92	1926	2737	5,11	17	0,88	0,62	1,91
f3	8,31	2085	2991	6,00	39	1,87	1,30	4,70
f4	8,01	1802	2591	5,39	16	0,89	0,62	2,00
f5	4,94	1172	1655	5,59	12	1,02	0,73	2,43
f6	6,31	1253	1802	4,76	26	2,08	1,44	4,12
f7	6,51	1551	2272	5,82	20	1,29	0,88	3,07
f8	5,94	1155	1776	4,98	15	1,30	0,84	2,53
f9	5,44	1107	1585	4,85	22	1,99	1,39	4,04
f10	8,24	2157	2993	6,05	24	1,11	0,80	2,91
f12	5,48	1355	1894	5,76	20	1,48	1,06	3,65
f13	4,33	969	1344	5,17	22	2,27	1,64	5,08
m1	9,09	1657	2548	4,67	27	1,63	1,06	2,97
m2	5,18	1326	1959	6,31	17	1,28	0,87	3,28
m3	10,88	2513	3779	5,79	64	2,55	1,69	5,88
m4	7,26	1737	2468	5,66	15	0,86	0,61	2,07
m5	4,52	996	1401	5,16	24	2,41	1,71	5,31
m6	8,14	1667	2438	4,99	62	3,72	2,54	7,62
m7	5,85	1198	1729	4,93	29	2,42	1,68	4,96
m8	6,41	1222	1796	4,67	22	1,80	1,22	3,43
m10	7,50	1850	2720	6,04	25	1,35	0,92	3,33
m13	8,51	1936	3001	5,88	63	3,25	2,10	7,40
m14	4,20	1227	1770	7,02	24	1,96	1,36	5,71
m15	10,12	1976	2936	4,84	114	5,77	3,88	11,27
Σ	166,7	37263	54174	–	735	–	–	–
\bar{x}	6,9	1552,6	2257,2	5,4	30,6	1,9	1,3	4,3
s	1,8	413,6	413,6	0,6	23,3	1,1	0,7	2,2
Σ_f	79,1	17958	25629	–	249	–	–	–
Σ_m	87,7	19305	28545	–	486	–	–	–
\bar{x}_f	6,6	1496,5	2135,8	5,4	20,8	1,4	1	3,2
s_f	1,5	404,5	566,9	0,5	7	0,5	0,3	1,1
\bar{x}_m	7,3	1608,8	2378,8	5,5	40,5	2,4	1,6	5,3
s_m	2,2	432,7	679,5	0,7	29,5	1,3	0,9	2,6

^a Ohne leere Token, extralinguistische Phänomene, stille Pausen und Clicks.

^b Alle mit fu und fy annotierte Füllpartikeln.

nen, die Anzahl der Füllpartikeln je Minute und der Anteil der Füllpartikeln je Token pro Sprecher im Gegensatz zu GECO-FP in BeDiaCo_c nicht normalverteilt ist, kann auch bei 24 Versuchspersonen nicht mit Sicherheit angegeben werden, welcher Verteilung die Daten folgen, da Tests auf Normalverteilung erst ab ca. 50 Datenpunkten robust werden.¹²⁶

Tabelle 6.2 enthält die produzierten Füllpartikelformen je Versuchsperson. Auch hier kann aufgrund der Stichprobengröße nicht mit Sicherheit angegeben werden, welcher Verteilung die Daten folgen.¹²⁷

Tabelle 6.2: Anzahl der glottalen (*fg*) und nicht-glottalen (*fv*) Füllpartikeln je Versuchsperson in BeDiaCo_c.

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f12	f13	Σ	%
fg	4	2	1	1	0	0	0	0	5	2	2	0	17	6,83
fv	12	15	38	15	12	26	20	15	17	22	18	22	232	93,17
Σ	16	17	39	16	12	26	20	15	22	24	20	22	249	
	m1	m2	m3	m4	m5	m6	m7	m8	m10	m13	m14	m15	Σ	%
fg	4	0	3	2	0	1	3	0	0	1	0	0	14	2,88
fv	23	17	61	13	24	61	26	22	25	62	24	114	472	97,12
Σ	27	17	64	15	24	62	29	22	25	63	24	114	486	
Σ													721	

6.1.2 Segmentale Formen

6.1.2.1 Nicht-glottale Formen

Auf der Ebene *segm* sind 87 verschiedene Formen von Segmentabfolgen vorhanden – 20 mehr als in GECO-FP. Durch Zusammenfassung konzeptuell ähnlicher Segmentabfolgen kann die Anzahl der Formen auf 74 segmental-abstrahierte Formen reduziert werden, das sind 19 segmental-abstrahierte Formen mehr als in GECO-FP. Tabelle 6.3 zeigt die zehn häufigsten segmental-abstrahierten Formen (für alle segmental-abstrahierten Formen s. Tabelle D.4 und D.5). Am häufigsten kommen – wie in

¹²⁶Ein Shapiro-Wilk-Test auf Normalverteilung zeigt an, dass die Anzahl der Füllpartikeln über alle Versuchspersonen nicht normalverteilt ist ($W = 0,67, p < 0,001$). Selbiges gilt für die Anzahl an Füllpartikeln je Minute ($W = 0,86, p < 0,01$) und der Anteil der Füllpartikeln je Token pro Sprecher ($W = 0,8, p < 0,001$).

¹²⁷Über alle Versuchspersonen zeigen diese Verteilungen sowohl für die glottalen Füllpartikeln ($W = 0,88, p = 0,1$) als auch für die vokalischen Füllpartikeln ($W = 0,67, p < 0,001$) Abweichungen von der Normalverteilung (Shapiro-Wilk-Test auf Normalverteilung).

GECO-FP auch – ?VN, ?V, VN und V¹²⁸ vor, allerdings in veränderter Rangfolge. Erst nach dieser Gruppe finden sich auch hier Formen, in denen dem Vokal eine Sequenz glottaler Plosive vorausgeht (GVN, GV).

Tabelle 6.3: Anzahl der zehn häufigsten segmental-abstrahierten Formen aller als *fv* annotierten Füllpartikeln in BeDiaCo_c (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal, GT = glottale Transition, # = unvollständige Glottisschwingung).

Segmente	Anzahl
VN	145
V	132
?VN	131
?V	54
GVN	29
GV	27
#VN	17
GTVN	15
#V	12
N	12

Abbildung 6.1 zeigt die Verteilung der abstrahierten Segmentformen und ihre Darstellung auf einer Log-log-Skala. Die logarithmisierten Häufigkeitswerte und ihre Ränge nähern sich einer Geraden mit negativer Steigung ($y = -x$) an, was auf eine Zipf-Verteilung der Daten deutet (Zipf 1949) und damit die Distribution für die Füllpartikeln in GECO-FP repliziert. Segmentale Füllpartikelformen verhalten sich somit ähnlich wie viele andere sprachliche Phänomene.

Abbildung 6.2 zeigt die Relation der segmental-abstrahierten Formen zu den orthographischen Repräsentationen.¹²⁹ Es lässt sich gut erkennen, dass die als *äh*, *ähm* und *hm* wiedergegebenen orthographischen Formen wie auch schon in GECO-FP phonetisch divers sind (in GECO-FP enthält orthographisches *äh* besonders häufig ?V, V und GV; orthographisches *ähm* ?VN, VN, GVN und ?VNG). In BeDiaCo_c enthält orthographisches *äh* besonders häufig V, gefolgt von ?V und GV. Die orthographische Form *ähm* repräsentiert VN und ?VN, gefolgt von GVN. Verglichen mit GECO-FP ist das Bild für die häufigsten Formen beider Korpora nicht einheitlich – unklar bleibt, woran dies liegt. Besonders zu Anfang und Ende einer Füllpartikel

¹²⁸Vokalische und vokalisch-nasale Formen mit und ohne vorangehendem glottalen Plosiv.

¹²⁹Für alle Formen, die mehr als fünf Vorkommen haben (was 121 abstrahierte Formen beziehungsweise 17,2% der 704 nicht-glottalen Füllpartikeln ausschließt).

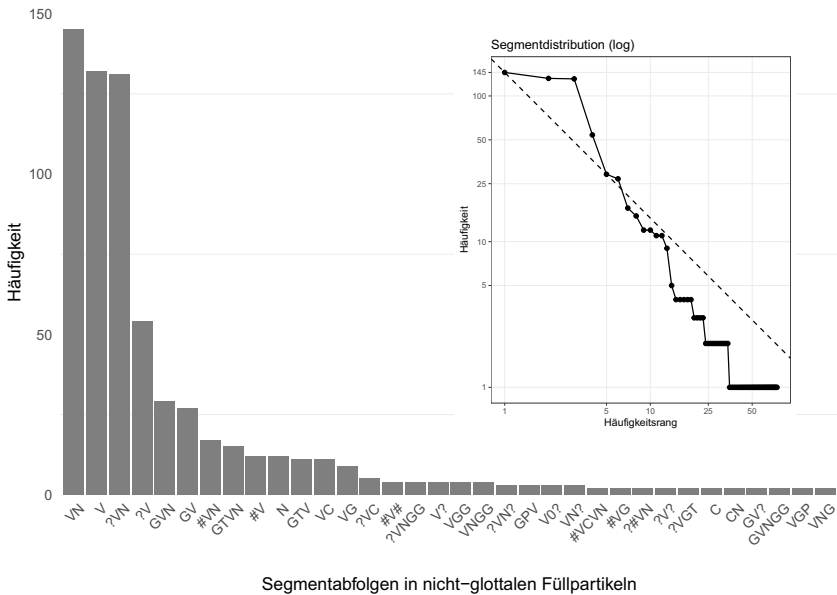


Abbildung 6.1: Anzahl der abstrahierten Segmentabfolgen ($n > 1$) in Füllpartikeln (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal, C = weitere Konsonanten, # = unvollständige Glottisschwingung, 0 = Stille, GT = glottale Transition, GG = glottalisiert mit geschlossenem Mund, GP = Glottissequenz mit mind. 50 ms Abstand). Der innere Graph enthält alle Rangfrequenzen der Füllpartikelformen (logarithmisch) und eine ideale Zipf-Verteilung (gestrichelt).

treten ein oder mehrere glottale Plosive auf. Die häufigsten *hm*-Formen in BeDiaCo_c enthalten keinen glottalen Frikativ [h].

Die drei häufigsten orthographischen Vorkommen sind *ähm* (361 mal), *äh* (306 mal) und *hm* (11 mal). *ähm* kommt somit über alle Versuchspersonen gerechnet 1,2 mal häufiger vor als *äh*, was signifikant häufiger ist (gerichteter exakter Binomialtest, $p < 0,05$). In GECO-FP ist der Faktor 1,3, allerdings sind dort nur Frauen enthalten.

Tabelle 6.4 zeigt die orthographischen Formen je Geschlecht. Frauen produzieren in BeDiaCo_c VN-Formen 1,3 mal häufiger als V-Formen, Männer nur 1,1 mal häufiger. Interessanterweise entspricht der Wert für die Frauen also exakt dem Wert in GECO-FP. Die Verhältnisse der V- und VN-Formen von Frauen und Männern sind

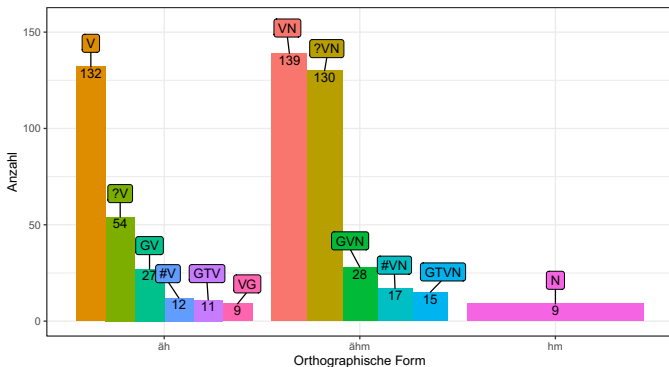


Abbildung 6.2: Zusammenfassung verschiedener segmentaler Realisierungen ($n > 5$) unter zwei orthographischen Formen (? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, GT = glottale Transition, V = Vokal, N = Nasal, # = unvollständige Glottisschwingung).

in BeDiaCo_c jedoch nicht signifikant unterschiedlich ($\chi^2 = 0,4, df = 1, p = 0,54$). Zum Vergleich ist nochmals die Relation von Füllpartikeln zu Token angegeben, die zeigt, dass Männer eine um 168 % erhöhte FP/Token-Rate haben, obwohl sie in BeDiaCo_c insgesamt nur 1347 beziehungsweise 8 % mehr Tokens produzieren wie Frauen (vgl. Tabelle 6.1).

Tabelle 6.4: Anzahl der orthographischen nicht-glottalen Füllpartikeln je Geschlecht in BeDiaCo_c, mit Kennzahl FP/Token.

	äh	ähm	hm	Σ	FP/Token
f	99	126	3	228	1,4
m	207	235	8	450	2,4
Σ	306	361	11	678	–

6.1.2.2 Glottale Formen

In BeDiaCo_c sind nur wenige glottale Formen vorhanden (vgl. Tabelle D.6 im Anhang), und auch nicht für jede Versuchsperson. Die häufigsten sind vereinzelt wahrnehmbare Abfolgen glottaler Plosive GP (die oben in GECO-FP mit ?+ markiert sind und dort 10mal vorkommen), gefolgt von der zweithäufigsten Kategorie G

(Sequenzen glottaler Plosive). Trotz der wenigen Vorkommen lässt sich eine Zipf-Distribution erahnen (vgl. Abbildung 6.3).

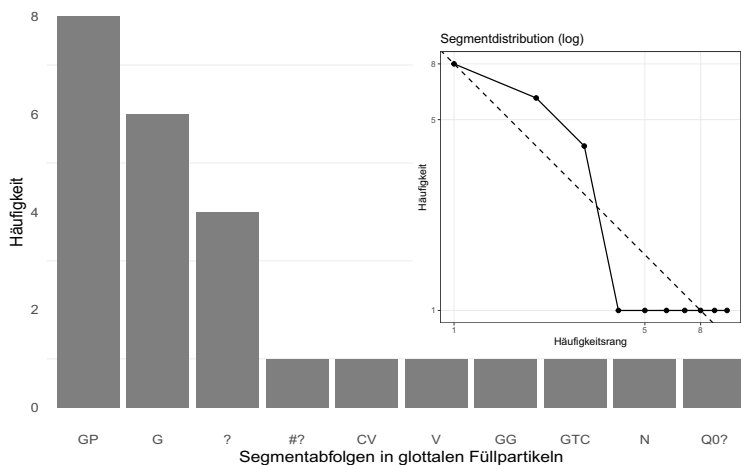


Abbildung 6.3: Anzahl der abstrahierten glottalen Segmentabfolgen in Füllpartikeln (GP = perzeptiv vereinzelt wahrnehmbare Abfolge glottaler Plosive, G = Sequenz glottaler Plosive, ? = ein bis drei glottale Plosive, # = unvollständige Glottisschwingung, C = Konsonant, V = Vokal, GG = Glottisaktivität mit geschlossenem Mund, GT = glottale Transition, N = Nasal, Q = hohe gepresste Stimme, 0 = Stille). Der innere Graph enthält alle Rangfrequenzen der Füllpartikelformen (logarithmisch) und eine ideale Zipf-Verteilung (gestrichelt).

6.1.3 Dauer

6.1.3.1 Nicht-glottale Formen

Die geringste Dauer einer nicht-glottalen Füllpartikel in BeDiaCo_c beträgt 26 ms (E beziehungsweise abstrahiert V), die längste 2387,3 ms (E→h→p\ [ɛhϕ]) beziehungsweise abstrahiert VCC. Abbildung 6.4 zeigt die Distribution der auf Silben normalisierten nicht-glottalen Füllpartikeln für die häufigsten Formen ($n > 5$).

Wie in GECO-FP besteht auch hier eine positive Korrelation (nach Pearson) zwischen der Anzahl an Segmentanteilen in der phonetischen Realisierung und der Länge von Füllpartikeln ($r = 0,39, t = 11,14, df = 702, p < 0,001, KI_{95\%} = [0,32; 0,45]$). Je mehr Segmente phonetisch realisiert werden, desto länger ist ihre Dauer. Mithi-

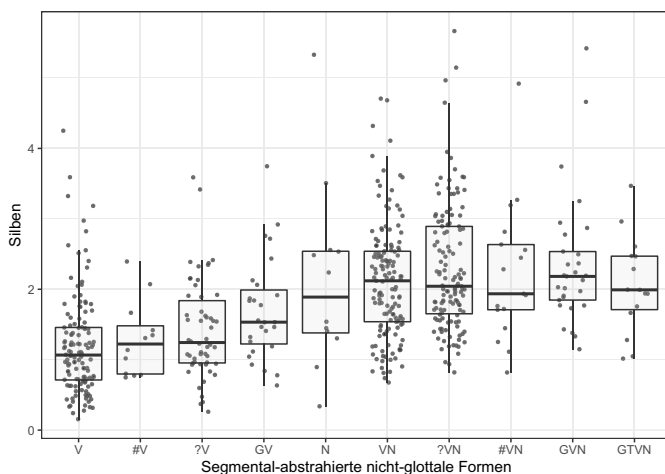


Abbildung 6.4: Dauer (in Silben) der häufigsten ($n > 5$) segmental-abstrahierten nicht-glottalen Füllpartikelformen in BeDiaCo_c (# = unvollständige Stimmklappenschwingung, ? = ein bis drei glottale Plosive, G = Sequenz von mehr als drei glottalen Plosiven, V = Vokal, N = Nasal).

fe von bonferronikorrigierten *Post-hoc*-Test werden die Formen ermittelt, die sich signifikant voneinander unterscheiden (vgl. Tabelle 6.5).

Tabelle 6.5: Bonferronikorrigierte p-Werte für paarweise t-Tests der Mittelwerte der acht segmental-abstrahierten nicht-glottalen Füllpartikelformen in Abbildung 6.4.

	V	#V	?V	GV	N	VN	?VN	#VN	GVN
#V	1	–	–	–	–	–	–	–	–
?V	1	1	–	–	–	–	–	–	–
GV	0,144	1	1	–	–	–	–	–	–
N	<0,01	0,522	0,297	1	–	–	–	–	–
VN	<0,001	<0,05	<0,001	0,777	1	–	–	–	–
?VN	<0,001	<0,01	<0,001	<0,05	1	1	–	–	–
#VN	<0,001	0,111	<0,05	1	1	1	1	–	–
GVN	<0,001	<0,01	<0,001	0,119	1	1	1	1	–
GTVN	<0,01	0,673	0,342	1	1	1	1	1	1

Die V-Form ist signifikant kürzer als die vokalischnasalen Formen VN, ?VN, #VN, GVN, GTVN und die nasale Form N. #V ist signifikant kürzer als die Formen

VN, ?VN und GVN. Die Form mit vorangehendem glottalen Plosiv ?V ist signifikant kürzer als die vokalisch-nasalen Formen VN, ?VN und GVN. Die GV-Form (mit vorangehenden Sequenzen glottaler Plosive) ist nur zu ?VN signifikant kürzer. Tabelle 6.6 gibt die durchschnittliche Dauer der acht häufigsten Formen in der auf Sprechgeschwindigkeit normalisierten Dauer in Silben und der nicht-normalisierten Dauer in Millisekunden an. Wie in GECO-FP liegt auch hier eine systematische Unterscheidung der V- und VN-Formen bzgl. ihrer Dauer vor.

Tabelle 6.6: Mittelwerte und Standardabweichungen der normalisierten Dauer in Silben (σ) pro Sekunde und der nicht-normalisierten Dauer (in Millisekunden) der acht häufigsten segmental-abstrahierten nicht-glottalen Füllpartikelformen in Abbildung 6.4.

Form	\bar{x} (in σ)	s (in σ)	\bar{x} (in ms)	s (in ms)
V	1,2	0,7	220,6	131,8
#V	1,3	0,5	241,1	96,1
?V	1,4	0,7	259,8	124,8
GV	1,7	0,7	319,4	135,2
GTVN	2,1	0,7	391,9	135,5
VN	2,1	0,8	392,5	156,7
N	2,1	1,3	417,8	197,9
#VN	2,2	1,0	423,8	168,3
?VN	2,3	0,9	434,3	279,2
GVN	2,4	0,9	455,2	199,5

Tabelle 6.7 vergleicht die Mittelwerte und Standardabweichungen der drei großen Gruppen vokalischer, vokalisch-nasaler und nasaler Füllpartikeln in BeDiaCo_c mit den Werten in Leeuw (2007) und GECO-FP. Die Werte der nicht-normalisierten Dauern von Leeuw (vgl. Tabelle 2.5) und BeDiaCo_c weichen nicht signifikant voneinander ab (t-Test mit ungleichen Stichprobengrößen und ungleichen Varianzen nach Welch-Satterthwaite). Nicht-normalisierte vokalische Formen sind in BeDiaCo_c zwar mit im Schnitt 12 ms signifikant länger als in GECO-FP ($t = 2,1, df = 558, 4, p < 0,05$), allerdings verschwindet dieser Unterschied, wenn man die normalisierte Dauer in Silben je Sekunde betrachtet. Da diese Normalisierung eben gerade die individuellen Sprechgeschwindigkeiten berücksichtigt, liefert sie sehr wahrscheinlich das reliablere Ergebnis als ein messdauerbasierter Vergleich, nach welchem in diesem Fall fälschlicherweise nicht vorhandene Längeneffekte interpretiert worden wären. Sowohl in GECO-FP also auch in BeDiaCo_c beträgt der Mittelwert einer vokalischen Füllpartikel 1,5 Silben/Sekunde und der einer vokalisch-nasalen 2,3 Silben/Sekunde.

Nasale Füllpartikeln finden sich im Bereich von 2,5–2,7 Silben/Sekunde.

Tabelle 6.7: Mittelwerte und Standardabweichung der nicht-normalisierten (ms) und normalisierten (Silben/Sekunde) Dauern vokalischer, vokalisch-nasaler und nasaler Füllpartikeln in den Daten von Leeuw (2007), GECO-FP und BeDiaCo_c.

	Vokalisch				Vokalisch-nasal				Nasal			
	ms		σ/s		ms		σ/s		ms		σ/s	
	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s
Leeuw	317,0	113,0			457,0	161,0			470,0	234,0		
GECO-FP	262,2	121,0	1,5	0,7	395,8	139,5	2,3	0,8	450,4	182,6	2,5	1,1
BeDiaCo _c	272,1	140,9	1,5	0,8	424,3	182,7	2,3	0,9	534,1	249,2	2,7	1,2

Die Vokaldauer ist in GECO-FP in vokalischen Füllpartikeln signifikant länger als in vokalisch-nasalen Füllpartikeln – dies wird für BeDiaCo_c auch angenommen. Zur Überprüfung wird dasselbe lineares gemischte Modell mit Dauer als abhängiger Variable und Form als unabhängiger Variable gerechnet. Versuchspersonen gehen als Random Intercepts ein, mit Random Slopes für die Form je Versuchsperson. Abbildung 6.5 visualisiert die Vorhersagen des Modells, Tabelle 6.8 enthält die Werte. Das Modell wird folgendermaßen in R dargestellt.

```
1 | lmer(Vokaldauer ~ Form + (1+Form|VP), data ...
```

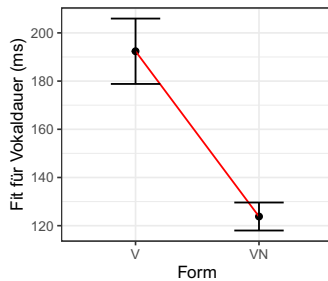


Abbildung 6.5: Effekt der Vokaldauer in Abhängigkeit von der Füllpartikelform in BeDiaCo_c.

Die Vokaldauer in vokalischen Füllpartikeln ist mit 31,7ms signifikant länger als in vokalisch-nasalen Füllpartikeln. Die Effektstärke ist allerdings kleiner als bei GECO-FP (dort ca. 69ms). Das beobachtete Phänomen wird von Maddieson (1985) als Vokalkürzung in geschlossener Silbe (Engl. ‚closed syllable vowel shortening‘)

in typologisch verschiedenen Sprachen beschrieben. Der Effekt wird in rezenter Literatur auch für englische Nicht-Wörter beschrieben (Katz 2012); Katz nennt die Kürzung von Vokalen in Silben mit einem Konsonanten in der Koda im Vergleich zu offenen Silben Simplexausgleichskürzung (Engl. ‚simplex compensatory shortening‘).

Tabelle 6.8: Lineares gemischtes Modell für die Vokaldauer in Abhängigkeit von *Form* + (*1+Form/VP*), mit erklärter Varianz R^2 .

Vokaldauer	
(Intercept)	196,9 (10,3)***
FormVN	-31,7 (10,0)**
AIC	7588,6
Num. obs.	652
Num. groups: vp	24
R_m^2/R_c^2	0,03/0,27

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

6.1.3.2 Glottale Formen

Die häufigsten glottalen Formen mit mehr als drei Vorkommen sind ?, G und GP.¹³⁰ Die Form ? ist in BeDiaCo_c (anders als in GECO-FP) jedoch nicht signifikant kürzer als die Form G und GP (paarweiser bonferronikorreierter t-Test, jeweils $p > 0,05$). Tabelle 6.9 enthält die durchschnittliche Dauer der drei Formen. Der geringste Wert einer glottalen Füllpartikel liegt bei 29,9 ms (ein einzelner glottaler Plosiv), der größte Wert bei 701,7 ms (eine Sequenz glottaler Plosive).

Tabelle 6.9: Mittelwerte und Standardabweichungen der normalisierten (in Silben σ) und der nicht-normalisierten Dauer (in Millisekunden) der drei häufigsten segmental-abstrahierten glottalen Füllpartikelformen in BeDiaCo_c.

Form	\bar{x} (in σ)	s (in σ)	\bar{x} (in ms)	s (in ms)
?	0,3	0,2	58,0	32,7
GP	1,4	0,8	272,3	182,2
G	1,5	0,9	293,4	206,4

¹³⁰GP entspricht in BeDiaCo_c der Form ?+ in GECO-FP.

6.1.4 Grundfrequenz

Für die Analyse der Grundfrequenz (f_0) werden nur die nicht-glottalen Formen betrachtet, da in den glottalen Formen die Stimmlippenschwingung zu unregelmäßig für die reliable Ermittlung der Grundfrequenz ist. Die Grundfrequenz messe ich wieder in der Mitte der Füllpartikel, wobei vorangehende und nachfolgende mögliche andere Lautsegmente nicht berücksichtigt werden. Dennoch kann es zu Fällen kommen, in denen die Grundfrequenz in Teilen nicht ermittelt werden kann. Zur ersten Übersicht zeigt das Histogramm in Abbildung 6.6 die Häufigkeit der Grundfrequenz über die Anzahl der vokalischen und vokalisch-nasalen Füllpartikeln. In 55 Fällen misst der Algorithmus 0 Hz.

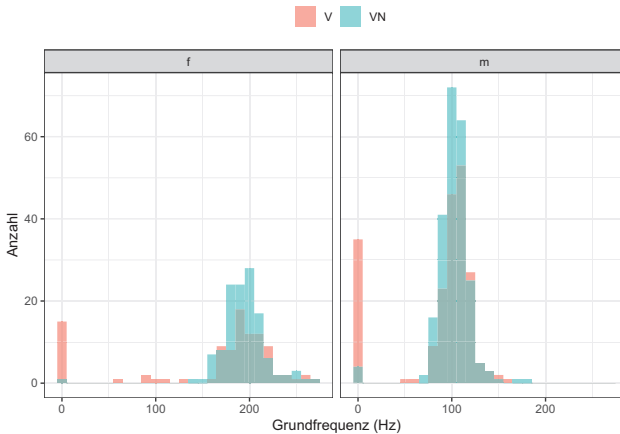


Abbildung 6.6: Histogramm der im Mittelpunkt aller FV gemessenen Grundfrequenz mit einer Klassenbreite von 10 Hz in BeDiaCo_c für Frauen (f) und Männer (m).

Die Füllpartikeln der weiblichen Versuchspersonen bilden zwar wie in GECO-FP eine Datengruppen zwischen 0 und 140 Hz (der Mittelwert dieser Häufung beträgt 103,8). Bei den männlichen Versuchspersonen ist keine Gruppierung zu erkennen.

Werden Füllpartikeln auch in BeDiaCo_c mit tieferer f_0 als der Mittelwert aller Wörter einer Versuchsperson produziert? Hierfür werden wieder alle Füllpartikeln je Versuchsperson $\bar{f}_{0,vp}^{FP}$ mit der mittleren $\bar{f}_{0,vp}$ je Versuchsperson verglichen. Für Frauen werde ich aufgrund der bimodalen Gruppierung weiterhin nur Grundfrequenzen

über 140 Hz betrachten.

Tabelle 6.10: Anzahl der in der Mitte gemessenen V- und VN-Füllpartikeln und Wörter je Grundfrequenzschwelle in BeDiaCo_c.

Sexus	Grundfrequenz	Füllpartikeln	Wörter
f	0 Hz	16	4719
f	> 0 und < 140 Hz	6	496
f	≥ 140 Hz	203	12553
f	Σ	225	17768
m	0 Hz	39	5924
m	≥ 140 Hz	402	13132
m	Σ	441	19056

Tabelle 6.11 stellt die Mittelwerte und Standardabweichungen der Einzelmesswerte der beiden Füllpartikelgrundfrequenzgruppen und die über alle Äußerungen gemittelte $\bar{f}_{0,vp}$ mit dem gleichen Wertebereich dar. Eine Evaluierung erfolgt mithilfe paarweiser Vergleiche eines gerichteten t-Tests.¹³¹ Gerichtete t-Test werden durchgeführt, da sowohl die Ergebnisse von GECO-FP als auch die in der Literatur gefundenen Werte eine niedrigere $\bar{f}_{0,vp}^{FP}$ ergaben als die mittlere Grundfrequenz einer Versuchsperson (vgl. Braun & Rosin 2015). Füllpartikeln zeigen für Frauen (oberhalb von 140 Hz) signifikant niedrigere Grundfrequenzen ($t = -3,5$, $p < 0,01$, $KI_{95\%} = [-Inf; -5,44]$) als andere Wörter ($n = 12553$). Nur eine weibliche Versuchsperson (f5) produziert die Füllpartikeln höher als ihre mittlere Grundfrequenz (dies war auch in GECO-FP für genau eine Sprecherin der Fall). Auch die Füllpartikeln der Männer sind signifikant niedriger ($t = -6,1$, $p < 0,001$, $KI_{95\%} = [-Inf; -3,68]$) als ihre sonstigen Wörter ($n = 13132$).

¹³¹Die Wahl des t-Tests statt des nicht-parametrischen Wilcoxon-Rangsummentest wird durch die Normalverteilung der Differenzen der zu vergleichenden Werte motiviert. Somit kann der mächtigere t-Test verwendet werden.

Tabelle 6.11: Mittelwerte und Standardabweichung für die in der Mitte von Füllpartikeln ($\bar{f}_{0_{vp}}^{FP}$) und von allen Wörtern ($\bar{f}_{0_{vp}}$) einer Versuchsperson gemessene Grundfrequenz in Hertz in BeDiaCoC für Frauen ($f_0 > 140$ Hz) und Männer ($f_0 > 0$ Hz).

VP	Füllpartikeln		Wörter		VP	Füllpartikeln		Wörter	
	\bar{x}	s	\bar{x}	s		\bar{x}	s	\bar{x}	s
f1	171,7	12,0	185,9	31,3	m1	97,4	14,5	100,7	22,7
f2	199,5	60,0	205,8	36,1	m2	83,2	8,5	86,4	24,3
f3	193,5	12,2	213,2	40,9	m3	98,9	11,9	104,5	23,3
f4	187,0	13,4	197,4	22,6	m4	97,6	6,1	99,1	21,6
f5	233,1	24,0	221,8	34,1	m5	109,3	18,3	113,2	18,8
f6	206,8	19,0	239,4	35,6	m6	106,7	12,3	114,2	19,4
f7	198,0	16,4	210,9	27,8	m7	110,5	6,7	114,8	11,8
f8	221,0	14,7	222,9	21,2	m8	132,7	16,1	142,8	26,7
f9	206,4	9,3	209,7	22,5	m10	90,9	9,4	99,4	19,7
f10	194,4	22,2	204,7	32,8	m13	106,6	9,4	106,9	9,3
f12	189,2	19,5	205,5	30,7	m14	97,3	7,6	104,1	26,0
f13	189,4	27,8	204,7	40,2	m15	107,6	10,1	115,2	18,7

6.1.5 Formanten

Zur Ermittlung der Formanten und damit der Vokalqualität der Füllpartikeln extrahiere ich aus jeder Füllpartikel einer Versuchsperson F_1 und F_2 in der Mitte des Vokals; ebenso für die Referenzvokale. Alle Formanten wurden zuvor manuell in der EMU-Datenbank korrigiert. Alle Referenzvokale einer Qualität werden je Versuchsperson zu einem einzigen Vertreter dieser Klasse gemittelt. Anschließend wird mit den Werten der Referenzvokale und der Füllpartikelvokale eine formant-intrinsische, sprecher-intrinsische und vokal-extrinsische Lobanovnormalisierung durchgeführt (vgl. Abschnitt 4.4.3). Die so normalisierten Vokalwerte werden für F_1 und F_2 in Abbildung 6.7 in einem Vokalraum dargestellt, der die Streuung der im quasi-stationären Vokalteil (50 %) gemessenen Referenz und Füllpartikelvokale als Datenellipsen zeigt.

Die Vokalqualitäten verhalten sich ähnlich zu denen in GECO-FP. Die Vokalqualität für vokalisches Füllpartikeln bewegt sich hauptsächlich im Bereich von [e e ə] und [ɛ]. Die Vokale vokalisch-nasaler Füllpartikeln überlappen mit den Bereichen von [œ e ə] und [a]. Im untermittelhohen Bereich zwischen [ɛ] und [a] könnte die Vokalqualität des im Deutschen nicht phonologisch vorhandenen [æ] getroffen werden. Als Maß für die Vokalüberlappung berechne ich den Pillai-Wert (vgl. Abschnitt 4.4.3). Je höher der Pillai-Wert, desto größer ist die Distanz von F_1 , F_2 und

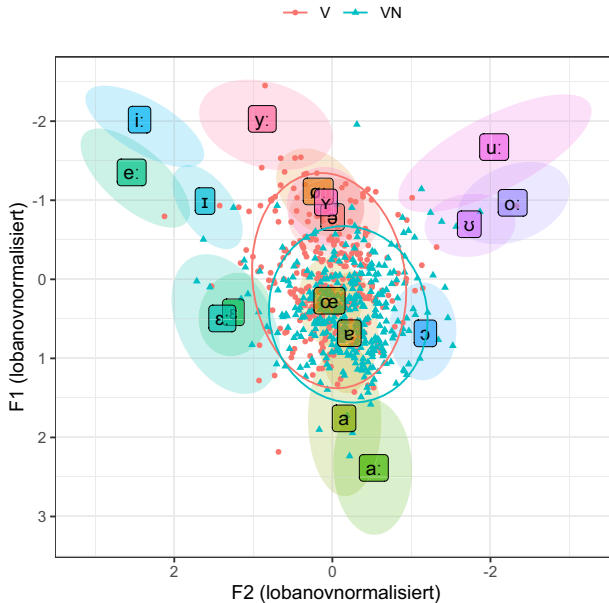


Abbildung 6.7: Lobanovnormalisierter Vokalraum aller Versuchspersonen in BeDiaCo_c mit Datenellipsen für die Distributionen der Referenzvokale (Ellipsen enthalten zur besseren Sichtbarkeit 80 % aller Referenzvokalmittelwerte). Vokalische Füllpartikeln sind mit roten Punkten, vokalisch-nasale Füllpartikeln mit blauen Dreiecken gekennzeichnet (mittige Messung im Vokal). Die entsprechenden Datenellipsen kennzeichnen 95 % aller Messwerte.

F₃ zwischen zwei Vokalen. Der in einer MANOVA ermittelte Pillai-Wert mit F₁, F₂ und F₃ als abhängiger Variable für die Distribution der Vokale in V- und VN-Formen in Abbildung 6.7 beträgt 0,15 ($Df = 669, F = 40,26, p < 0,001, \eta^2 = 0,15$). Die Vokale beider Formen überlappen sich also zum größten Teil (F₃ als Maß für die Lippenrundung kann in der zweidimensionalen Graphik nicht dargestellt werden, wird allerdings weiter unten nochmals diskutiert). Allerdings ist bei der Interpretation des errechneten Pillai-Wertes für alle Füllpartikeln in Abbildung 6.7 Zurückhaltung geboten, da sich die einzelnen Versuchspersonen sehr unterschiedlich verhalten, was in einer MANOVA nicht modelliert werden kann. Die Effektstärke η^2 ist größer als

in GECCO-FP und liegt schon im Bereich eines großen Effektes.¹³² Abbildung 6.8 zeigt die räumliche Verteilung zwischen vokalischen und vokalischnasalen Füllpartikeln aller 24 Versuchspersonen in BeDiaCo_c zum 50%-Messpunkt, ihre Pillai-Werte enthält Tabelle 6.12.

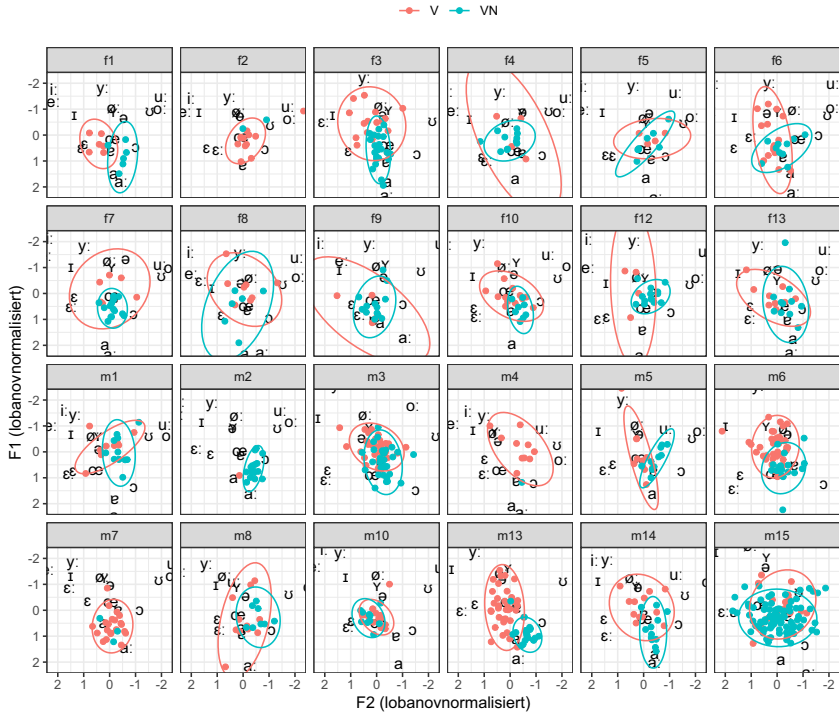


Abbildung 6.8: Vokalische (rot) und vokalischnasale (blau) Füllpartikeln in BeDiaCo_c je Sprecherin in lobanovnormalisierten Formantträumen (mittige Messung im Vokal) mit Zoom in die Vokalraummitte. Datenellipsen umspannen 95% der Messpunkte.

Höhere Pillai-Werte beschreiben eine größere Distanz zwischen zwei Vokalverteilungen. Somit unterscheidet Versuchsperson f1 die beiden Formen am deutlichsten. Die tiefere Realisierung des ersten Formanten in VN-Formen in Abbildung 6.8 und

¹³²Zur Orientierung kann man von einem kleinen Effekt bei $\eta^2 = 0,01$, von einem mittleren Effekt bei $\eta^2 = 0,06$ und von einem großen Effekt bei $\eta^2 = 0,14$ sprechen (Cohen 1988: 188 f.).

Tabelle 6.12: Freiheitsgrade (Df), Pillai-, F- und p-Werte für die Vergleiche der Vokaldistribution in V- und VN-Formen je Versuchsperson (VP) in BeDiaCo_c mittels einer multivariaten Varianzanalyse mit F₁, F₂ und F₃ als abhängiger Variable, sortiert nach Pillai-Werten.

VP	Df	Pillai	F	p	VP	Df	Pillai	F	p
f1	10	0,8	16,6	<0,01	m13	54	0,7	52,9	<0,001
f12	16	0,5	7,6	<0,01	m5	19	0,5	10,9	<0,01
f7	18	0,5	7,7	<0,01	m14	22	0,5	9,0	<0,01
f3	37	0,5	15,4	<0,001	m6	52	0,4	19,7	<0,001
f10	19	0,3	4,8	<0,05	m2	15	0,4	4,5	<0,05
f8	12	0,3	2,4	0,138	m4	9	0,4	2,5	0,142
f13	18	0,3	3,0	0,078	m8	17	0,3	4,3	<0,05
f2	12	0,3	1,9	0,203	m1	19	0,3	4,1	<0,05
f6	23	0,2	2,9	0,078	m3	60	0,3	11,3	<0,001
f9	15	0,2	1,3	0,297	m10	23	0,2	2,9	0,077
f4	13	0,1	0,8	0,483	m15	108	0,1	4,9	<0,01
f5	10	0,1	0,3	0,748	m7	22	0,0	0,3	0,766

auch der Gesamteindruck in Abbildung 6.7 wirkt die Frage auf, ob F₁ möglicherweise durch den nachfolgenden Nasal in VN-Formen abgesenkt wird. Jedoch gibt es weder im explorativen Korpus GECO-FP noch in der Literatur (Styler 2017: 2476) Evidenz für eine koartikulatorische Absenkung zum Zeitpunkt der Messung in der Mitte des Vokals, die nur die Vokale vor Nasalen in VN-Formen betrifft. Abbildung 6.9 zeigt außerdem, dass F₁ sowohl für V-Formen (rot) als auch für VN-Formen (blau) ab etwa 75 % (gestrichelte vertikale Linie) einen Knick nach unten aufweist, also erst nach der Messung. Fast alle Versuchspersonen zeigen fallende Formanten für VN-Formen, was eine koartikulatorische Folge des nachfolgenden bilabialen Nasals sein könnte, für den der Mund geschlossen wird.

Der dritte Formant F₃ ist ein Maß für die Lippenrundung; je tiefer F₃, desto stärker ist die Lippenrundung (Stevens 1997: 474). Abbildung 6.10 gibt die F₃-Werte der Füllpartikeln wieder und vergleicht sie mit ausgewählten Referenzvokalen aller Versuchspersonen in BeDiaCo_c. Die Auswahl erfolgte anhand der Überschneidung der Füllpartikel- mit den Referenzvokalen in Abbildung 6.7. F₃ zeigt wie in GECO-FP besondere Ähnlichkeit mit [ɐ] und [a:], wobei außerhalb der ersten Standardabweichung auch gerundete Werte (Ähnlichkeit mit [œ]) und ungerundete Werte (Ähnlichkeit mit [ə] und [ɛ]) auftreten.

Die Vokaldauer für vokalische Füllpartikeln ist signifikant länger als für vokalischnasale Füllpartikeln (Ausgleichskürzung, vgl. Abschnitt 6.1.3). Es ist daher möglich,

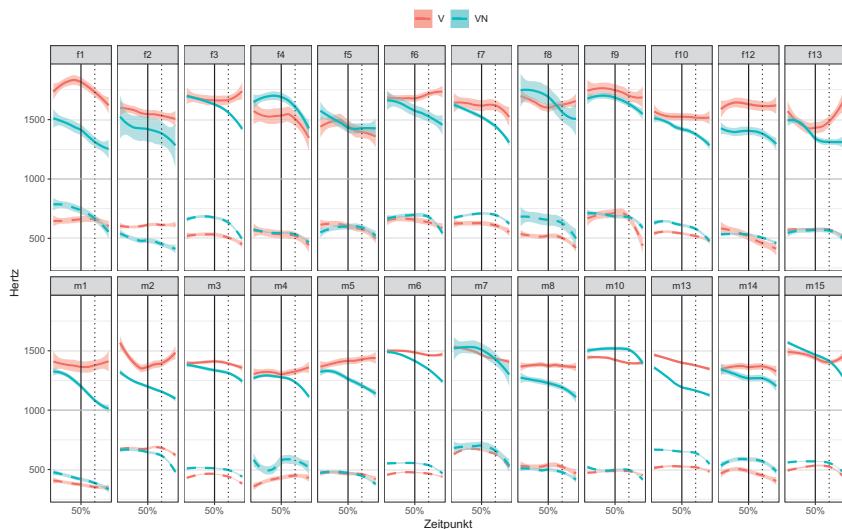


Abbildung 6.9: Gemittelte Formanttrajektorien von F_1 (unter 1000 Hz, gestrichelt) und F_2 (über 1000 Hz) für den Vokal in V- (rot) und VN-Füllpartikeln (blau) je Versuchsperson in BeDiaCo v.1 mit Zeitpunkten für mittige Messung der Formanten zum 50 %-Zeitpunkt (vertikale durchgezogene Linie). Die vertikale gepunktete Linie markiert den 75 %-Zeitpunkt und häufig den F_1 -Knick.

dass sich die Vokaldauer auch auf die Formantwerte auswirkt. Abbildung 6.11 inspiert nun den Zusammenhang zwischen Vokaldauer und den ersten beiden Formanten.

Für F_1 in Abbildung 6.11a visualisieren zwei getrennte lineare Regressionslinien die verschiedenen Korrelationen zwischen dem ersten Formanten und Vokallängen unter und über 100 ms. Für F_1 steigt die Regressionslinie unter 100 ms stark an, ist allerdings wie in GECO-FP für vokalische Füllpartikelvokale nicht signifikant (gerichtete Pearson's product-moment correlation, $r = 0,23, t = 1,39, Df = 36, p = 0,1$), für Vokale in vokalischnasalen Füllpartikeln hingegen durchaus (gerichtete Spearman's rank correlation rho, $\rho = 0,35, S = 28\,582, p < 0,01$).¹³³ Über 100 ms ist die Korrelation zwischen F_1 und Vokaldauer zwar klein, aber dennoch

¹³³ Aufgrund der nicht normalverteilten vokalischnasalen Vokaldauern unter 100 ms (Shapiro-Wilk normality test, $W = 0,89, p < 0,001$) wird eine Spearman-Korrelation verwendet.

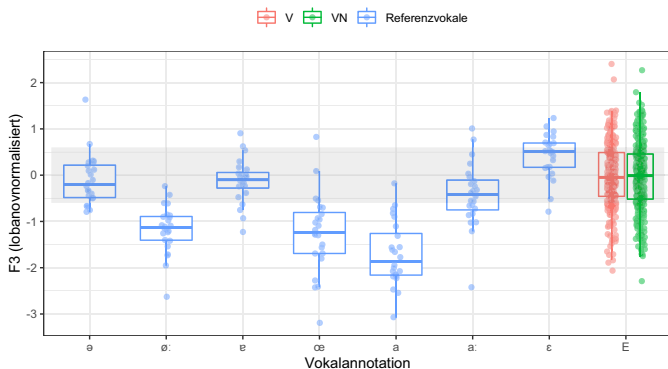


Abbildung 6.10: F₃-Werte der Vokale in vokalischen (rot) und vokalischnasalen (grün) Füllpartikeln in BeDiaCo_c für ausgewählte Referenzvokale (mittige Messung im Vokal). Der graumarkierte Bereich kennzeichnet eine Standardabweichung um den F₃-Mittelwert der Füllpartikelvokale.

signifikant (für Vokale in V-Formen: gerichtete Spearman's rank correlation rho, $\rho = 0,17$, $S = 2810901$, $p < 0,01$; für Vokale in VN-Formen: gerichtete Spearman's rank correlation rho, $\rho = 0,24$, $S = 3298828$, $p < 0,001$).¹³⁴ Möglicherweise erreichen die F₁-Werte in VN-Formen unter 100ms wie in GECO-FP aufgrund ihrer Kürze nicht ihr eigentliches Ziel, was einen sogenannten *target undershoot* darstellt. Dass die Korrelation über 100ms signifikant ist, könnte eine Folge des größeren Datensamples sein, da dies in GECO-FP nicht gefunden wird.

Für F₂ ergeben sich wie in GECO-FP keine visuellen Anzeichen für eine Zweiteilung der Regressionsgeraden. Für F₂ vokalischer Füllpartikeln ergibt sich eine negative Korrelation zwischen der Vokaldauer und der F₂-Höhe, d. h. je länger der Vokal, desto niedriger wird F₂ (gerichtete Spearman's rank correlation rho, $\rho = -0,24$, $S = 6233229$, $p < 0,001$). Für F₂ und Vokaldauer vokalischnasaler Füllpartikeln besteht nur ein kleiner negativer Zusammenhang (ungerichtete Spearman's rank correlation rho, $\rho = -0,14$, $S = 8825941$, $p < 0,05$).

Nachdem sich also Füllpartikelvokallängen unter 100ms noch stark auf die Po-

¹³⁴Aufgrund der nicht normalverteilten V-Vokaldauern über 100ms (Shapiro-Wilk normality test, $W = 0,88$, $p < 0,001$) sowie der nicht normalverteilten VN-Vokaldauern über 100ms (Shapiro-Wilk normality test, $W = 0,98$, $p < 0,01$) wird in beiden Fällen eine Spearman-Korrelation verwendet.

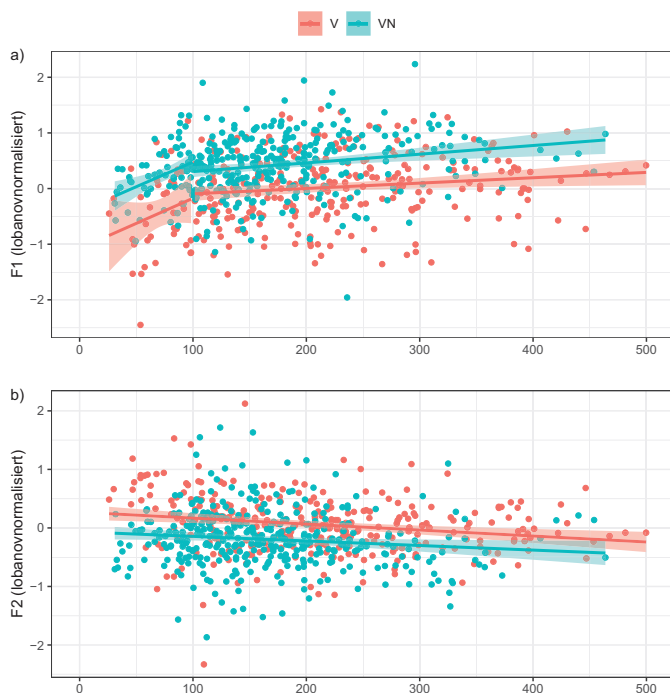


Abbildung 6.11: Zusammenhang zwischen F_1 in (a) beziehungsweise F_2 in (b) und Vokaldauer für vokalische (rot) und vokalischnasale (blau) Füllpartikeln. Lineare Regressionen in a) sind separat berechnet für Werte unter und über 100 ms.

sition des Formanten auswirken und möglicherweise ein *target undershoot* für VN-Formen vorliegt, liegt es nahe, den Formantraum über alle Füllpartikeln nochmal zweigeteilt darzustellen – jeweils für Vokale unter und über 100 ms Länge. Abbildung 6.12a veranschaulicht, dass sich die Variabilität für Dauern unter 100 ms besonders für V-Füllpartikeln stark erhöht, mit einem erhöhten Pillai-Wert als Indikator für die abweichenden Distributionen im Vokalraum und einer großen Effektstärke (Pillai = 0,38, $df = 100$, $F = 20,1$, $p < 0,001$, $\eta^2 = 0,38$). Dies mag daran liegen, dass ihr direkter segmentaler Kontext variabler als der von VN-Formen ist, die im rechten Segment immer den Nasal enthalten und somit ein anschließendes akusti-

sches Ziel vorgeben. Für die Vokaldauern über 100 ms in Abbildung 6.12 b ändert sich kaum etwas im Vergleich zur Gesamtschau in Abbildung 5.10 – sowohl der Pillai-Wert als auch die Effektstärke η^2 bleiben auf dem gleichen Niveau (Pillai = 0,14, $df = 567$, $F = 30$, $3, p < 0,001$, $\eta^2 = 0,14$). Dies repliziert die Ergebnisse aus GECO-FP.

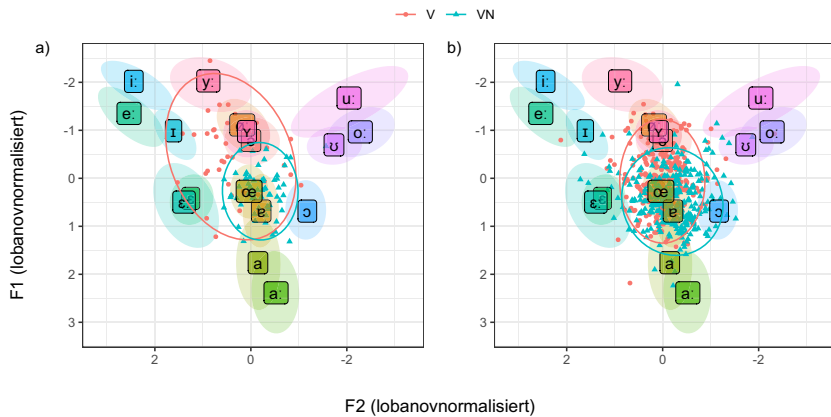


Abbildung 6.12: Lobanovnormalisierter Vokalraum aller Versuchspersonen in BeDiaCo_c mit Datenellipsen für die Distributionen der Referenzvokale (Ellipsen enthalten zur besseren Sichtbarkeit 80% aller Referenzvokalmittelwerte) für Vokaldauern unter 100 ms (a) und über 100 ms (b). V-Formen sind mit roten Punkten, VN-Formen mit blauen Dreiecken gekennzeichnet (mittige Messung im Vokal). Die entsprechenden Datenellipsen kennzeichnen 95% aller Messwerte.

Zur Quantifizierung der Überlappung zwischen V- und VN-Formantwerten wird für jeden Formanten ein lineares gemischtes Modell in Abhängigkeit von der Formausprägung der Füllpartikeln und der logarithmisierten Dauer gerechnet (Interaktionen verbessern das Modell nicht). Versuchspersonen gehen als Random Intercept, Vokaldauer der Füllpartikel als Random Slope in das Modell ein, ohne Korrelation zwischen dem Random Slope und dem Random Intercept. Das Modell ist wie folgt in R spezifiziert:

```
1 | lmer(F ~ Form + log(Dauer) + (1|VP) + (0+log(Dauer)|vp), data = ...
```

Abbildung 6.13 zeigt die jeweiligen Effekte für jeden Formanten. Nach der Analyse von GECO-FP war zu erwarten, dass F_1 für VN-Formen signifikant niedriger liegt,

Tabelle 6.13: Linear gemischte Modelle für F_1 , F_2 und F_3 in Abhängigkeit von *form* + *log(Dauer)* + $(1/vp)$ + $(0+\log(Dauer)/vp)$, mit erklärter Varianz R^2 .

	F1	F2	F3
(Intercept)	343,1 (26,1)***	1844,8 (57,7)***	2502,1 (61,0)***
FormVN	62,1 (5,2)***	-111,3 (11,6)***	2,6 (12,2)
log(Dauer)	38,1 (6,0)***	-61,8 (12,5)***	16,4 (14,4)
AIC	7266,7	8517,5	8586,0
Num. obs.	653	669	668
Num. groups: vp	24	24	24
R_m^2/R_c^2	0,1/0,69	0,08/0,57	0/0,69

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

während F_2 und F_3 keine Auswirkungen zeigen. Tabelle 6.13 zeigt die Ergebnisse der abhängigen Formanten F_1 , F_2 und F_3 ; sowohl der erste Formant als auch der zweite Formant zeigen signifikante Unterschiede zwischen V- und VN-Formen und ihrer Dauer: In VN-Formen ist F_1 ca. 62,1 Hz höher als in V-Formen (d. h. er liegt tiefer im Vokalraum) und steigt für längere Vokaldauern weiter an (bzw. bewegt sich im Vokalraum nach unten) – F_2 ist ca. -111,3 Hz niedriger für VN-Formen (d. h. er liegt posteriorer im Vokalraum) und bewegt sich weiter nach hinten, je länger er andauert. Für F_1 werden die Daten aus GECO-FP bestätigt; F_2 verhält sich jedoch anders als in den explorativen Daten. Sowohl in GECO-FP als auch in BeDiaCo_c wird F_3 nicht von der ihn beinhaltenden Füllpartikelform oder von der Dauer des Vokals beeinflusst.

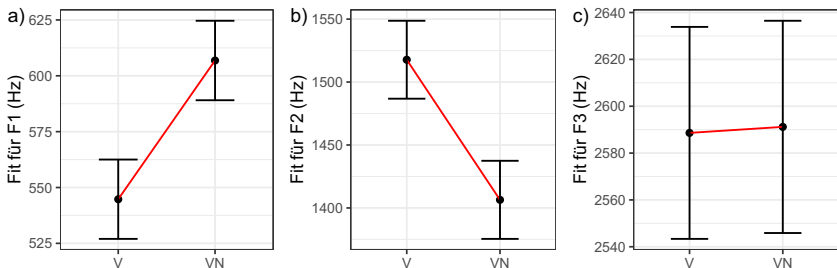


Abbildung 6.13: Effekte für a) F_1 , b) F_2 und c) F_3 des mittig gemessenen Vokals in V- und VN-Formen mit Form als Prädiktor.

6.1.6 Phonationsart

Abbildung 6.14 stellt die Verteilung der Vokale in V- und VN-Formen in BeDiaCo_c nach Phonationsart dar. Nasale wurden nicht annotiert. Aufgrund der wenigen laryngalisierten Formen im explorativen Korpus GECCO-FP wurde in BeDiaCo_c auf die Annotation von Laryngalisierung verzichtet. Stattdessen wurden behauchte Formen hinzugenommen, auch diese sind jedoch sehr niederfrequent. Insgesamt sind 13 % ($n = 86$) aller V- und VN-Füllpartikeln glottalisiert. Von den V-Formen sind es 18,3 %, von den VN-Formen 8,6 %. Dies entspricht den Werten in GECCO-FP.

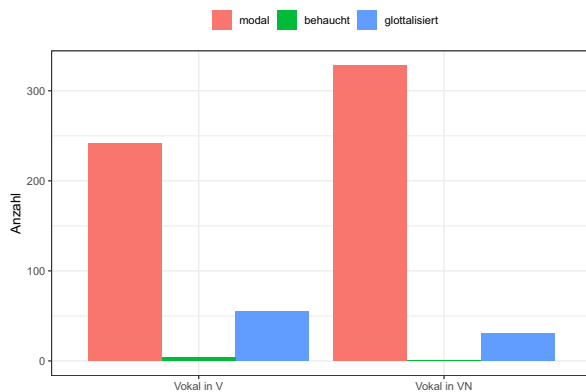


Abbildung 6.14: Anzahl der Vokale in V- und VN-Formen nach Phonationsart in BeDiaCo_c.

Die Vokaldauern je Füllpartikelform sind in Abbildung 6.15 dargestellt, mit den genauen Werten für normalisierte und nichtnormalisierte Dauer in Tabelle 6.14. Modale Vokale in V-Formen sind nach paarweisen Vergleichen mithilfe von t-Tests mit gepoolter Standardabweichung und bonferronikorrektiertem p-Wert signifikant länger als glottalisierte (ungerichteter Welch Two Sample t-test, $t = 3,1$, $df = 77,8$, $p < 0,01$, $KI_{95\%} = [0,12; 0,54]$), jedoch nicht als behauchte Vokale ($p = 1$). Modale Vokale in VN-Formen sind jedoch nicht signifikant länger als glottalisierte Vokale (ungerichteter Welch Two Sample t-test, $t = 0,5$, $df = 33,8$, $p = 0,6$, $KI_{95\%} = [-0,15; 0,26]$).

Das Histogramm in Abbildung 6.16 zeigt die Häufigkeit der Grundfrequenz über

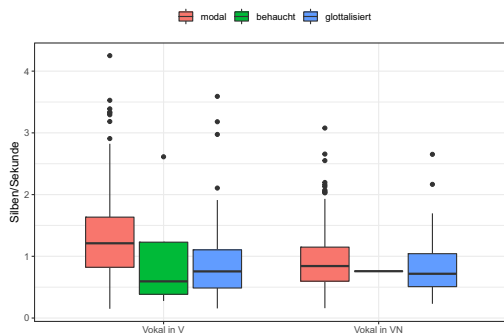


Abbildung 6.15: Normalisierte Dauer der Vokale in V- und VN-Formen nach Phonationsart in BeDiaCo_c.

Tabelle 6.14: Mittelwert und Standardabweichung für nichtnormalisierte Dauer (in ms) und normalisierte Dauer (in Silben σ) für Vokale je Füllpartikelform in BeDiaCo_c.

Form	Phonationsart	\bar{x} (ms)	s	\bar{x} (σ)	s
Vokal in V	modal	241,7	126,1	1,3	0,7
Vokal in V	behaucht	200,9	241,5	1,0	1,1
Vokal in V	glottalisiert	176,0	126,0	1,0	0,7
Vokal in VN	modal	173,6	85,0	0,9	0,5
Vokal in VN	behaucht	147,0		0,8	
Vokal in VN	glottalisiert	160,5	112,4	0,9	0,6

die Anzahl der modalen und glottalisierten Füllpartikeln. In 61 Fällen misst der Algorithmus 0 Hz. Die restlichen Daten zerfallen in zwei Gruppen (vgl. Abschnitt 5.1.4 und für BeDiaCo_c Abbildung 6.6), eine erste zwischen 0 und 140 Hz (der Mittelwert dieser Häufung beträgt 104,3 Hz) und eine zweite ab 140 Hz (der Mittelwert beträgt 198,6 Hz). In Abschnitt 5.1.4 wurde gezeigt, dass die Daten in der ersten Gruppe – wohlgemerkt nur für Frauen – auf glottalisierte Formen zurückzuführen sind. Die Frauen in BeDiaCo_c replizieren dies in Teilen, sie produzieren aber auch glottalisierte Füllpartikeln mit höherer f_0 . Für die Männer kann eine solche Gruppierung nicht durchgeführt werden, da ihre Grundfrequenz zu niedrig ist, so dass sich modale und glottalisierte Vokale in ähnlichen Frequenzbereichen befinden. Somit lassen sich für BeDiaCo_c die Ergebnisse, dass glottalisierte Formen hauptsächlich mit einer Grundfrequenz von weniger als 140 Hz produziert werden, nicht replizieren.

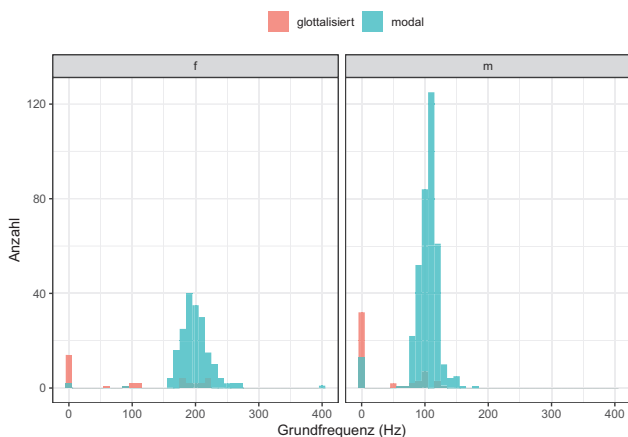


Abbildung 6.16: Phonationsarthistogramm der im Mittelpunkt aller nicht-glottalen Füllpartikeln gemessenen Grundfrequenz mit einer Klassenbreite von 10 Hz und den Mittelwerten der beiden Häufungen (gestrichelte Linien) über alle Formen je Häufung in BeDiaCoC.

6.2 Form und Kontext

6.2.1 Sequenzielle Kontexte

Sequenzielle Kontexte können auf Wort- oder auf Segmentebene analysiert werden. Zur Betrachtung und dem Vergleich der Wort-Pause-Kontexte rekategorisiere ich die sequenziellen Kontexte der Füllpartikelebene in Anlehnung an Leeuw (2007) in Wort- und Pausenkontexte (unabhängig von ihrer Art). Tabelle 6.15 enthält die Distribution der nicht-glottalen Füllpartikelformen in Wort-Pause-Kontexten für V-, VN- und N-Formen (orthographisch repräsentiert als *äh*, *ähm* und *hm*) im Vergleich mit GECO-FP.¹³⁵

Die für die Replikationsstudie geltende Nullhypothese geht davon aus, dass die Daten in beiden Korpora gleich verteilt sind. Für drei Kategorien ist dies auch zutreffend, allerdings nicht für Wort-FP-Wort-Kontexte. Während V- und VN-Formen in diesem Kontext in GECO-FP annähernd gleich verteilt sind (54 % vs. 46 %),

¹³⁵Diese Einteilung enthält die auf der Füllpartikelebene annotierten Kontexte (vgl. im Anhang Tabelle B.2) und ordnet sie den Kategorien ‚Wort‘ (*as/ps*) oder ‚Pause‘ (*ap/pp*, *ar/pr*, *ah/ph*, *at/pt*) zu (ohne Clicks, extralinguistische Entitäten und ohne Zweifelsfälle).

Tabelle 6.15: Anzahl vokalischer, vokalisch-nasaler und nasaler Füllpartikeln in Wort-Pause-Kontexten in GECO-FP und BeDiaCo_c mit p-Wert und Teststatistik für ihre Unterschiede je Zeile (p für χ^2 -Test basierend auf 2000 Simulationen).

Kontext	GECO-FP			BeDiaCo _c			p	Test
	V	VN	N	V	VN	N		
Wort-FP-Wort	73	61	0	133	55	0	< 0,01	Fisher
Wort-FP-Pause	28	56	1	38	66	0	= 0,59	$\chi^2 = 1,4$
Pause-FP-Wort	40	31	4	98	100	5	= 0,29	$\chi^2 = 2,4$
Pause-FP-Pause	5	36	2	14	101	3	= 0,84	$\chi^2 = 0,5$
Σ	146	184	7	283	322	8		

überwiegen in BeDiaCo_c die V-Formen deutlich (71 % vs. 29 %). In Pause-FP-Pause-Kontexten überwiegen hingegen die VN-Formen, und zwar in beiden Korpora (84 % in GECO-FP, 86 % in BeDiaCo_c). In Wort-FP-Pause-Kontexten reduziert sich der Anteil von VN-Formen (66 % in GECO-FP, 63 % in BeDiaCo_c). In Pause-FP-Wort sinkt er weiter (41 % VN in GECO-FP, 49 % VN in BeDiaCo_c). Abgesehen von Wort-FP-Wort-Kontexten scheinen die Füllpartikelformen also relativ robust verteilt zu sein.

Die Annotation der füllpartikeladjazenten Kontexte ermöglicht einen detaillierten Einblick in die Verteilung der Formen je Kontext. Im Gegensatz zu GECO-FP sind in BeDiaCo_c jedoch kaum glottale Füllpartikeln enthalten; daher betrachte ich nur nicht-glottale Formen. Abbildung 6.17 visualisiert ihre Verteilung in den zehn häufigsten Kontexten.

Der häufigste Kontext *as FP ps* enthält vor und nach der Füllpartikel ein Lautsegment. Frauen produzieren hier 72 % V-Formen, Männer 70 %. Ansonsten ist die Verteilung der Formen zwischen Männern und Frauen recht ähnlich, bis auf *as FP ph* (7 % V-Formen bei Frauen, 42 % bei Männern) und *at FP ps* (47 % V-Formen bei Frauen, 70 % bei Männern).

Die Distributionen der V- und VN-Formen der ersten fünf Kontexte in GECO-FP unterscheiden sich (nach Bonferronikorrektur für multiples Testen) nicht signifikant von denen in BeDiaCo_c, weder für Frauen noch für Männer. Auch für die Kontexte *ap FP pp* und *ah FP pp* gibt es keine signifikanten Unterschiede. In den restlichen Kontexten nehmen die enthaltenen absoluten Instanzen stark ab, weswegen ich hier von einer Interpretation absehe.

Hängt die Wahl einer Formvariante in BeDiaCo_c von ihrem sequenziellen Laut-

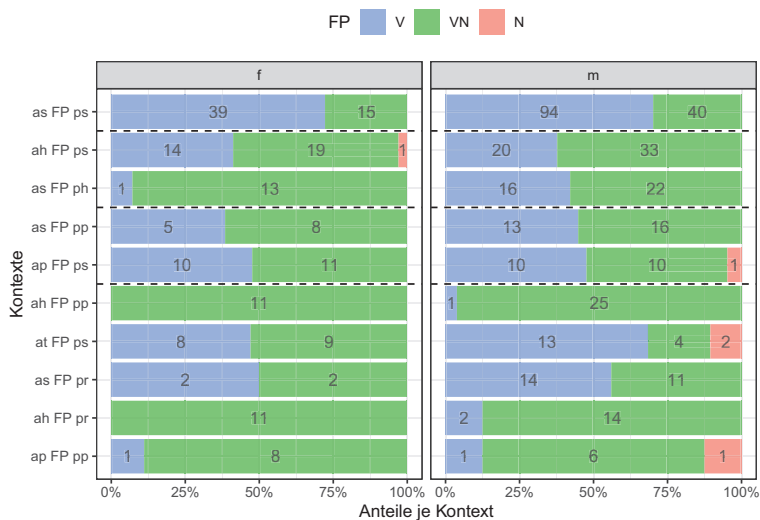


Abbildung 6.17: Anteil der V- und VN-Füllpartikelformen je Kontexte ($n > 10$ für mind. Frauen oder Männer) in BeDiaCo_c (FP = Füllpartikel, Antezedentia: at = Turn, as = Segment, ap = Pause, ah = Atmungspause; Postzedentia: ps = Segment, pp = Pause, ph = Atmungspause, pr = Pause mit phonetischen Reflexen). Gestrichelte Linien gliedern die Graphik in ähnliche Kontexte.

kontext ab? Abbildung 6.18 zeigt die Verteilung adjazenter Laute (in SAMPA) für den segmentumgebenden Kontext *as FP ps* für V- und VN-Formen. 15 Kontexte kommen sowohl bei V- als auch bei VN-Formen vor, einige Treffer häufiger als einmal. Um zu ermitteln, ob die beiden Tableaus unabhängig voneinander sind, werden 10 000 Permutationen des V-Tableaus simuliert und für jede Permutation die jeweils mit dem VN-Tableau übereinstimmenden Treffer gezählt. Anschließend ermittle ich mithilfe eines Binomialtests die Wahrscheinlichkeit, 15 oder mehr Übereinstimmungen in den 10 000 Vergleichen beider Tableaus zu erhalten. Da dies in 1 von 10 000 Fällen geglückt ist, ist $p < 0,001$ (der p-Wert entspricht hier der oberen Grenze des 95 %-Konfidenzintervalls). Die Hypothese, dass die beiden Tableaus unabhängig sind, muss somit abgelehnt werden. Dies bedeutet, dass es keine Evidenz dafür gibt, dass V-Formen einen anderen segmentalen Kontext haben als VN-Formen (sonst hätte es weniger gemeinsame Treffer geben müssen, nämlich maximal neun). Die

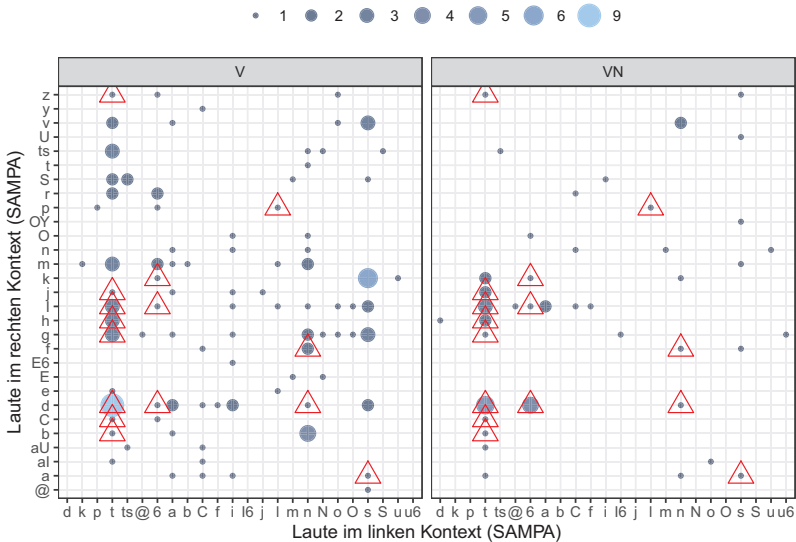


Abbildung 6.18: Anzahl und Verteilung der V- und VN-Füllpartikelformen je linkem und rechtem Kontext für alle Kontexte mit Segmenten ($n > 5$) in BeDiaCo_c (für SAMPA vgl. Tabelle D.1). Rote Dreiecke kennzeichnen identische Kontexte für V und VN.

Hypothese, dass die Wahl einer V- oder VN-Form von ihren direkt vorangehenden und nachfolgenden Lauten abhängt, muss demnach abgelehnt werden.

6.2.1.1 Dauer

In GECO-FP sind Füllpartikeln nach Segmenten und vor stillen Pausen und Atmungspausen (präpausal) sowie Füllpartikeln zwischen Atmungspausen und stillen Pausen (interpausal) signifikant länger als in intersegmentaler Position (vgl. Abschnitt 5.2.1.1). Die Hypothese lautet daher, dass auch in BeDiaCo_c Füllpartikeln in intersegmentaler Position kürzer als in linkssegmental-rechtspausaler und interpausaler Position sind. Zur Vergleichbarkeit mit den Kontextkategorien in GECO-FP wird die in BeDiaCo v.1 neu eingeführte Kategorie *ar/pr* (phonetische Reflexe in Pausen) zu Pausen (*ap/pp*) gezählt. Die Ergebnisse eines linearen gemischten Modells mit Dauer als abhängiger Variable werden in Abbildung 6.19 a–b und Tabel-

le 6.16ii dargestellt. Als unabhängige Variablen gehen wie in GECO-FP die sieben Kontexte Segment-FP-Segment, Turnpause-FP-Segment, Atmung-FP-Segment, Pause-FP-Segment, Segment-FP-Atmung, Segment-FP-Pause und Atmung-FP-Pause sowie die beiden Formen V und VN ein, als zufällige Effekte die Versuchspersonen. Random slopes und Interaktionen verbessern das Modell nicht; auch der Faktor Geschlecht erzielt keine Verbesserung.

$$1 \mid \text{lmer}(\text{FP-Dauer} \sim \text{Kontext} + \text{Form} + (1 \mid \text{VP}), \text{data} = \dots)$$

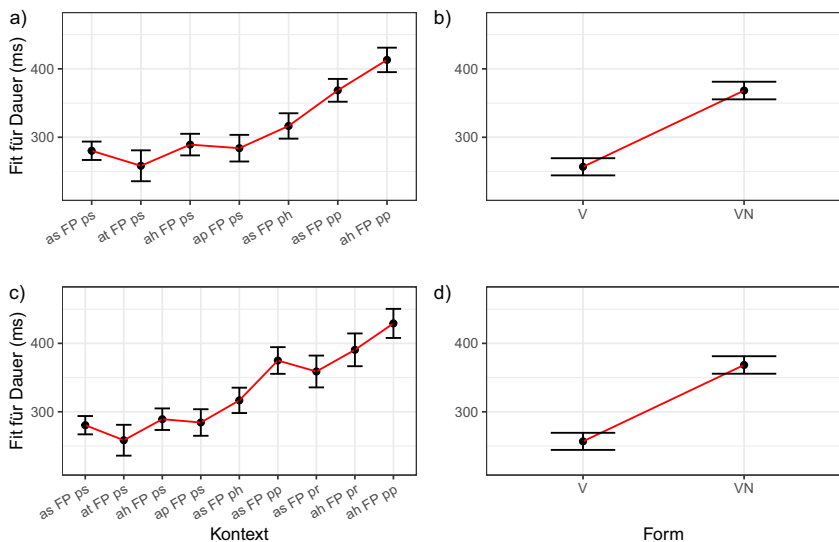


Abbildung 6.19: Modelleffekte für Füllpartikeldauern in adjazenten sequenziellen Kontexten der nicht-glottalen V- und VN-Füllpartikeln (FP) mit konfirmatorischer Analyse der Kontext- und Formeffekte in a) und b) und unter Einbeziehung neuer Kategorien in c) und d). Kontexte sind Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) sowie die neuen Kontexten *as FP pr* und *ah FP pr* (pr = Pause mit postzedentem phonetischen Reflex) in BeDiaCo_c.

Tatsächlich sind die Effekte mit denen von GECO-FP identisch. Füllpartikeln zwischen einem Segment und einer Pause sind mit ca. 88,3ms signifikant länger als intersegmentale Füllpartikeln; ebenso sind Füllpartikeln zwischen einem Segment im linken und Atmung im rechten Kontext mit ca. 36,3ms und zwischen Atmung

und Pause mit 132,8 ms signifikant länger als intersegmentale Füllpartikeln. Zudem sind, wie schon bekannt, VN-Füllpartikeln mit ca. 111,4 ms signifikant länger als V-Füllpartikeln.

Das zweite Modell bezieht die neueingeführten Kategorien *ar/pr* mit ein (vgl. Abbildung 6.19 c–d und Tabelle 6.16 i). Die Ergebnisse sind auch hier ähnlich mit denen in GECO-FP. Auch in den neuen Kontexte *as FP pr* und *ah FP pr* sind Füllpartikeln (wie schon für *as FP pp* und *ah FP pp*) signifikant länger als in intersegmentalen Kontexten. Die replizierte längere Dauer nach Segmenten und vor Pausen tritt an Positionen auf, die auch das Ende einer prosodischen Phrase definieren können. Dieser begründete Verdacht aus GECO-FP wird mit der Intonationsphrasenannotation untersucht. Die Ergebnisse im Hinblick auf diese Ebene werden in Abschnitt 6.2.3 diskutiert.

Tabelle 6.16: Linear gemischte Modelle für die Füllpartikeldauer in Abhängigkeit von *Kontext + Form + (1/VP)*, mit erklärter Varianz R^2 .

	i) Dauer (mit <i>pr</i>)	ii) Dauer (ohne <i>pr</i>)
(Intercept)	224,9 (13,4)***	224,9 (13,4)***
Kontextat FP ps	-22,0 (22,3)	-21,9 (22,3)
Kontextah FP ps	8,8 (14,9)	9,0 (14,9)
Kontextap FP ps	3,9 (19,0)	3,8 (19,0)
Kontextas FP ph	36,3 (17,7)*	36,3 (17,7)*
Kontextas FP pp	94,5 (18,4)***	88,3 (15,4)***
Kontextas FP pr	78,5 (22,5)***	
Kontextah FP pr	110,1 (24,0)***	
Kontextah FP pp	148,8 (21,2)***	132,8 (17,8)***
FormVN	111,6 (11,6)***	111,4 (11,6)***
AIC	6338,4	6353,5
Num. obs.	523	523
Num. groups: vp	24	24
R_m^2/R_c^2	0,35/0,46	0,35/0,46

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

In GECO-FP gibt es keine Evidenz dafür, dass die Länge der nachfolgenden Pause von der Länge oder der Form der Füllpartikel abhängt. Abbildung 6.20 zeigt die Länge der nachfolgenden Pause in Abhängigkeit von der Länge und Form der Füllpartikeln in BeDiaCoc. Die linearen Geraden zeigen für V-Formen einen schwachen negativen Zusammenhang ($\rho = 0,2$), für VN-Formen einen schwachen positiven Zusammenhang ($\rho = -0,1$), beide jedoch nicht signifikant. Die normalisierte mittlere

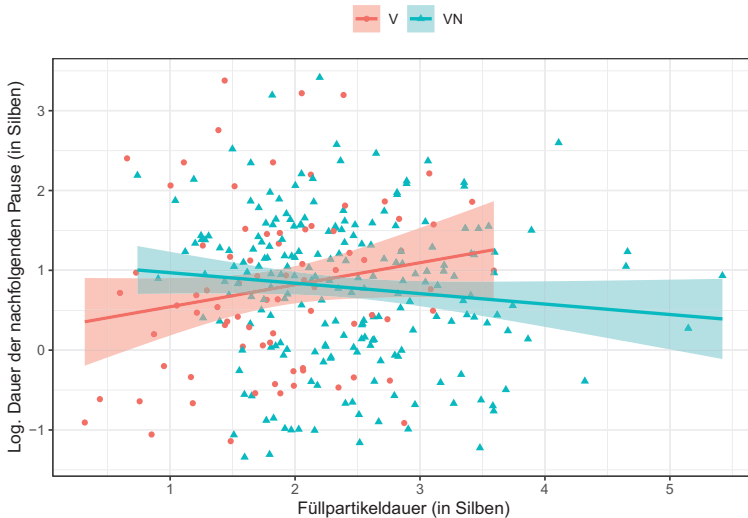


Abbildung 6.20: Füllpartikeldauer und Dauer der nachfolgenden Pause für V- und VN-Füllpartikelformen ($n_V = 80$, $n_{VN} = 210$) in BeDiaCo_c.

Dauer der nachfolgenden Pause nach V-Formen beträgt $\bar{x} = 3,88\sigma$ ($sd = 5,34$) und nach VN-Formen $\bar{x} = 3,29\sigma$ ($sd = 3,52$). Ein lineares gemischtes Modell mit der logarithmischen Füllpartikeldauer als abhängiger Variable, Form und Geschlecht als unabhängiger Variable und Versuchspersonen als Random Intercepts zeigt keine signifikanten Abhängigkeit der Pausendauer von der Form oder der Länge der Füllpartikel.

```
1 | lmer(log(Postposition) ~ Form + FP-Dauer + Sex + (1|VP), data = ...)
```

Es gibt allerdings einen (marginalen) Effekt für den Faktor Geschlecht, wonach die nachfolgende Pause bei Männern geschätzt 149,6ms länger ist als bei Frauen ($Std. Err. = 97,4$, $t = 2,1$, $p < 0,05$). Damit können weder in GECO-FP noch in BeDiaCo_c die Ergebnisse von Clark & Fox Tree (2002) repliziert werden. Auch das Maß der ‚totalen Verzögerung‘ in Anlehnung an Rose (2015) führt wieder zu keinem anderen Ergebnis. Es beträgt für die Dauer von V-Formen mit nachfolgender Pause $\bar{x} = 5,74\sigma/s$ ($sd = 5,42$), für VN-Formen mit nachfolgender Pause $\bar{x} = 5,71\sigma/s$ ($sd = 3,53$). Der Unterschied ist nicht signifikant (ungerichteter Welch Two Sample

t-test, $t = 0,04$, $df = 105,57$, $p = 0,97$, $KI_{95\%} = [-1,27; 1,32]$).

Die Ergebnisse zeigen also keinen Zusammenhang zwischen der nachfolgenden Pausenlänge und der Füllpartikelform oder -länge. Der umgebende Kontext hat allerdings sehr wohl einen Einfluss auf die Länge der Füllpartikel. So sind Füllpartikeln in Kontexten mit vorangehenden Segmenten und nachfolgender Pause (Stille, Atmung, oder mit phonetischen Reflexen) sowie interpausal signifikant länger als intersegmental. Die Art dieser Kontexte lässt vermuten, dass es hier möglicherweise einen Zusammenhang zwischen der Prosodie und der Füllpartikellänge gibt. Diese Hypothese wird in Abschnitt 6.2.3 überprüft.

6.2.1.2 Grundfrequenz

Kann der kontextbedingte f_0 -Effekt aus GECO-FP repliziert werden? In GECO-FP (vgl. Abschnitt 5.2.1.2) werden f_0 in den Kontexten *at FP ps* und *ah FP ps* mit signifikant höherer f_0 geäußert. Für die Kontexte in BeDiaCo_c wird die Analyse nun zunächst mit den Kategorie von GECO-FP wiederholt, wobei die in BeDiaCo_c neu hinzugekommenen Kontexte *ar/pr* als *ap/pp* rekategorisiert werden.

Für jede Füllpartikel ($n_V = 306$, $n_{VN} = 361$) wird der Mittelwert gebildet. Werte unter 50 Hz für Männer und 140 Hz für Frauen (Messfehler und glottalisierte Phasen) und über 350 Hz sowie nichtnormalverteilte Residuen werden ausgeschlossen (insgesamt 3%). Die Abhängigkeit der f_0 -Mittelwerte wird in einem linearen gemischten Modell berechnet, mit Form, Position und Geschlecht als unabhängigen Faktoren sowie Random Intercepts für Versuchspersonen. Interaktionen und Random Slopes verbessern das Modell nicht. Abbildung 6.21 a–c visualisiert die Modellvorhersagen in Tabelle 6.17 ii im direkten Vergleich mit GECO-FP. Abbildung 6.21 d–f visualisiert die Modellvorhersagen in Tabelle 6.17 i mit der neuen Kontextkategorie *pr*.

```
1 | lmer(F0 ~ Form + Kontext + Sex + (1|VP), data = ...)
```

V- und VN-Füllpartikeln unterscheiden sich wie in GECO-FP nicht in ihrer Grundfrequenz; Frauen produzieren Füllpartikeln mit einer signifikant höheren f_0 als Männer. Ausgehend von Füllpartikeln in intersegmentaler Position (*as FP ps*) sind Füllpartikeln, denen eine Turnpause (die andere Person spricht) vorangeht und ein Segment nachfolgt (*at FP ps*) signifikant höher (7,9 Hz in Abbildung 6.21 a, 8,8 Hz in Abbildung 6.21 d). Selbiges gilt für Füllpartikeln, denen eine Atmungspause vorangeht und ein Segment nachfolgt (4,5 Hz höher in Abbildung 6.21 a, 5,1 Hz höher

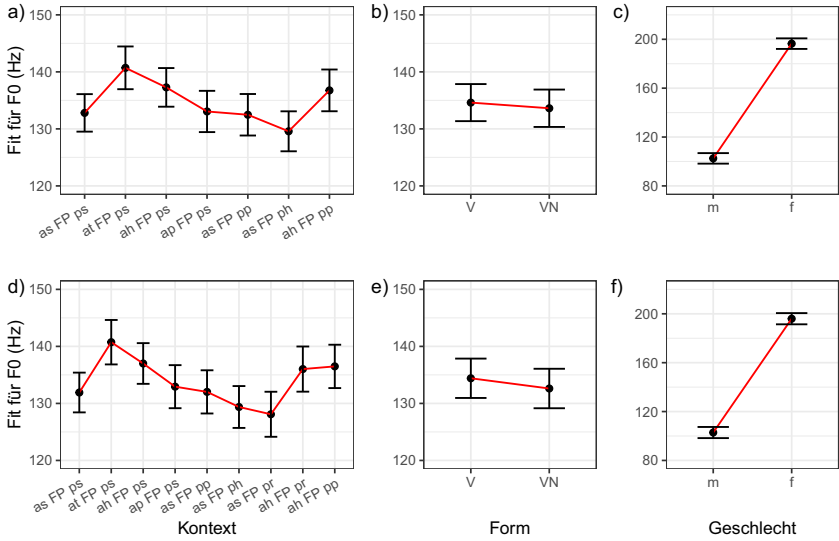


Abbildung 6.21: Modelleffekte für f_0 -Mittelwerte in adjazenten sequenziellen Kontexten der nicht-glottalen V- und VN-Füllpartikeln (FP) mit konfirmatorischer Analyse der Kontext-, Form- und Geschlechtereffekte in a)–c) und unter Einbeziehung neuer Kategorien in c)–e) (Geschlechtereffekte mit anderer Ordinatenskala). Kontexte sind Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) sowie die neuen Kontexten *as FP pr* und *ah FP pr* (*pr* = Pause mit postzedentem phonetischen Reflex) in BeDiaCo_c.

in Abbildung 6.21 d. Dies repliziert die Ergebnisse aus GECO-FP, wenngleich die Effektstärke hier geringer ist.

Post-hoc-Tests (vgl. Tabelle 6.18) zeigen zudem, dass Füllpartikeln im Kontext *at FP ps* mit einer um 8,7 Hz signifikant höheren f_0 produziert werden als nach Segmenten und vor stillen Pausen (*as FP pp*), 11,4 Hz höher als nach Segmenten und vor Atmungspausen (*as FP ph*) und 12,6 Hz höher als nach Segmenten und vor Pausen mit phonetischen Reflexen (*as FP pr*). Die ersten beiden Unterschiede replizieren die Ergebnisse aus GECO-FP, während der Kontext *as FP pr* neu hinzukommt. Der in GECO-FP zusätzlich vorhandene signifikante Unterschied zwischen *at FP ps* und *ah FP ps* beziehungsweise *ap FP ps* kann mit den Daten aus BeDiaCo_c nicht repliziert

Tabelle 6.17: Modellvorhersagen für $\bar{x}_{f_0} \sim \text{Form} + \text{Kontext} + \text{Sex} + (1|\text{VP})$, mit erklärter Varianz R^2 .

	i) f_0 (mit pr)	ii) f_0 (ohne pr)
(Intercept)	102,2 (4,6)***	101,7 (4,4)***
FormVN	-1,8 (1,2)	-1,0 (1,3)
Kontextat FP ps	8,8 (2,2)***	7,9 (2,3)***
Kontextah FP ps	5,1 (1,4)***	4,5 (1,5)**
Kontextap FP ps	1,0 (1,9)	0,2 (2,0)
Kontextas FP pp	0,1 (1,9)	-0,3 (2,0)
Kontextas FP ph	-2,5 (1,7)	-3,2 (1,8)
Kontextas FP pr	-3,8 (2,2)	
Kontextah FP pr	4,1 (2,3)	
Kontextah FP pp	4,6 (2,0)*	3,9 (2,1)
sexf	93,2 (6,5)***	93,9 (6,1)***
AIC	3831,1	3499,0
Num. obs.	505	455
Num. groups: vp	24	24
R_m^2/R_c^2	0,84/0,93	0,85/0,93

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

werden.

Tabelle 6.18: Signifikante Kontraste der Position für f_0 des Modells in Tabelle 6.17 i (ohne Referenzniveau *as FP ps*), mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95%-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
at FP ps - as FP pp	8,7	2,6	476,0	0,47	16,96	3,30	<0,05
at FP ps - as FP ph	11,4	2,5	476,1	3,66	19,07	4,59	<0,001
at FP ps - as FP pr	12,6	2,9	475,1	3,75	21,53	4,43	<0,001
ah FP ps - as FP ph	7,6	1,8	473,8	1,91	13,35	4,16	<0,01
ah FP ps - as FP pr	8,9	2,3	473,3	1,60	16,21	3,80	<0,01
as FP ph - ah FP pp	-7,1	2,3	473,9	-14,17	-0,07	-3,15	<0,05
as FP pr - ah FP pp	-8,4	2,7	472,9	-16,70	-0,09	-3,15	<0,05

Die Ergebnisse einer höheren f_0 aus GECO-FP für die Kontexte *at FP ps* und *ah FP ps* konnten also gegenüber der intersegmentalen Position *as FP ps* in BeDiaCo_c bestätigt werden. In Abschnitt 6.2.3 (vgl. Abbildung 6.33 a) wird gezeigt, dass diese beiden Kontexte tatsächlich (wie schon in Abschnitt 5.2.1.2 vermutet) überwiegend am Anfang prosodischer Phrasen stehen.

Werden Füllpartikeln in Fünfgrammen auch in BeDiaCo_c mit niedrigerer f_0 pro-

duziert als ihr Kontext? Für jedes Wort in dem Fünfgramm Wort-Wort-FP-Wort-Wort ($n_V = 5 \cdot 133$, $n_{VN} = 5 \cdot 55$) wird der Mittelwert gebildet. Werte unter 50 Hz für Männer und 140 Hz für Frauen (Messfehler und glottalisierte Phasen) und über 350 Hz sowie nichtnormalverteilte Residuen werden ausgeschlossen (insgesamt 24,1 %). Die Abhängigkeit der f_0 -Mittelwerte wird in einem linearen gemischten Modell berechnet, mit Form, Position und Geschlecht als unabhängige Faktoren sowie Random Intercepts für Versuchspersonen und einem Index, der für jedes Fünfgramm gleich ist. Interaktionen verbessern das Modell nicht. Abbildung 6.22 visualisiert die Effekte für die vorhergesagten Werte in Tabelle 6.19.

```

1 lmer(F0 ~ Form + Position + Sex + (1|VP) + (1|Index),
2 data = ...)

```

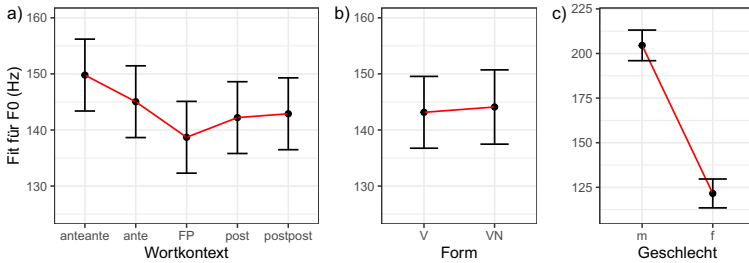


Abbildung 6.22: Modelleffekte der f_0 -Mittelwerte a) zwei Wörter (*anteante*) und ein Wort (*ante*) vor sowie ein Wort (*post*) und zwei Wörter (*postpost*) nach nicht-glottalen Füllpartikeln (FP), b) für Formen und c) für Geschlecht (andere Ordinalskala) in BeDiaCo_c.

Das Modell und die weiteren signifikanten Kontraste in Tabelle 6.20 zeigen einen signifikanten f_0 -Abfall von -4,8 Hz von *anteante* zu *ante* sowie von 6,3 Hz von *ante* zur Füllpartikel, wonach die Grundfrequenz zu den nächsten beiden Wörtern wieder zwischen 3,5 Hz (von FP zu *post*) und 4,2 Hz (von FP zu *postpost*) signifikant ansteigt, aber auf einem um -6,9 Hz niedrigeren Niveau als zum Zeitpunkt *anteante* verbleibt. Dies passt zu dem bekannten Konzept einer allgemein sinkenden f_0 -Deklinaton innerhalb einer prosodischen Phrase. Im Gegensatz zu GECO-FP zeigen die hier vorliegenden Daten keine Interaktion zwischen V- und VN-Formen. Somit wurde repliziert, dass Füllpartikeln in Fünfgrammen mit niedrigerer f_0 produziert werden als ihr Kontext.

Tabelle 6.19: Modellvorhersagen für $\bar{x}_{f_0} \sim \text{Form} + \text{Position} + \text{Form} * \text{Position} + (1|\text{VP}) + (1|\text{Id})$, mit erklärter Varianz R^2 .

	Mittlere f_0
(Intercept)	167,1 (10,9)***
FormVN	0,7 (2,4)
Positionante	-4,8 (1,0)***
PositionFP	-11,1 (1,0)***
Positionpost	-7,6 (1,0)***
Positionpostpost	-6,9 (1,0)***
AIC	5410,2
Num. obs.	713
Num. groups: id	171
Num. groups: vp	21
R_m^2/R_c^2	0/0,98

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Tabelle 6.20: Signifikante Kontraste der Position für f_0 ohne jene in Tabelle 6.19, mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
ante - FP	6,3	0,9	540,9	3,93	8,76	7,20	<0,001
ante - post	2,8	0,9	543,5	0,38	5,27	3,16	<0,05
FP - post	-3,5	0,9	543,0	-5,97	-1,07	-3,93	<0,001
FP - postpost	-4,2	0,9	543,9	-6,74	-1,66	-4,52	<0,001

6.2.1.3 Formanten

Beeinflusst der Kontext die Vokalqualität einer Füllpartikel? In GECO-FP ist dies für einige Kontexte der Fall (vgl. Abschnitt 5.2.1.3). Die Vokalqualität wird in der Mitte des Füllpartikelvokals in der Ebene \mathbf{fp} in BeDiaCo v. 1 gemessen. Als erklärenden Faktor verwende ich wieder die sieben häufigsten Kontexte (vgl. Abschnitt 6.2.1) mit 535 Instanzen ($n_V = 262, n_{VN} = 273$). Für die Kontexte in BeDiaCo_c wird die Analyse zunächst mit denselben Kategorien wie in GECO-FP wiederholt, wobei die in BeDiaCo_c neu hinzugekommene Kontext pr als pp rekategorisiert wird. Für F_1 und F_2 werden zwei separate gemischte lineare Modelle gerechnet, mit Kontext, Form und der logarithmisierten Dauer als Prädiktoren sowie Versuchspersonen als Random Intercepts und die unkorrelierte logarithmische Länge des Vokals als Random Slope je Versuchsperson. Interaktionen verbessern das Modell nicht. 2,1 % der

F₁- und 0,9% der F₂-Daten mussten aufgrund ihrer nichtnormalverteilten Residuen ausgeschlossen werden. Abbildung 6.23 visualisiert die Modellvorhersagen für F₁ und F₂ in Tabelle 6.21 iii-iv.

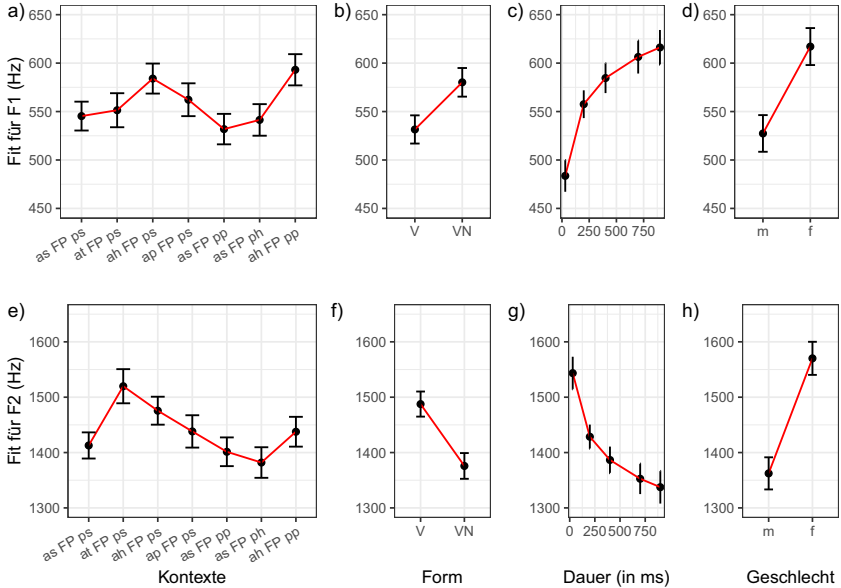


Abbildung 6.23: Modelleffekte der F₁- und F₂-Werte für a)/e) sequenzielle Kontexte Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) und Atmung-FP-Pause-Kontexte (*ah FP pp*), b)/f) Form, c)/g) Dauer und d)/h) Geschlecht in BeDiaCo_c.

Ausgehend von Füllpartikeln in intersegmentaler Position (*as FP ps*) liegen die Füllpartikelvokale, denen eine Turnpause (die andere Person spricht) vorangeht und ein Segment nachfolgt (*at FP ps*) anders als in GECO-FP nicht signifikant tiefer, wohl aber signifikant anteriorer (107 Hz) im Vokalraum. Füllpartikelvokale, denen eine Atmungspause vorangeht und ein Segment nachfolgt, sind jedoch wie in GECO-FP signifikant tiefer (38,7 Hz) im Vokalraum und in BeDiaCo_c auch signifikant anteriorer (62,8 Hz) als intersegmentale Füllpartikelvokale. Füllpartikelvokale zwischen Segment und Atmung (*as FP ph*) sind in BeDiaCo_c nicht signifikant ante-

riorer als intersegmentale Füllpartikelvokale, wie es in GECCO-FP der Fall ist. Anders als in GECCO-FP führt hier auch die interpausale Stellung einer Füllpartikel zwischen Atmung und stiller Pause (*ah FP pp*) zu einer signifikant tieferen Realisierung (47,8 Hz) als in intersegmentaler Stellung.

Vokale in V-Formen werden wie in GECCO-FP signifikant höher und anteriorer realisiert als VN-Formen. Die Dauer hat einen signifikant positiven Einfluss auf F_1 (je länger, desto tiefer) und einen signifikant negativen Einfluss auf F_2 (je länger, desto posteriorer) und repliziert damit auch die Analyse aus GECCO-FP. Frauen haben signifikant höhere Formantwerte als Männer, was aufgrund ihres kürzeren Ansatzrohres erwartet wird.

Tabelle 6.21: Linear gemischte Modelle für F_1 und F_2 in Abhängigkeit von *Kontext + Form + log(Dauer) + Sex + (1/VP) + (0+log(Dauer)/VP)* und für F_3 in Abhängigkeit von *Kontext + Form + Sex + (1/VP) + (0+log(Dauer)/VP)*, mit erklärter Varianz R^2 .

	i) F_1 (mit <i>pr</i>)	ii) F_2 (mit <i>pr</i>)	iii) F_1	iv) F_2	v) F_3 (mit <i>pr</i>)
(Intercept)	366,6*** (34,3)	1930,4*** (66,6)	375,6*** (34,1)	1929,5*** (66,7)	2701,2*** (38,3)
Kontextat FP ps	5,2 (11,8)	107,7*** (24,9)	6,1 (11,7)	107,0*** (24,9)	86,2*** (29,3)
Kontextah FP ps	35,7*** (8,2)	63,0*** (17,1)	38,7*** (8,1)	62,8*** (17,1)	53,6** (19,9)
Kontextap FP ps	16,2 (10,8)	25,6 (22,7)	16,9 (10,7)	25,4 (22,6)	21,0 (26,4)
Kontextas FP pp	-11,3 (10,2)	-27,7 (21,6)	-13,5 (8,4)	-11,4 (17,9)	18,3 (25,2)
Kontextas FP ph	-3,8 (9,6)	-31,3 (20,4)	-4,0 (9,5)	-30,9 (20,4)	-22,4 (23,8)
Kontextas FP pr	-19,0 (12,3)	13,4 (25,7)			8,5 (30,1)
Kontextah FP pr	59,9*** (13,3)	28,2 (28,1)			59,3 (32,7)
Kontextah FP pp	35,8** (11,9)	23,4 (24,8)	47,8*** (10,0)	24,8 (21,0)	95,4*** (28,6)
FormVN	50,1*** (6,7)	-111,6*** (14,0)	48,7*** (6,6)	-111,8*** (14,0)	-21,3 (15,9)
log(Dauer)	40,4*** (5,9)	-60,7*** (11,8)	39,0*** (5,8)	-60,6*** (11,8)	
sexm	-86,6** (26,6)	-209,2*** (41,4)	-89,7*** (26,7)	-207,8*** (41,6)	-287,6*** (51,5)
AIC	5757,4	6580,8	5750,0	6596,1	6820,9
Num. obs.	525	530	524	530	535
Num. groups: vp	24	24	24	24	24
R_m^2/R_c^2	0,29/0,68	0,4/0,63	0,3/0,69	0,4/0,63	0,35/0,65

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Tabelle 6.22 enthält die *Post-hoc*-Tests der Kontexte für F_1 und F_1 . Obwohl nicht

alle Ergebnisse aus GECO-FP repliziert werden, wirken sich doch die Kontexte, die typisch für den Beginn prosodischer Phrasen sind (linkerhand Atmung, rechterhand Segment) oder sogar eigenständige Phrasen konstituieren (interpausal), signifikant auf die Vokalqualität von Füllpartikeln aus.

Tabelle 6.22: Signifikante Kontraste für multiple paarweise Vergleiche der positionalen Faktorausprägungen, mit Schätzer β und Standardabweichung (s) für F_1 und F_2 , Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiples Testen nach Tukey (Turnpause-FP-Segment = *at FP ps*, Atmung-FP-Segment = *ah FP ps*, Segment-FP-Atmung = *as FP ph*).

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
F_1							
at FP ps - ah FP pp	-41,7	13,6	499,0	-82,05	-1,44	-3,07	<0,05
ah FP ps - as FP pp	52,2	9,5	494,1	23,94	80,41	5,47	<0,001
ah FP ps - as FP ph	42,7	10,5	496,1	11,46	73,93	4,05	<0,01
as FP pp - ah FP pp	-61,2	10,6	493,8	-92,61	-29,88	-5,78	<0,001
as FP ph - ah FP pp	-51,8	11,5	495,8	-85,83	-17,69	-4,50	<0,001
F_2							
at FP ps - as FP pp	118,4	27,7	509,0	36,41	200,47	4,27	<0,001
at FP ps - as FP ph	137,9	29,4	510,0	50,95	224,79	4,70	<0,001
ah FP ps - as FP pp	74,2	20,2	502,2	14,41	134,03	3,67	<0,01
ah FP ps - as FP ph	93,6	22,5	504,3	27,12	160,17	4,17	<0,001

Abbildung 6.24 visualisiert die Kontexteffekte für F_1 und F_2 in Tabelle 6.21 i–ii mit dem zusätzlichen Kontext *pr* (postzedenter Pause mit phonetischen Reflexen). Die Werte ändern sich im Vergleich zu den Modellen ohne diesen Kontext nur marginal. Allerdings ist in dem neuen Modell der Kontext *ah FP pr* (ebenso wie *ah FP pp*) signifikant tiefer (59,9 Hz) im Vokalraum als Füllpartikelvokale in intersegmentaler Position. Der Kontext *ah FP pr* verhält sich also ungefähr wie *ah FP pp*, während der Kontext *as FP pr* sich wie *as FP pp* verhält (keine signifikanten Unterschiede in *Post-hoc*-Tests nach Tukey).

In GECO-FP haben die Füllpartikelkontexte keinen signifikanten Einfluss auf den dritten Formanten. Abbildung 6.25 visualisiert die Kontexteffekte für F_3 in BeDiaCo_c (vgl. Tabelle 6.21 v). F_3 liegt sowohl im Kontext zwischen Turnpause und Segment (*at FP ps*, 86,2 Hz) als auch im Kontext zwischen Atmung und Segment (*ah FP ps*, 53,6 Hz) und im Kontext zwischen Atmung und Pause (*ah FP pp*, 95,4 Hz) signifikant höher als intersegmental. In multiplen *Post-hoc*-Tests mit Korrektur nach

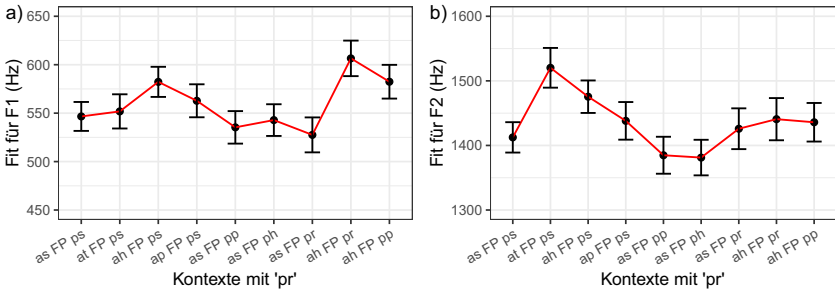


Abbildung 6.24: Modelleffekte der F₁- und F₂-Werte für a) sequenzielle Kontexte Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*) und Atmung-FP-Pause-Kontexte (*ah FP pp*) sowie zusätzlich mit b) Segment-FP-Reflexpause (*as FP pr*) und Atmung-FP-Reflexpause (*ah FP pr*) in BeDiaCo_c.

Tukey ist zudem der Kontext *as FP ph* signifikant niedriger als *at FP ps* und *ah FP pp*.¹³⁶ F₃ ist für Frauen signifikant höher als für Männer, während es sich für die Füllpartikelform nicht signifikant unterscheidet.

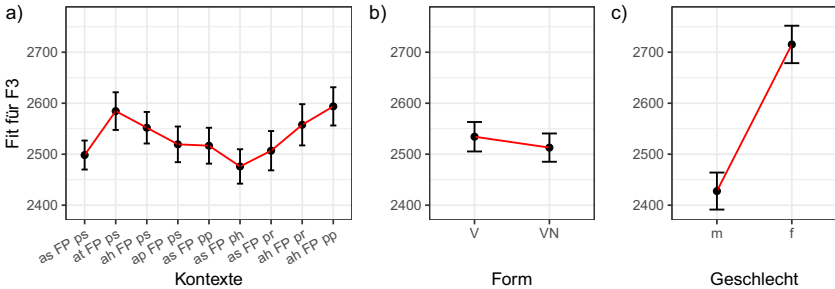


Abbildung 6.25: Modelleffekte der F₃-Werte a) in sequenziellen Kontexten Segment-FP-Segment- (*as FP ps*), Turnpause-FP-Segment- (*at FP ps*), Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*), Segment-FP-Reflexpause (*as FP pr*), Atmung-FP-Reflexpause (*ah FP pr*) und Atmung-FP-Pause-Kontexte (*ah FP pp*), b) für Form und c) für Geschlecht in BeDiaCo_c.

¹³⁶Post-hoc-Tests mit Korrektur nach Tukey: *at FP ps* - *as FP ph*, $\beta = 108,6, s = 34,3, df = 511,9$, Unteres KI = 1,7, Oberes KI = 215,5, $t = 3,2, p < 0,05$; *as FP ph* - *ah FP pp*, $\beta = -117,8, s = 32,3, df = 508,1$, Unteres KI = -218,5, Oberes KI = -17,2, $t = -3,6, p < 0,01$.

6.2.1.4 Phonationsart

Für die direkten sequenziellen Kontexte mit der absoluten Häufigkeit von $n > 5$ zeigt Abbildung 6.26 die absoluten und relativen Werte je Phonationsart und Geschlecht. Während glottalisierte Vokale in GECO-FP fast nur in Kontexten stehen, in denen im Ante- oder Postzedens ein Segment oder Click vorkommt, ist das Bild in BeDiaCo_c diverser. Für Frauen gibt es tatsächlich keine signifikante Abweichung von der erwarteten Verteilung (beidseitiger exakter Fisher-Test basierend auf 2000 Simulationen, $p = 0,6$), für Männer hingegen schon (beidseitiger exakter Fisher-Test basierend auf 2000 Simulationen, $p < 0,05$). Somit lassen sich für die Abhängigkeit der Phonationsart von sequenziellen Kontexten bisher keine konsistenten Schlüsse aus den beiden Studien ziehen.

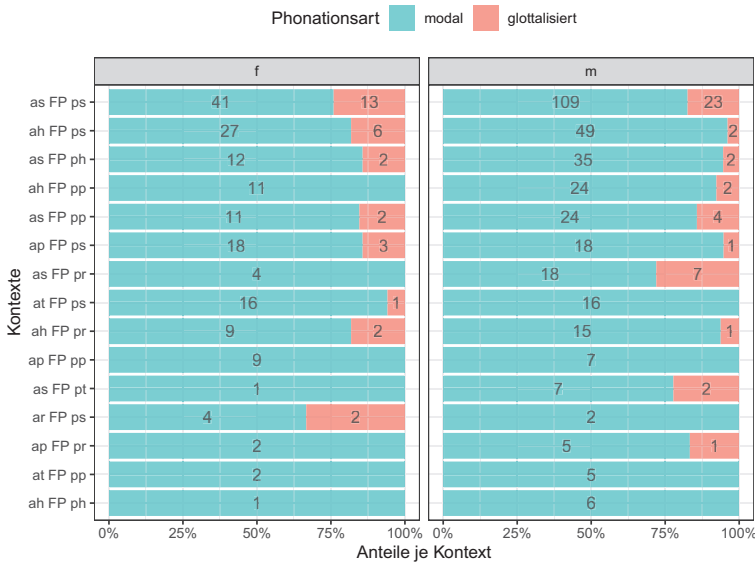


Abbildung 6.26: Absolute und relative Werte der Phonationsart des Vokals in sequenziellen Kontexten ($n > 5$) in BeDiaCo_c (FP = nicht-glottale Füllpartikel, Antezedentia = a, Postzedentia = p, zweite Stelle: t = Turnpause, s = Segment, p = Pause, h = Atmungspause, r = Pause mit artikulatorischem Reflex).

6.2.2 Dialogzüge

Tabelle 6.23 zeigt die Häufigkeiten der Annotationswerte je Dialogzugkategorie.¹³⁷ Die Anzahl der verschiedenen Dialogzüge und die Anzahl der Wörter je Dialogzug hängen stark von den Annotationsrichtlinien ab. Zur Vergleichbarkeit mit anderen Studien gebe ich daher die prozentuale Verteilung der Wörter über die Dialogzüge an. Die in den Dialogzügen enthaltenen Anteile glottaler (FG) und nicht-glottaler Füllpartikeln (FV) unterscheiden sich hier jedoch im Gegensatz zu GECO-FP nicht signifikant von den erwarteten Häufigkeiten (Fishers zweiseitiger exakter Text, $p = 0, 1$).

Tabelle 6.23: Annotationswerte (n), Wörter (W), Anteil der Wörter (W/n), Anteil der Wörter je Dialogzug am Datensample, Anzahl glottaler (FG) und nicht-glottaler (FV) Füllpartikeln, relative Häufigkeit nicht-glottaler Füllpartikeln (FV/W), sowie für FV die Anzahl der vokalischen (V), vokalisch-nasalen (VN) und nasalen (N) Formen und ihre Summe je Dialogzug in BeDiaCo_c, sortiert nach FV/W (an/ap/au/aw = extscNeg./pos./unklare/komplexe Antwort, b = BACKCHANNELING, e = ERZÄHLUNG, fw/fe = W-/ENTSCHEIDUNGSFRAGE, r = REDEBEREITSCHAFT).

DZ	n	Wörter (W)	W/n	W/Σ(W)	FG	FV	FV/W	V	VN	N	Σ
r	10	18	1,8	0%	0	5	27,8%	0	4	1	5
au	11	482	43,8	1,3%	1	16	3,3%	8	6	1	15
aw	114	4513	39,6	12,3%	5	145	3,2%	67	67	3	137
ap	106	1684	15,9	4,6%	4	46	2,7%	18	25	2	45
an	56	1004	17,9	2,7%	1	27	2,7%	17	7	1	25
e	506	24430	48,3	66,5%	15	420	1,7%	173	233	4	410
fw	126	855	6,8	2,3%	0	13	1,5%	9	4	0	13
fe	176	1574	8,9	4,3%	2	17	1,1%	10	7	0	17
b	546	2204	4	6%	3	9	0,4%	2	4	2	8
Σ	1651	36764		100%	31	698		304	357	14	675

Die relative Häufigkeit nicht-glottaler Füllpartikeln über die Wörter in einer Dialogzugkategorie (Spalte FV/W) ist für REDEBEREITSCHAFT am höchsten (27,8%), gefolgt von UNSICHEREN ANTWORTEN (3,3%), KOMPLEXEN ANTWORTEN (3,2%) sowie POSITIVEN und NEGATIVEN ANTWORTEN (jeweils 2,7%). Dies ist (bis auf REDEBEREITSCHAFT) ganz ähnlich zu GECO-FP. Wie in GECO-FP sind auch die Anzahl der Wörter und die Anzahl der nicht-glottalen Füllpartikeln nicht unabhängig voneinander ($\chi^2 = 139,5$, $df =$, $p < 0,001$). Hierfür sind besonders POSITIVE,

¹³⁷Ohne die ambigen Kategorien x.b.e (3), x.aw.e (1), x.aw.fe (1), x.e (1), x.e.fe (1), und x.e.fw (1).

UNKLARE und KOMPLEXE ANTWORTEN, BACKCHANNELING, ERZÄHLUNGEN und ENTSCHEIDUNGSFRAGEN verantwortlich, die alle systematisch und signifikant zu wenig nicht-glottale Füllpartikeln enthalten, als erwartet. REDEBEREITSCHAFT hingegen enthält signifikant mehr nicht-glottale Füllpartikeln als erwartet. Zwischen der beobachteten und erwarteten Häufigkeit von Wörtern und glottalen Formen besteht hingegen kein Missverhältnis ($\chi^2 = 9, 2, p = 0, 2$).

Die nicht-glottalen Füllpartikeln in Tabelle 6.23 sind nochmals detaillierter für ihre drei häufigsten abstrahierten Formen V, VN und N genauer aufgelistet.¹³⁸ Ihre beobachteten Häufigkeiten unterscheiden sich hier (anders als in GECO-FP) signifikant von den erwarteten Häufigkeiten, d. h. ihre Verteilung hängt von den Dialogzügen ab (χ^2 -Text mit simuliertem p-Wert, basierend auf 2000 Wiederholungen; $\chi^2 = 51, 63, p = 0, 002$). In *Post-hoc*-Tests nach Shan & Gerstenberger (2017) geht dieser Unterschied insbesondere auf NEGATIVE ANTWORTEN zurück, in denen V-Formen signifikant häufiger (17 mal statt 11 mal, $p < 0, 05$) und VN-Formen signifikant seltener (7 mal statt 13 mal, $p < 0, 05$) als erwartet vorkommen, sowie auf ERZÄHLUNGEN, in denen VN-Formen (233 mal statt 217 mal, $p < 0, 05$) und N-Formen (4 mal statt 9 mal, $p < 0, 05$) signifikant häufiger als erwartet vorkommen. Zuletzt sind auch in BACKCHANNELING-Zügen N-Formen signifikant zu häufig (2 mal statt 0 mal, $p < 0, 05$).

Welche Positionen nehmen Füllpartikeln in den Dialogzügen ein? Tabelle 6.24 zeigt die Positionen von V- und VN-Füllpartikeln in jedem Dialogzug. Füllpartikelpositionen können in der EMU-Datenbank für Anfang, Mitte und Ende eines Dialogzuges ermittelt werden, wobei ‚Anfang‘ als Gleichheit der linken Intervallgrenze von Dialogzug und Füllpartikel und ‚Ende‘ als Gleichheit der rechten Intervallgrenze definiert ist. ‚Mitte‘ bezeichnet alle sonstigen Positionen. V- und VN-Formen sind für sich genommen über die drei Positionen signifikant unterschiedlich verteilt (für V: $\chi^2 = 55, 15$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0, 01$; für VN: $\chi^2 = 180, 69$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0, 001$).

Aus *Post-hoc*-Tests nach Shan & Gerstenberger (ebd.) für alle Positionen und Dialogzüge sind V-Formen wie in GECO-FP am Anfang KOMPLEXER ANTWORTEN

¹³⁸Die Summe für V, VN und N differiert von FV aufgrund von anderen Formen wie beispielsweise Sequenzen von Vokalen und Konsonanten, die hier jedoch aufgrund ihrer kleinen Zahl nicht berücksichtigt werden.

signifikant häufiger als erwartet (17 mal statt 8 mal, $p < 0,001$), in ihrer Mitte jedoch seltener (52 mal statt 59 mal, $p < 0,01$). In ERZÄHLUNGEN sind V-Formen (wie in GECO-FP) anfangs signifikant seltener (3 mal statt 21 mal, $p < 0,001$) und mittig signifikant häufiger als erwartet (170 mal statt 153 mal, $p < 0,001$). Abweichend von den Ergebnissen in GECO-FP sind V-Formen in BeDiaCo_c zudem signifikant häufiger zu Beginn von UNKLAREN ANTWORTEN (4 mal statt 1 mal, $p < 0,01$), in der Mitte von UNKLAREN ANTWORTEN jedoch signifikant seltener als erwartet (4 mal statt 7 mal, $p < 0,05$). Zu Beginn von ENTSCHEIDUNGSFRAGEN sind V-Formen signifikant häufiger als erwartet (4 mal statt 2 mal, $p = 0,1$), und in der Mitte von W-FRAGEN sind sie signifikant seltener als erwartet (5 mal statt 8 mal, $p < 0,05$).

VN-Formen zeigen für ERZÄHLUNGEN das gleiche Verhalten wie V-Formen, sie sind nämlich anfangs signifikant seltener (5 mal statt 25 mal, $p < 0,001$) und mittig signifikant häufiger als erwartet (225 mal statt 201 mal, $p < 0,001$). In Antworten auf W-Fragen sind VN-Formen zu Beginn signifikant häufiger als erwartet (22 mal statt 7 mal, $p < 0,001$); zudem sind sie gleichzeitig mittig signifikant seltener als erwartet (44 mal statt 58 mal, $p < 0,001$). Wie auch die V-Formen sind die VN-Formen in ERZÄHLUNGEN anfangs signifikant seltener (5 mal statt 25 mal, $p < 0,001$) und mittig signifikant häufiger als erwartet (225 mal statt 201 mal, $p < 0,001$). Bis hier stimmen alle über- und unterhäufigen VN-Vorkommen mit denen aus GECO-FP überein. Anders als in GECO-FP weichen VN-Formen in BeDiaCo_c in UNKLAREN ANTWORTEN nicht von ihren erwarteten Häufigkeiten ab. Zudem sind sie in BeDiaCo_c in der Mitte NEGATIVER ANTWORTEN signifikant seltener (3 mal statt 6 mal, $p < 0,01$), dafür am Ende negativer Antworten signifikant häufiger als erwartet (2 mal statt 0 mal, $p < 0,05$). Ein ähnliches Muster lässt sich in BACKCHANNELING-Zügen beobachten, nämlich mittig signifikant seltener (1 mal statt 4 mal, $p < 0,01$) und am Ende signifikant häufiger als erwartet (3 mal statt 0 mal, $p < 0,001$). Auch für W-FRAGEN und REDEBEREITSCHAFT sind die Muster ähnlich. VN-Formen sind zu Beginn von W-FRAGEN (4 mal statt 0 mal, $p < 0,001$) sowie zu Beginn der REDEBEREITSCHAFT (3 mal statt 0 mal, $p < 0,01$) signifikant häufiger als erwartet, in der Mitte von W-FRAGEN (0 mal statt 3 mal, $p < 0,001$) und von REDEBEREITSCHAFTzügen (1 mal statt 3 mal, $p < 0,01$) jedoch signifikant seltener als erwartet.

Vergleicht man V- und VN-Formen jeweils bezüglich ihrer Vorkommen für jede Position einzeln miteinander, so unterscheiden sie sich nur in ihren mittigen Positio-

Tabelle 6.24: Häufigkeit vokalischer und vokalisch-nasaler Füllpartikeln ($\Sigma(V + VN) > 7$) nach Position im Dialogzug (DZ) und relativem Anteil am absoluten Vorkommen eines DZ in BeDiaCo_c, sortiert nach Σ (an/ap/au/aw = NEG./POS./UNKLARE/KOMPLEXE ANTWORT, b = BACKCHANNELING, e = ERZÄHLUNG, fw/fe = W-/ENTSCHEIDUNGSFRAGE, r = REDEBEREITSCHAFT).

DZ	äh						ähm						Σ
	Anfang (%)		Mitte (%)		Ende (%)		Anfang (%)		Mitte (%)		Ende (%)		
e	3	(0,6)	170	(33,6)	7	(1,4)	5	(1)	225	(44,5)	4	(0,8)	414
aw	17	(14,9)	52	(45,6)	1	(0,9)	22	(19,3)	44	(38,6)	1	(0,9)	137
ap	3	(2,8)	15	(14,2)	1	(0,9)	1	(0,9)	23	(21,7)	1	(0,9)	44
an	2	(3,6)	15	(26,8)	1	(1,8)	2	(3,6)	3	(5,4)	2	(3,6)	25
fe	4	(2,3)	9	(5,1)	0	(0)	1	(0,6)	6	(3,4)	0	(0)	20
au	4	(36,4)	4	(36,4)	0	(0)	0	(0)	6	(54,5)	0	(0)	14
fw	3	(2,4)	5	(4)	1	(0,8)	4	(3,2)	0	(0)	0	(0)	13
b	1	(0,2)	1	(0,2)	0	(0)	1	(0,2)	1	(0,2)	3	(0,5)	7
r	0	(0)	0	(0)	0	(0)	3	(30)	1	(10)	0	(0)	4
Σ	37		271		11		39		309		11		678

nen ($\chi^2 = 22,6$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p < 0,001$), anders als in GECO-FP, wo es in keiner Position Unterschiede gibt. Keine Unterschiede zwischen beobachteten und erwarteten V- und VN-Formen sind am Anfang ($\chi^2 = 11$, simulierter p-Wert basierend auf 2000 Wiederholungen: $p = 0,2$) und am Ende (exakter Test nach Fisher, $p = 0,42$) von Dialogzügen zu beobachten. Offen bleibt, inwiefern der Beginn eines Dialogzuges beziehungsweise seine ‚Mitte‘ konfundierende Variablen wie beispielsweise Intonationsphrasen enthält, weswegen die Aussagekraft dieser Daten mit Vorsicht zu behandeln ist.

Tabelle 6.25 zeigt die Häufigkeitsverteilung der häufigsten nicht-glottalen Füllpartikelkontexte, in denen mindestens ein Dialogzug mehr als zehn Treffer hat, auf ausgewählte Dialogzüge. Je Dialogzug sind die Kontexte recht unterschiedlich verteilt; so kommt beispielsweise der Kontext Segment-FP-Pause fast nur in ERZÄHLUNGEN vor (nur 3 mal in KOMPLEXEN ANTWORTEN). Problematisch sind die geringen oder fehlenden Werte in den einzelnen Zellen besonders für die Auswertung kontinuierlicher Parameter wie Dauer, Grundfrequenz und Formanten. Daher betrachte ich für die weitere Auswertung nur die mit Pfeilen markierten Kontexte und die Dialogzüge *aw* und *e*.

Tabelle 6.25: Häufigkeiten aller nicht-glottalen Kontexte, in denen mindestens ein Dialogzug mehr als zehn Treffer hat, über Dialogzüge (an/ap/au/aw = NEG./POS./UNKLARE/KOMPLEXE ANTWORT, e = ERZÄHLUNG, fw/fe = W-/ENTSCHEIDUNGSFRAGE). Ante- und Postzedentia in den Kontexten beinhalten Atmung (ah/ph), Pausen (ap/pp), Segmente (as/ps), Turnpausen (at) und phonetische Reflexe (pr). Pfeile kennzeichnen Kontexte, die in die vergleichenden Modelle mit GECO-FP für die Variablen Dauer, t_0 und Formanten eingehen.

Kontexte	Dialogzüge							Σ
	an	ap	au	aw	e	fe	fw	
ah FP pr	0	0	1	7	17	0	1	26
as FP pr	0	0	0	9	16	2	1	28
ah FP pp	0	4	1	4	23	0	0	32
at FP ps	4	1	1	21	2	1	2	32
→ ap FP ps	0	4	0	8	27	1	0	40
as FP pp	1	3	0	3	33	2	0	42
→ as FP ph	4	2	1	8	36	0	1	52
→ ah FP ps	3	3	2	21	54	1	2	86
→ as FP ps	8	13	3	29	123	8	3	187
Σ	20	30	9	110	331	15	10	525

6.2.2.1 Dauer

Tabelle 6.26 enthält die analysierten Vorkommen von V- und VN-Füllpartikeln in den zwei Dialogzügen ERZÄHLUNG und KOMPLEXE ANTWORT.

Tabelle 6.26: Häufigkeit vokalischer und vokalisch-nasaler Füllpartikeln für Dialogzüge in ERZÄHLUNGEN (e) und KOMPLEXEN ANTWORTEN (aw) in den vier ausgewählten Kontexten (vgl. Tabelle 6.25) in BeDiaCo_c.

	V	VN
aw	42	25
e	125	115

Im Folgenden wird die Analyse aus GECO-FP mit den vier Kontexten *ah FP ps*, *ap FP ps*, *as FP ps* sowie *as FP ph* und den Dialogzügen ERZÄHLUNG sowie KOMPLEXE ANTWORT wiederholt. Ein Unterschied ist jedoch die Logarithmisierung der abhängigen Variable Dauer, um normalverteilte Residuen zu erhalten. Ansonsten gehen in das lineare gemischte Modell wieder Füllpartikelform, Kontext und Dialogzüge als unabhängige Variablen ein. Versuchspersonen werden als zufällige Intercepts modelliert. Interaktionen und Random Slopes verbessern das Modell nicht.

Abbildung 6.27 a–d visualisiert die Modellvorhersagen in Tabelle 6.27 im direkten Vergleich mit GECO-FP.

```
1 lmer(log(FP-Dauer) ~ Form + Kontext + Dialogzug + Sex + (1|VP),
2 data = ...)
```

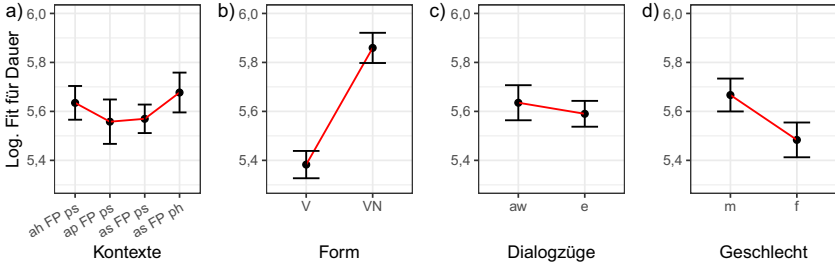


Abbildung 6.27: Modelleffekte für Füllpartikeldauern in a) sequenziellen Kontexten für Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*) und Segment-FP-Atmung-Kontexte (*as FP ph*), b) für Form, c) für Dialogstruktur (*aw* = KOMPLEXE ANTWORT, *e* = ERZÄHLUNG) und d) für Geschlecht in BeDiaCo_c.

Die Effekte aus GECO-FP für die Dialogzüge und die Kontexte können mit den Daten in BeDiaCo_c nicht repliziert werden. Der einzige signifikante Unterschied ist konsistenterweise der zwischen V- und VN-Formen. Rechnet man für alle sequenziellen Kontexte aus Tabelle 6.25 dasselbe Modell nochmals, bleibt das Ergebnis für Dialogzüge, Form und Geschlecht ähnlich. Die Effekte für Kontexte sind dann jedoch fast deckungsgleich mit dem Modell aus Abschnitt 6.2.1.1 für die sequenziellen Kontexte (vgl. Tabelle 6.16 i), weswegen sie hier nicht nochmals wiedergegeben werden.

Dass KOMPLEXE ANTWORTEN und ERZÄHLUNGEN in BeDiaCo_c keinen Einfluss auf die Dauer einer Füllpartikel haben, sollte allerdings kein Effekt der unterschiedlichen Dialogstruktursegmentierung sein. Eher liegt es möglicherweise an der Stichprobengröße. In BeDiaCo_c sind doppelt so viele VN- und viermal so viele V-Formen in KOMPLEXEN ANTWORTEN enthalten wie in GECO-FP. Insgesamt gibt es also keine konsistente Evidenz für Planungseffekte in KOMPLEXEN ANTWORTEN, die sich in der Füllpartikeldauer widerspiegeln.

Tabelle 6.27: Linear gemischte Modelle für die logarithmische Füllpartikeldauer in Abhängigkeit von *Dialogzug + Kontext + Form + Sex + (1|VP)*, mit erklärter Varianz R^2 .

	log(Dauer)
(Intercept)	5,3 (0,1)***
FormVN	0,5 (0,1)***
Kontextap FP ps	-0,1 (0,1)
Kontextas FP ps	-0,1 (0,1)
Kontextas FP ph	0,0 (0,1)
Dialogzüge	-0,0 (0,1)
sexm	0,2 (0,1)
AIC	433,6
Num. obs.	307
Num. groups: vp	24
R_m^2/R_c^2	0,23/0,35

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

6.2.2.2 Grundfrequenz

Da aus Abschnitt 6.2.1.2 und 6.2.1.2 hervorgeht, dass sich V- und VN-Formen nicht signifikant unterscheiden, werden diese Formen nicht separat im Modell betrachtet. f_0 geht als abhängige Variable in ein gemischtes lineares Modell ein; als unabhängige Variablen werden Kontext, Dialogzug und Geschlecht betrachtet. Versuchspersonen werden als Random Intercept modelliert. Interaktionen und Random Slopes verbessern das Modell nicht. Abbildung 6.28 visualisiert die Modelleffekte aus Tabelle 6.28.

```
1 | lmer(F0 ~ Kontext + Dialogzug + Sex + (1|VP), data = ...)
```

Anders als in GECO-FP ist hier kein Unterschied zwischen der Füllpartikelgrundfrequenz in ERZÄHLUNGEN und KOMPLEXEN ANTWORTEN vorhanden. Während in GECO-FP der Unterschied zwischen Atmung-FP-Segment (*ah FP ps*) und Segment-FP-Atmung (*as FP ph*) nur innerhalb KOMPLEXER ANTWORTEN und zwischen KOMPLEXEN ANTWORTEN und ERZÄHLUNGEN signifikant ist, gilt dies hier für den Unterschied von -8,2 Hz für beide Dialogzüge in gleicher Weise. Auch unterscheiden sich Füllpartikeln im Kontext *ah FP ps* signifikant um 5,3 Hz von denen in intersegmentalen Kontexten (*as FP ps*). Auch diese Ergebnisse deuten wieder auf den konfundierenden Faktor der Prosodie hin, da Füllpartikeln nach Atmung und vor Segmenten höher produziert werden als nach Segmenten und vor Atmung. Für die Dialogstruktur können die Ergebnisse aus GECO-FP jedoch nicht repliziert wer-

den. Das hier verwendete Modell erklärt zudem einen größeren Anteil der Varianz ($R_c^2 = 0,94$) als das Modell für GECO-FP ($R_c^2 = 0,58$).

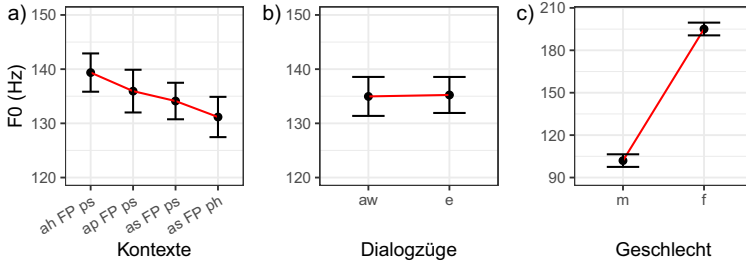


Abbildung 6.28: V- und VN-Füllpartikel- f_0 in Abhängigkeit von a) Kontexten für Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*) und Segment-FP-Atmung (*as FP ph*) sowie von b) Dialogzügen KOMPLEXE ANTWORTEN (*aw*) und ERZÄHLUNGEN (*e*) und von c) Geschlecht (freie Ordinalskala) in BeDiaCo_c.

Tabelle 6.28: Linear gemischte Modelle für die Füllpartikeldauer in Abhängigkeit von *Dialogzuge + Kontext + Form + (1/VP)*, mit erklärter Varianz R^2 .

	Grundfrequenz
(Intercept)	106,0 (4,8)***
Kontextap FP ps	-3,4 (2,8)
Kontextas FP ps	-5,3 (1,8)**
Kontextas FP ph	-8,2 (2,4)***
Dialogzuge	0,3 (2,0)
Sexf	93,1 (6,3)***
AIC	2392,8
Num. obs.	300
Num. groups: vp	24
R_m^2/R_c^2	0,85/0,94

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

6.2.2.3 Formanten

Wirkt sich die Dialogstruktur auf die Vokalqualität einer Füllpartikeln in BeDiaCo_c aus? In Tabelle 6.25 sind die Kontexte, die für die kontinuierlichen abhängigen phonetischen Parameter genauer analysiert werden, mit Pfeilen markiert, nämlich *ah FP ps*, *ap FP ps*, *as FP ps* und *as FP ph*. Im Folgenden vergleiche ich zwei Mo-

delle, nämlich das aus Tabelle 6.25 mit einem neuen Modell, das zusätzlich den Prädiktor Dialogzug enthält, für alle Faktorstufen mit $n > 6$: POSITIVE, NEGATIVE, UNKLARE UND KOMPLEXE ANTWORTEN sowie ERZÄHLUNGEN, ENTSCHEIDUNGS- und W-FRAGEN.¹³⁹ Der zusätzliche Prädiktor Dialogzug verbessert das Modell aus Abschnitt 6.2.1.3 in Likelihood-Vergleichen nicht signifikant ($p = 0, 1$). Auch für F_2 verbessert sich das Modell nicht ($p = 1$). Für F_3 wird ein Modell mit Dialogzügen gegen ein Nullmodell verglichen. Das Nullmodell ist signifikant besser ($p < 0, 05$), d. h. Dialogstruktur hat keinen Einfluss.

Aus Tabelle 6.23 geht hervor, dass ERZÄHLUNGEN und KOMPLEXE ANTWORTEN absolut die größte Anzahl an Füllpartikeln zeigen. Zusätzlich zeigt Tabelle 6.25, dass intersegmentale Füllpartikeln absolut in den neun häufigsten Kontexten jeweils am häufigsten sind. Zum Vergleich mit GECO-FP rechne ich ein neues Modell mit diesen reduzierten Daten in Tabelle 6.29, in das als Dialogzugausprägungen nur ERZÄHLUNGEN und KOMPLEXE ANTWORTEN eingehen und Füllpartikeln in intersegmentalen Kontexten mit allen anderen Kontexten in einem binären Faktor *bin.Kontext* kontrastiert werden.

Tabelle 6.29: Anzahl der für Formanten analysierten Füllpartikelformen in KOMPLEXEN ANTWORTEN und ERZÄHLUNGEN in intersegmentalen (*as FP ps*) und nicht-intersegmentalen (anderen) Kontexten in BeDiaCo v. 1.

Dialogzug	bin.Kontext	V	VN
aw	as FP ps	24	4
aw	andere	43	62
e	as FP ps	86	37
e	andere	87	196

Für F_1 ist ein Modell mit den Daten in Tabelle 6.29 nicht signifikant besser als ein Modell, welches nur sequenzielle Kontexte enthält ($p = 0, 2$). Auch für F_2 verbessert der Faktor Dialogzug das Modell nicht signifikant ($p = 0, 1$). Im Gegensatz zu GECO-FP ist hier ein Modell für F_3 mit den beiden Dialogzügen ERZÄHLUNG und KOMPLEXE ANTWORT jedoch signifikant besser als ein Modell ohne Dialogzüge ($p < 0, 05$). Abbildung 6.29 visualisiert die Effekte für die F_3 -Werte in Tabelle 6.30 aus folgendem Modell:

¹³⁹Hierzu müssen allerdings alle neun sequenziellen Kontexte aus Tabelle 6.25 berücksichtigt werden, also auch *as FP pp*, *at FP ps*, *ah FP pr*, *as FP pr* und *ah FP pp*. Abhängige Variable ist F_1 (in der Mitte des Vokals gemessen).


```
1 lmer(F3 ~ Form + bin.Kontext + Dialogzug + Sex + log(Dauer) + (1|VP) +
  (0+log(Dauer)|VP), data = ...)
```

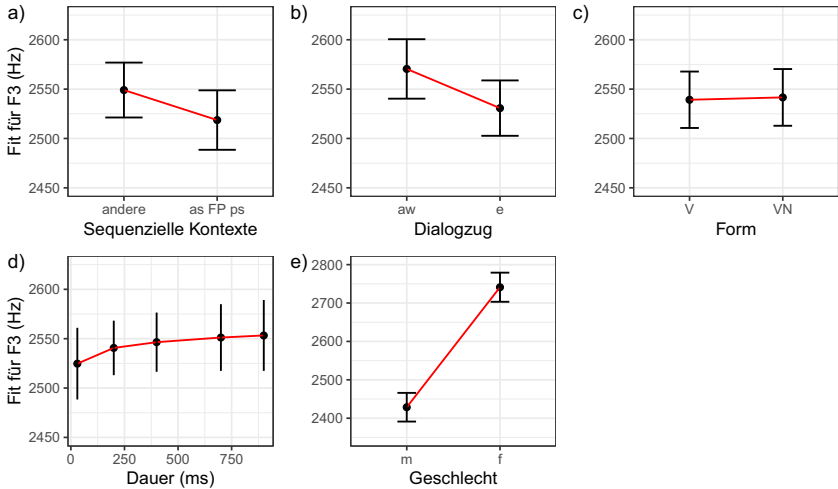


Abbildung 6.29: Anpassung für F_3 in Abhängigkeit von a) sequenziellen Kontexten (intersegmental = *as FP ps*; nicht-intersegmental = *andere*), b) Dialogzug (*aw* = KOMPLEXE ANTWORTEN, *e* = ERZÄHLUNGEN), c) Form (*V* = vokalisch, *VN* = vokalisch-nasal), d) Dauer und e) Geschlecht (andere Ordinalskala) in BeDiaCo_c.

Weder Füllpartikelform, noch -dauer oder sequenzieller Kontext wirkt sich signifikant auf F_3 aus. Allerdings ist F_3 in ERZÄHLUNGEN signifikant um 39,7 Hz niedriger als in KOMPLEXEN ANTWORTEN. Unklar bleibt, woran das liegt. Bis hierher können die Ergebnisse aus GECO-FP nicht repliziert werden. Männer zeigen signifikante niedrigere F_3 -Werte als Frauen, was wieder eine Folge der natürlichen Ansatzrohrvariation ist.

Trotz der nichtreplizierten Unterschiede für die Interaktion der beiden Dialogzüge KOMPLEXE ANTWORT und ERZÄHLUNG sowie die beiden inter- und nicht-intersegmentalen Kontextbedingungen für im mittleren Teil des Vokals gemessene Formanten werde im Folgenden zur Vergleichbarkeit mit GECO-FP die beiden Kontext- und Dialogzuggruppen auf Unterschiede in ihren Formantrajektorien untersuchen. Hierzu nutze ich wieder generalisierte additive Modelle (GAM), die die

Tabelle 6.30: Lineares gemischtes Modell für F_3 in Abhängigkeit von *Dialogzug + bin.Kontext + Form + Sex + log(Dauer) + (1/VP) + (0+log(Dauer)/VP)*, mit erklärter Varianz R^2 .

	F_3
(Intercept)	2703,5 (80,5)***
Dialogzuge	-39,7 (17,2)*
FormVN	2,4 (15,6)
bin.Kontextandere	30,4 (15,7)
log(Dauer)	8,4 (13,7)
sexm	-312,6 (53,2)***
AIC	6917,7
Num. obs.	539
Num. groups: vp	24
R_m^2/R_c^2	0,39/0,65

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

zeitnormalisierten und lobanovnormalisierten F_1 -Werte in Abhängigkeit von Füllpartikelform, Kontext und Dialogzug in dem Interaktionsprädiktor *DZ.Form.Kontext* berücksichtigen und mit Likelihood-Tests (ML-gefittet) verglichen werden. Das beste Modell (vgl. R-Code unten) enthält die Interaktion aller drei Prädiktoren (Z. 1), Kurvenglättungen für diese Prädiktoren (Z. 2) sowie zufällige Kurvenglättungen über die Prädiktoren je Versuchsperson (Z. 3) und eine Korrektur für Autokorrelation (Z. 4).

```

1 bam(lob.f1 ~ DZ.Form.Kontext +
2     s(Zeit, by = DZ.Form.Kontext, k = 10) +
3     s(Zeit, VP, by = DZ.Form.Kontext, bs = "fs", m = 1),
4     rho = rho.acf, AR.start = dat$start.event,
5     data = dat)

```

Abbildung 6.30 stellt die im Modell angepassten Trajektorien für F_1 und die Differenzen ihrer Kurven dar. Signifikante Unterschiede der Trajektorien sind vorhanden, wenn die Differenzkurve von der Nulllinie abweicht. Dies ist nur für 6.30 f der Fall, dort weicht die Trajektorie der V-Füllpartikeln in nicht-intersegmentalen Kontexten in KOMPLEXEN ANTWORTEN (6.30 e) zwischen 0% und 38,3% der normalisierten Zeit signifikant von der in ERZÄHLUNGEN ab. Damit werden die Ergebnisse aus BeDiaCo v. 1 nur zum Teil bestätigt; dort sind auch Teile der F_1 -Trajektorie in nicht-intersegmental stehenden VN-Füllpartikeln signifikant verschieden, hier jedoch nicht. Tabelle 6.31 zeigt die Werte für die im Modell enthaltenen Glättungsdrücke.

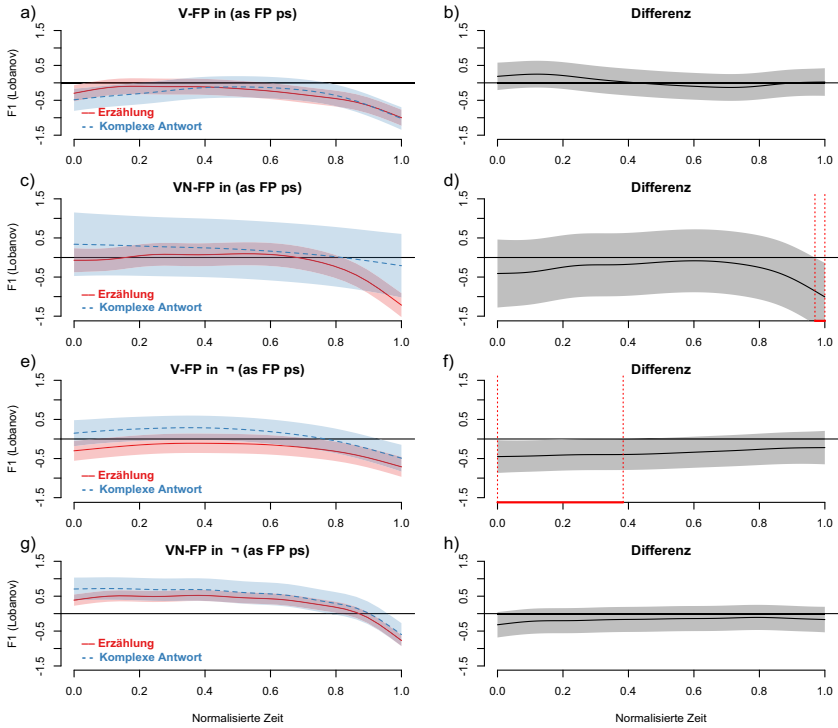


Abbildung 6.30: Angepasste F_1 -Trajektorien der Vokale in V- und VN-Füllpartikel in intersegmentalen Kontexten (as FP ps) für die Dialogzüge ERZÄHLUNG (rot) und KOMPLEXE ANTWORT (blau) in a)–d) in nicht-intersegmentalen Kontexten (\neg as FP ps) in e)–h) in BeDiaCoC. Die rechten Bilder zeigen die Differenz der jeweils links abgebildeten Trajektorien; rote Balken kennzeichnen signifikante Abweichungen von der Nulllinie.

Für F_2 gibt es wie schon für die Vokalmittelpunkte keine Dialogzugeffekte für V- und VN-Formen in intersegmentalen und nicht-intersegmentalen Kontexten. Von der Analyse der F_3 -Trajektorien wird abgesehen, da ihre Interpretation unklar bleibt.

Tabelle 6.31: Effektive Freiheitsgrade (edf), F- und p-Werte für die unabhängigen (DZ.Form.Kontext) und zufälligen (DZ.Form.Kontext je Versuchsperson) Glättungsausdrücke der Trajektorien für F_1 über die normalisierte Zeit.

	edf	F	p
s(Zeit):DZ.Form.Kontextaw.äh.intersegmental	5,23	11,69	< 0,001
s(Zeit):DZ.Form.Kontexte.äh.intersegmental	7,11	32,51	< 0,001
s(Zeit):DZ.Form.Kontextaw.ähm.intersegmental	1,89	1,40	n, s,
s(Zeit):DZ.Form.Kontexte.ähm.intersegmental	6,12	15,54	< 0,001
s(Zeit):DZ.Form.Kontextaw.äh.andere	3,70	6,38	< 0,001
s(Zeit):DZ.Form.Kontexte.äh.andere	4,90	10,96	< 0,001
s(Zeit):DZ.Form.Kontextaw.ähm.andere	7,29	23,66	< 0,001
s(Zeit):DZ.Form.Kontexte.ähm.andere	8,15	58,85	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.äh.intersegmental	13,07	0,30	< 0,001
s(Zeit,vp):DZ.Form.Kontexte.äh.intersegmental	33,47	0,79	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.ähm.intersegmental	16,93	0,67	< 0,001
s(Zeit,vp):DZ.Form.Kontexte.ähm.intersegmental	55,01	0,81	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.äh.andere	90,39	1,04	< 0,001
s(Zeit,vp):DZ.Form.Kontexte.äh.andere	109,66	1,46	< 0,001
s(Zeit,vp):DZ.Form.Kontextaw.ähm.andere	97,01	1,49	< 0,001
s(Zeit,vp):DZ.Form.Kontexte.ähm.andere	138,28	1,96	< 0,001

6.2.2.4 Phonationsart

Für die Verteilung von Füllpartikeln in Dialogzügen mit einer absoluten Häufigkeit von $n > 3$ je Dialogzug zeigt Abbildung 6.31 die absoluten und relativen Werte je Phonationsart. Ohne getrennte Betrachtung für Männer und Frauen weicht die Verteilung wie in GECO-FP nicht signifikant von den erwarteten Werten ab ($\chi = 12, 8$, p-Wert basierend auf 2000 Wiederholungen: $p = 0, 1$). Auch Frauen zeigen keine Abweichungen von der erwarteten Verteilung (Fishers exakter zweiseitiger Test, p-Wert basierend auf 2000 Wiederholungen: $p = 0, 1$). Für Männer weicht die Verteilung hingegen signifikant von der erwarteten ab ($\chi = 25, 1$, p-Wert basierend auf 2000 Wiederholungen: $p < 0, 01$). *Post-hoc*-Tests nach Shan & Gerstenberger (2017) zeigen, dass dies jedoch nur auf die übererwartet häufigen glottalisierten Füllpartikeln in ENTSCHEIDUNGSFRAGEN (*fe*) zurückgeht (4 statt 1,1). Insgesamt kann also auch aus den Ergebnissen von BeDiaCo_c kein Effekt der Dialogstruktur auf die Füllpartikeln postuliert werden.

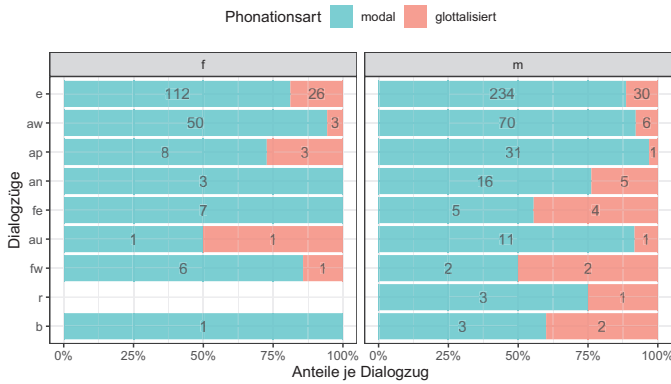


Abbildung 6.31: Absolute und relative Anteile von modalen und glottalisierten Füllpartikeln Dialogzügen ($n > 3$) in BeDiaCo_c, sortiert nach Häufigkeit je DZ (e = ERZÄHLUNG, ap/an/au/aw = POSITIVE/NEGATIVE/UNKLARE/KOMPLEXE ANTWORT, fe/fw = ENTSCHEIDUNGS-/W-FRAGE, r = REDEBEREITSCHAFT, b = BACKCHANNELING).

6.2.3 Intonationsphrasen

Die Muster sequenzieller Kontexte in GECO-FP lassen vermuten, dass die prosodische Struktur für die Formvariation eine erhebliche Rolle spielt. Daher sind in BeDiaCo v. 1 zusätzlich Intonationsphrasen (IP) annotiert. Diese Annotation wird in diesem Kapitel für Dauer, Grundfrequenz, Formanten und Phonationsart bezüglich Füllpartikelform (hier nur V- vs. VN-Formen), sequenzieller Kontexte, Dialogzüge und Geschlecht ausgewertet.

Insgesamt sind in 9% aller Intonationsphrasen Füllpartikeln enthalten (in 708 aus 7876). Dies sind wesentlich weniger als bei Swerts (1998) (45%). Die IP-Position von Füllpartikeln kann vier Ausprägungen annehmen: initial (direkt zu Beginn der IP), medial (weder initial noch final), final (direkt am Ende der IP) und kongruent (in kongruenter Position stimmen Anfang und Ende einer Füllpartikel mit Anfang und Ende einer IP überein, sie bildet also eine eigenständige IP).¹⁴⁰ Füllpartikeln in IPn kommen 159 mal in initialer, 131 mal in medialer, 180 mal in kongruenter und 136 mal in finaler Position vor.¹⁴¹ Abbildung 6.32 stellt die Anteile von Füllpar-

¹⁴⁰Diese IP-Positionen ergeben sich aus einer kombinierten Abfrage der Intonationsphrasenebene und der Füllpartikelebene.

¹⁴¹Füllpartikeln innerhalb des Annotationswertes *ipp* sind seltener (2 mal initial, 2 mal medial,

tikelform und Geschlecht an den verschiedenen Positionen von Intonationsphrasen dar. Es sind keine signifikanten Unterschiede feststellbar.

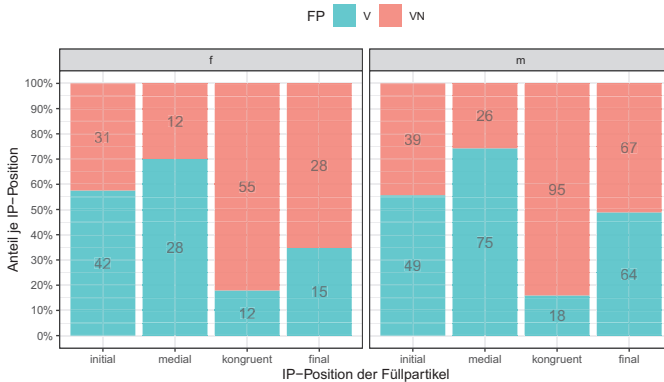


Abbildung 6.32: Absolute und relative Anteile der V- und VN-Füllpartikelformen je Intonationsphrasenposition je Geschlecht in BeDiaCo_c.

Abbildung 6.33 a stellt die Anteile der IP-Positionen von Füllpartikeln an den neun häufigsten sequenziellen Kontexten, Abbildung 6.33 b an Dialogzügen dar. Für den intersegmentalen Kontext bedeutet beispielsweise eine kongruente oder finale Position, dass sich danach die nächste Intonationsphrase ohne Pause anschließt, während in initialer IP-Position eine Intonationsphrase direkt vorangeht. Da Pausen oft IP-Grenzen bilden, sind diese Instanzen intersegmental jedoch nicht so häufig.

Füllpartikeln stehen in den Intonationsphrasen der Kontexte Atmung-FP-Pause (*ah FP pp*) und Atmung-FP-Reflex (*ah FP pr*) fast nur in kongruenter Position, d. h. sie bilden zwischen zwei (wie auch immer ausgeprägten) Pausen eine eigene Intonationsphrase. In den Kontexten Atmung-FP-Segment (*ah FP ps*), Pause-FP-Segment (*ap FP ps*) und Turnpause-FP-Segment (*at FP ps*) kommen sie fast nur in initialer IP-Position vor (stehen sie dort in kongruenten Fällen, bedeutet dies, dass direkt im Anschluss die nächste Intonationsphrase folgt). Fast spiegelbildlich dazu stehen sie in den Kontexten Segment-FP-Atmung (*as FP ph*), Segment-FP-Pause (*as FP pp*) und Segment-FP-Reflex (*as FP pr*) überwiegend in finaler IP-Position, während sie IP-medial in nennenswerter Größe besonders in intersegmentalen Kontexten auf 38 mal final) und wird daher nicht weiter betrachtet, was 7,6 % der Formen ausschließt.

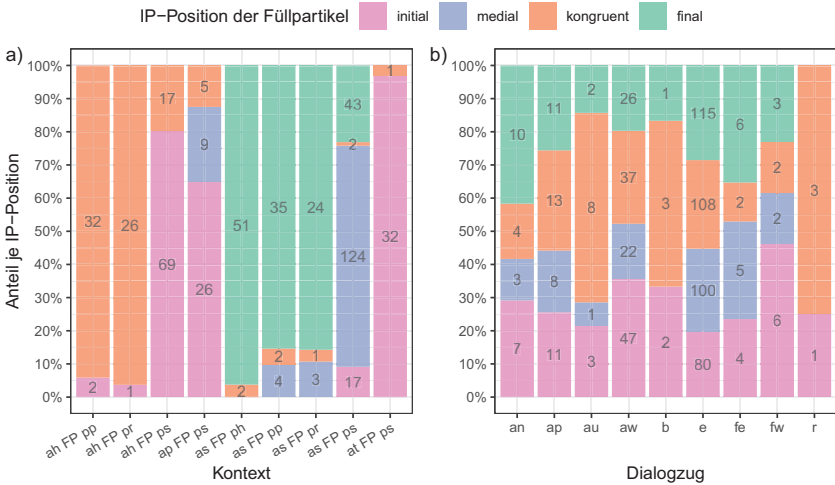


Abbildung 6.33: Absolute und relative Anteile der Füllpartikelposition in Intonationsphrasen je a) Kontext Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*), Segment-FP-Reflex (*as FP pr*) und Atmung-FP-Reflex (*ah FP pr*) sowie b) Dialogzug (e = ERZÄHLUNG, ap/an/au/aw = POS./NEG./UNKLARE/KOMPLEXE ANTWORT, fw/fe = W-/ENTSCHEIDUNGSFRAGE, r = REDEBEREITSCHAFT, b = BACKCHANNELING) in BeDiaCo_c.

treten. Somit besteht hier ein deutlicher Zusammenhang zwischen den sequenziellen Kontexten und der Position in Intonationsphrasen, wenngleich er nicht immer in Gänze determiniert ist (vgl. die Aufteilung in *as FP ps* und *ap FP ps*). Damit wird die Hypothese aus Abschnitt 5.2.1.2 und Abschnitt 6.2.1.2 wahrscheinlicher, dass eine erhöhte Grundfrequenz nach Atmung oder Turnpausen und vor Segmenten einer initialen Intonationsphrasenposition zugeschrieben werden kann.

Die IP-Positionen sind über die Dialogzüge in Abbildung 6.33 b signifikant unterschiedlich verteilt ($\chi^2 = 41,9$, p-Wert nach 2000 Simulationen: $p < 0,01$). IP-initiale Füllpartikeln treten in KOMPLEXEN ANTWORTEN signifikant häufiger auf als erwartet (47 mal statt 32 mal, $p < 0,01$), in ERZÄHLUNGEN hingegen signifikant seltener (80 mal statt 99 mal, $p < 0,001$). Ebenso in ERZÄHLUNGEN kommen sie IP-medial signifikant zu häufig vor (100 mal statt 87 mal, $p < 0,05$). In IP-kongruenter Positi-

on sind Füllpartikeln in UNSICHEREN ANTWORTEN signifikant häufiger als erwartet (8 mal statt 4 mal, $p < 0,05$) und in IP-finaler Form in KOMPLEXEN ANTWORTEN signifikant zu selten (26 mal statt 35 mal, $p < 0,05$). IP und Dialogstruktur zeigen also gemeinsam ein deutlicheres Bild: In Antworten stehen Füllpartikeln eher IP-initial, in ERZÄHLUNGEN eher IP-medial.

6.2.3.1 Dauer

Ein Modell für die Dauer, das die Position einer Füllpartikel berücksichtigt, ist signifikant besser als ein Modell ohne IP-Position ($p < 0,001$). Als unabhängige Variable gehen der sequenzielle Kontext, die Form und die Position in der Intonationsphrase ein, mit Random Intercepts für Versuchspersonen und unkorrelierten Random Slopes für die Form je Versuchsperson. Weder Interaktionen zwischen den Faktoren noch der Faktor Geschlecht oder der Faktor Dialogzug verbessern das Modell.

```
lmer(FP-Dauer ~ Kontext + Form + IP-Position + (1|VP) + (0+Form|VP),
data ...)
```

Abbildung 6.34 a–c visualisiert die Modellvorhersagen in Tabelle 6.32. Ausgehend

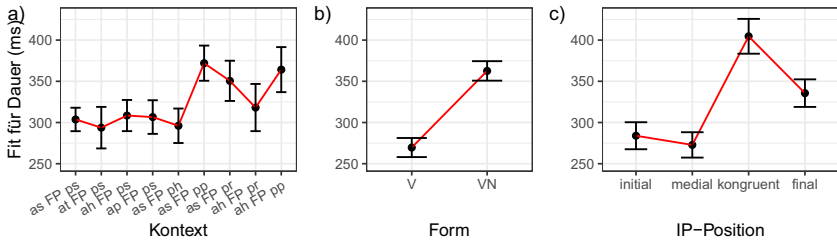


Abbildung 6.34: Modelleffekte für Füllpartikeldauern in adjazenten sequenziellen Kontexten der nicht-glottalen V- und VN-Füllpartikeln (FP) für a) die Kontexte Atmung-FP-Segment- (*ah FP ps*), Pause-FP-Segment- (*ap FP ps*), Segment-FP-Segment- (*as FP ps*), Segment-FP-Pause- (*as FP pp*), Segment-FP-Atmung- (*as FP ph*), Segment-FP-Reflex (*as FP pr*) und Atmung-FP-Reflex (*ah FP pr*), b) Form und c) IP-Position in BeDiaCo_c.

von intersegmentalen Kontexten sind Füllpartikeln in Segment-FP-Pause-Kontexten (*as FP pp*) signifikant (68,3ms) länger. In einem *Post-hoc*-Vergleich (vgl. Tabelle 6.33) ist eine Füllpartikel im Kontext Segment-FP-Pause außerdem signifikant länger als im Kontext Segment-FP-Atmung (*as FP ph*, 75,9ms) – und das obwohl beide fast nur in finaler IP-Position vorkommen (vgl. Abbildung 6.33 a). Die

IP-Position moduliert den Effekt der sequenziellen Kontexte: Dass hier nur noch Segment-FP-Pause-Kontexte (*as FP pp*) signifikant länger als intersegmentale Füllpartikeln (anders als in Abschnitt 6.2.1.1), hängt damit zusammen, dass Füllpartikeln mit rechtsstehenden Pausen (Stille, Atmung oder mit phonetischem Reflex) fast nur in finaler IP-Position stehen (vgl. Abbildung 6.33 a). Der alleinige Effekt sequenzieller Kontexte fällt weg, stattdessen kann die Dauer einer Füllpartikel die Position, an der ihr sequenzielle Kontext auftritt, und damit über ihre Position in einer IP erklärt werden.

Tabelle 6.32: Lineares gemischtes Modell für die Füllpartikeldauer in Abhängigkeit von *Kontext + Form + IP-Position + (1/VP)*, mit erklärter Varianz R^2 .

	Dauer
(Intercept)	225,6 (21,7)***
Kontextat FP ps	-10,0 (28,1)
Kontextah FP ps	4,7 (22,5)
Kontextap FP ps	2,9 (22,8)
Kontextas FP ph	-7,7 (20,8)
Kontextas FP pp	68,3 (20,8)**
Kontextas FP pr	46,9 (24,0)
Kontextah FP pr	14,4 (32,4)
Kontextah FP pp	60,5 (31,1)
FormVN	93,0 (11,4)***
IPmedial	-11,1 (20,9)
IPkongruent	120,6 (22,0)***
IPfinal	51,7 (23,9)*
AIC	6053,2
Num. obs.	505
Num. groups: vp	24
R_m^2/R_c^2	0,4/0,49

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

VN-Formen sind signifikant länger als V-Formen. Füllpartikeln in einem intersegmentalen Kontext sind in kongruenter (120,6 ms) und finaler Position (51,7 ms) signifikant länger als zu Beginn einer Intonationsphrase. *Post-hoc*-Vergleiche (Tabelle 6.33) zeigen, dass Füllpartikeln in finaler Position signifikant länger (62,8 ms) als in medialer Position sind. Dieser prosodische Effekt der finalen Längung lässt sich auch in anderen Wörtern des Deutschen beobachten (fürs Deutsche Kohler 1983; Peters 2003; vgl. für einen Überblick Cho 2015). Auch kongruente IP-Positionen sind signifikant länger als mediale (131,8 ms). Sie bilden aber einen (bisher unbeschriebenen)

Sonderfall, da sie aus genau einer Intonationsphrase bestehen.

Tabelle 6.33: Signifikante Kontraste der Kontexte und IP-Position für die Dauer (ohne Referenzniveau), mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95%-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
as FP ph - as FP pp	-75,9	21,9	479,9	-144,18	-7,70	-3,47	<0,05
medial - kongruent	-131,8	26,2	483,6	-199,30	-64,24	-5,03	<0,001
medial - final	-62,8	17,0	481,1	-106,74	-18,88	-3,69	<0,01

Da sich Füllpartikeln also an IP-Grenzen längen, liegt es nahe, den Effekt der Längung *en detail* zu analysieren. Für finale Längung an IP-Grenzen wird in vielen Sprachen beobachtet, dass der Längungseffekt graduell ist, d. h. dass die Segmente direkt an der Grenze stärker gedehnt werden als Segmente, die weiter von der Grenze entfernt sind (für einen Überblick vgl. Krivokapić 2014). Dieses graduelle Längung wird innerhalb der Artikulatorischen Phonologie (Browman & Goldstein 1992) durch eine prosodische Geste modelliert, die π -Geste (Byrd & Saltzman 2003), die an der Grenze aktiv ist und die Äußerungszeit dort lokal verlangsamt. Da in VN-Formen zwei Segmente vorhanden sind, sagt die π -Geste vorher, dass der Nasal stärker gelängt wird als der Vokal. Zur Überprüfung rechne ich ein Modell mit der logarithmischen Dauer des Vokals und Nasals (um normalverteilte Residuen zu erhalten) in VN-Formen als abhängiger und Segmentart in Interaktion mit IP-Position als unabhängiger Variable, mit Random Intercepts für Versuchspersonen und Random Slopes für die Segmente je Versuchsperson. Das Geschlecht verbessert das Modell nicht. Abbildung 6.35 visualisiert die Modellvorhersagen in Tabelle 6.34.

```

1 | lmer(log(Dauer) ~ Segment * IP-Position + (1+Segment|VP) +
  | (0+Form|VP), data ...)
```

Abbildung 6.35 zeigt, dass Nasale in VN-Formen signifikant gelängt werden: von initialer zu finaler Position um 41 %, $p < 0,001$ (Vokale werden sogar um 7 % kürzer, n. s.) ; von medialer zu finaler Position um 56 %, $p < 0,001$ (Vokale um 13 %, n. s.). Die Unterschied zwischen finaler zu kongruenter Position ist für Nasale in Tukey-korrigierten *Post-hoc*-Tests nicht signifikant (15 %), aber für Vokale (37 %, $p < 0,001$). In kongruenter Position sind Nasale (81 %, $p < 0,001$) und Vokale (56 %, $p < 0,001$) signifikant länger als in medialer Position.

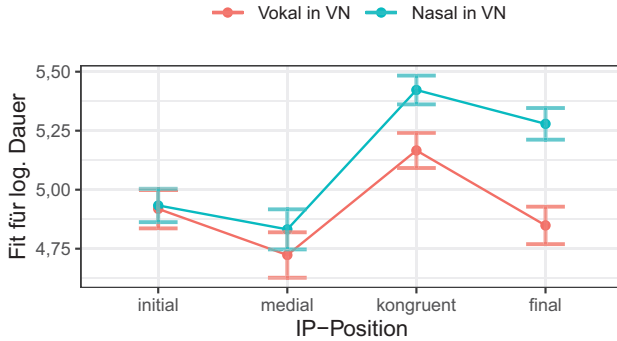


Abbildung 6.35: Modelleffekte für die Dauer des Vokals und Nasals in VN-Formen für IP-Positionen in BeDiaCo_e.

Tabelle 6.34: Linear gemischtes Modell für die logarithmische Dauer der Segmente in VN-Formen in Abhängigkeit von *IP-Position + (1/VP)*, mit erklärter Varianz R^2 .

	log (Dauer)
(Intercept)	4,9 (0,1)***
IPmedial	-0,2 (0,1)*
IPkongruent	0,2 (0,1)***
IPfinal	-0,1 (0,1)
Segmentm	0,0 (0,1)
IPmedial:Segmentm	0,1 (0,1)
IPkongruent:Segmentm	0,2 (0,1)**
IPfinal:Segmentm	0,4 (0,1)***
AIC	839,3
Num. obs.	718
Num. groups: vp	24
R_m^2/R_c^2	0,19/0,46

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

6.2.3.2 Grundfrequenz

Ein Modell für die Grundfrequenz, das die Position einer Füllpartikel berücksichtigt, ist nicht signifikant besser als ein Modell ohne IP-Position ($p = 0,1$). Es kann die Daten für die Grundfrequenz also nicht besser erklären, als das Modell in Tabelle 6.17. Allerdings ist fraglich, ob tatsächlich alle bisherigen Faktoren benötigt werden. Ein

Nullmodell, welches ausgehend von dem Faktor Geschlecht noch IP-Position hinzufügt, wird signifikant besser ($p < 0,01$). Das weitere Hinzufügen von Kontext ($p = 0,7$), Form ($p = 0,5$) oder Dialogzügen ($p = 0,1$) trägt zu keiner signifikanten Verbesserung bei. Versuchspersonen werden als Random Intercepts modelliert. Abbildung 6.36 visualisiert die Modellvorhersagen in Tabelle 6.35.

```
lmer(FP-Dauer ~ Sex + IP-Position + (1|VP), data ...)
```

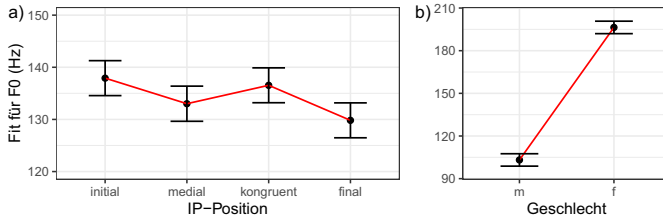


Abbildung 6.36: Modelleffekte für die Füllpartikelgrundfrequenz für a) IP-Position und b) Geschlecht in BeDiaCo_c.

Ausgehend von initialen IP-Positionen werden Füllpartikeln in medialer (-4,9 Hz) und finaler (-8,1 Hz) Position signifikant tiefer geäußert. In einem *Post-hoc*-Vergleich (vgl. Tabelle 6.36) ist eine Füllpartikel in kongruenter Position signifikant höher als in finaler Position (6,7 Hz). Die Grundfrequenz von Füllpartikeln sinkt also über die Intonationsphrase hinweg ab.

Tabelle 6.35: Lineares gemischtes Modell für die Füllpartikelgrundfrequenz in Abhängigkeit von *IP-Position + Sex + (1|VP)*, mit erklärter Varianz R^2 .

Dauer	
(Intercept)	200,0 (4,4)***
sexm	-93,2 (6,2)***
IPmedial	-4,9 (1,4)***
IPkongruent	-1,4 (1,3)
IPfinal	-8,1 (1,3)***
AIC	4912,6
Num. obs.	628
Num. groups: vp	24
R_m^2/R_c^2	0,85/0,95

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Tabelle 6.36: Signifikante Kontraste der Kontexte und IP-Position für die Grundfrequenz (ohne Referenzniveau), mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
medial - kongruent	-3,5	1,4	604,5	-7,05	-0,02	-2,59	<0,05
kongruent - final	6,7	1,3	604,1	3,37	10,07	5,16	<0,001

6.2.3.3 Formanten

Bisher gibt es in den Auswertungen für Dauer und Grundfrequenz keine signifikanten Interaktionen zwischen IP-Positionen und sequenziellen Kontexten. Wie in Abbildung 6.33 a zu sehen ist, kommen bestimmte Kontextausprägungen fast ausschließlich in bestimmten IP-Positionen vor. Es kann also davon ausgegangen werden, dass beide Faktoren, sequenzielle Kontexte und IP-Position, zumindest manche Formvariation in ähnlicher Weise erklären, jedoch möglicherweise nicht immer eine additive Erklärung beitragen (außer in dem Modell für die Dauer). Auch Effekte in den Modellen für F_1 , F_2 und F_3 , in denen IP-Position hinzugefügt wird, ergeben tatsächlich in weiten Teilen ähnliche Ergebnisse wie in Tabelle 6.21 i, ii und v. Der zusätzliche Prädiktor IP-Position verbessert diese Modelle nicht signifikant. Für die Formanten beziehe daher wie für die Grundfrequenz statt der sequenziellen Kontexte die IP-Position in die Modelle ein. Der Prädiktor Dialogzug¹⁴² verbessert dieses Modell signifikant (für F_1 $p < 0,01$). Abbildung 6.37 visualisiert die Modellvorhersagen für F_1 in Tabelle 6.37 i; das Modell ist folgendermaßen aufgebaut.

```
1 lmer(F1 ~ Form + log(Dauer) + Sex + IP-Position + Dialogzug + (1|VP) +
  (0+log(Dauer)|VP), data ...)
```

Ausgehend von initialer Position in Intonationsphrasen werden Füllpartikeln in medialer (-38,4 Hz) und finaler (-43,9 Hz) Position mit einem signifikant tieferem F_1 geäußert, liegen also höher im Vokalraum. *Post-hoc*-Vergleiche (vgl. Tabelle 6.38) zeigen außerdem, dass der Vokal in kongruenter Position signifikant tiefer im Vokalraum liegt als in medialer und finaler Position.

Füllpartikelvokale liegen in NEGATIVEN ANTWORTEN signifikant tiefer im Vokalraum als in POSITIVEN (47,5 Hz), UNKLAREN (79,8 Hz) und KOMPLEXEN AN-

¹⁴²Aufgrund der wenigen Instanzen ohne BACKCHANNELING und REDEBEREITSCHAFT.

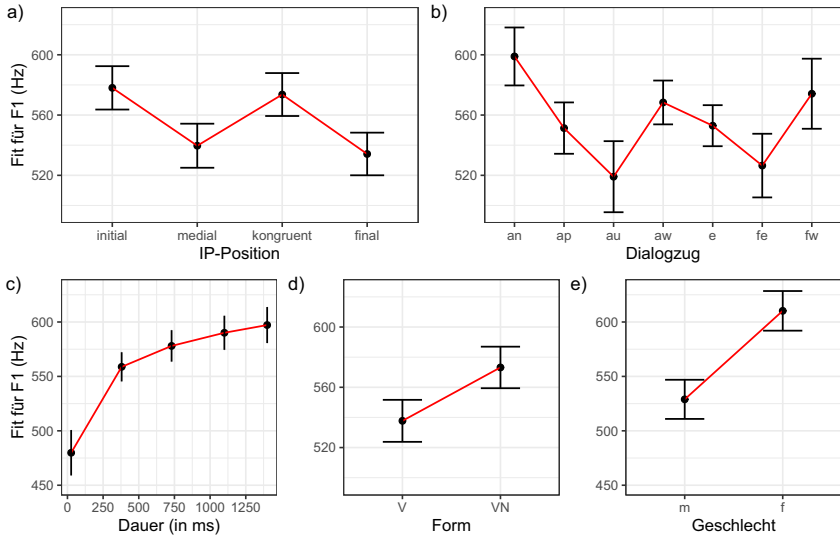


Abbildung 6.37: Modelleffekte von F_1 für a) IP-Position, b) Dialogzug (e = ERZÄHLUNG, ap/an/au/aw = POSITIVE/NEGATIVE/UNKLARE/KOMPLEXE ANTWORT, fe/fw = ENTSCHEIDUNGS-/W-FRAGE), c) Form, d) Dauer und e) Geschlecht in BeDiaCoC.

WORTEN (30,5 Hz) sowie in ERZÄHLUNGEN (45,9 Hz) und ENTSCHEIDUNGSFRAGEN (72,4 Hz). VN-Formen liegen signifikant tiefer im Vokalraum als V-Formen (35,4 Hz). Je länger der Vokal andauert, desto tiefer liegt er (vgl. Abbildung 6.37 d). Frauen zeigen einen signifikant höheren ersten Formanten als Männer, was aufgrund ihres oft kürzeren Ansatzrohres erwartbar ist.

Für F_2 ergibt sich folgendes Modell. Der Prädiktor Dialogstruktur verbessert das Modell nicht signifikant. Abbildung 6.38 visualisiert die Modellvorhersagen für F_2 in Tabelle 6.37 ii.

$$1 \quad \text{lmer}(F_2 \sim \text{Form} + \log(\text{Dauer}) + \text{Sex} + \text{IP-Position} + (1|\text{VP}) + (0+\log(\text{Dauer})|\text{VP}), \text{data} \dots)$$

Ausgehend von initialer Position in Intonationsphrasen werden Füllpartikeln in medialer (81,1 Hz) und finaler (74,2 Hz) Position mit einem signifikant tieferem F_2 geäußert, liegen also posteriorer im Vokalraum. *Post-hoc*-Vergleiche für F_2 (vgl. Tabelle 6.38) zeigen außerdem, dass der Vokal in kongruenter Position signifikant ante-

Tabelle 6.37: Linear gemischte Modelle für Formanten (IP = Intonationsphrasenposition, DZ = Dialogzug), mit erklärter Varianz R^2 .

	i) F ₁	ii) F ₂	iii) F ₃
(Intercept)	482,2 (41,2)***	2085,7 (74,6)***	2756,3 (39,9)***
FormVN	35,4 (7,1)***	-59,9 (14,5)***	
log(Dauer)	29,4 (6,5)***	-79,0 (13,7)***	
sexm	-81,3 (25,6)**	-192,6 (39,4)***	-282,8 (54,8)***
IPmedial	-38,4 (8,4)***	-81,1 (16,9)***	-55,8 (17,6)***
IPkongruent	-4,4 (8,5)	-19,4 (17,4)	3,1 (16,5)
IPfinal	-43,9 (8,1)***	-74,2 (16,4)***	-54,2 (16,8)**
DZap	-47,5 (17,9)**		
DZau	-79,8 (24,1)***		
DZaw	-30,5 (15,5)*		
DZe	-45,9 (14,6)**		
DZfe	-72,4 (21,8)***		
DZfw	-24,7 (23,8)		
AIC	7226,6	8163,7	8251,9
Num. obs.	642	642	642
Num. groups: vp	24	24	24
R_m^2/R_c^2	0,28/0,6	0,34/0,56	0,35/0,64

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

riorer im Vokalraum liegt als in medialer und finaler Position. Vokale in VN-Formen liegen signifikant posteriorer im Vokalraum als in V-Formen (59,9 Hz). Je länger der Vokal andauert, desto posteriorer liegt er (vgl. Abbildung 6.38 c). Frauen zeigen einen signifikant höheren zweiten Formanten als Männer, was aufgrund ihres oft kürzeren Ansatzrohres erwartbar ist.

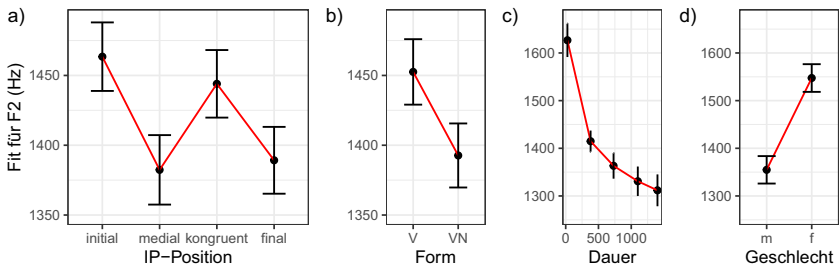


Abbildung 6.38: Modelleffekte von F_2 für a) IP-Position, b) Form, c) Dauer und d) Geschlecht (unterschiedliche Ordinalskala in a–b und c–d) in BeDiaCo_c.

Tabelle 6.38: Signifikante Kontraste der IP-Position für F_1 , F_2 und F_3 (ohne Referenzniveau), mit Schätzer β und Standardabweichung (s) für f_0 , Freiheitsgraden (df), 95 %-Konfidenzintervall (KI) und t-Statistik, korrigiert für multiple Vergleiche nach Tukey.

Vergleich	β	s	df	Unteres KI	Oberes KI	t	p
F_1							
medial - kongruent	-34,0	9,2	611,9	-57,77	-10,18	-3,68	<0,01
kongruent - final	39,4	7,9	612,3	18,98	59,92	4,96	<0,001
F_2							
medial - kongruent	-61,6	18,8	621,3	-110,13	-13,12	-3,27	<0,01
kongruent - final	54,8	16,2	619,1	12,96	96,60	3,37	<0,01
F_3							
medial - kongruent	-58,9	17,2	620,9	-103,18	-14,72	-3,43	<0,01
kongruent - final	57,3	16,5	621,1	14,82	99,78	3,47	<0,01

Für F_3 ergibt sich folgendes Modell. Form, Dauer, Dialogzüge und Kontexte verbessern das Modell nicht signifikant. Abbildung 6.39 visualisiert die Modellvorhersagen für F_3 in Tabelle 6.37 iii.

```
1 | lmer(F3 ~ Sex + IP-Position + (1|VP), data ...)
```

Ausgehend von initialer Position in Intonationsphrasen werden Füllpartikeln in medialer (55,8 Hz) und finaler (54,2 Hz) Position mit einem signifikant höheren F_3 geäußert. *Post-hoc*-Vergleiche für F_2 (vgl. Tabelle 6.38) zeigen außerdem, dass F_3 in kongruenter Position signifikant höher ist als in medialer und finaler Position. Frauen zeigen einen signifikant höheren dritten Formanten als Männer, was aufgrund ihres oft kürzeren Ansatzrohres erwartbar ist.

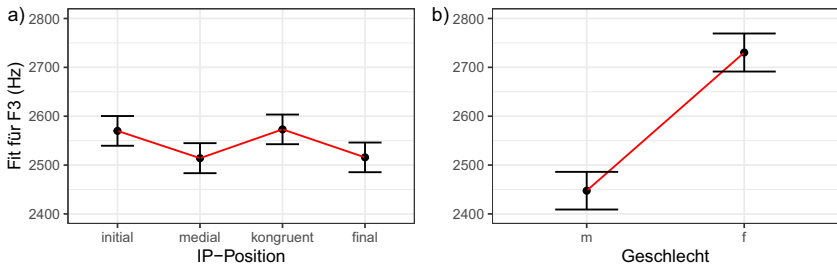


Abbildung 6.39: Modelleffekte von F_3 für a) IP-Position und b) Geschlecht in BeDiaCo_c.

Abschließend zeigt Abbildung 6.40 die Verteilung von Füllpartikeln im Vokalraum je IP-Position. Dabei werden initiale und kongruente sowie mediale und finale Positionen zusammengefasst, da sie ähnliche Effekte zeigen. Zur besseren Ansicht wird nur ein vergrößerter Ausschnitt dargestellt.

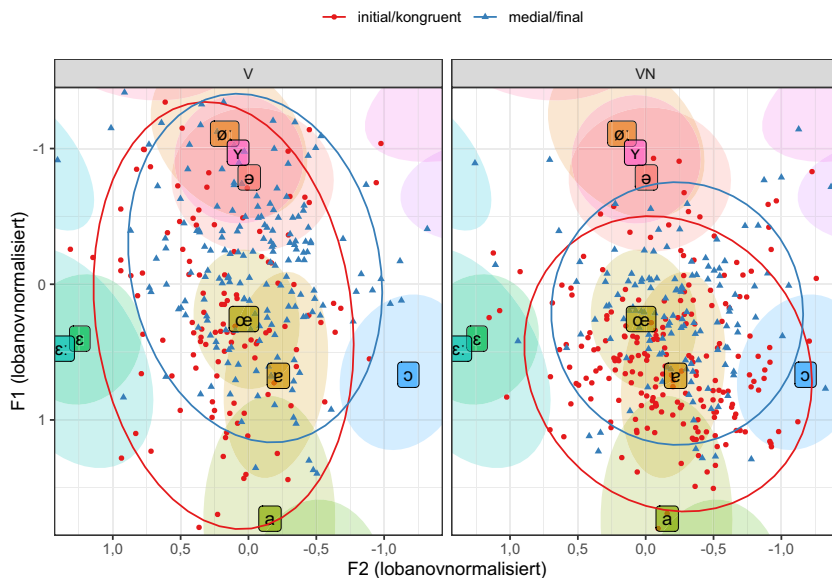


Abbildung 6.40: Intonationsphrasenpositionen für vokalische (V) und vokalisch-nasale (VN) Füllpartikeln (mittige Messung im Vokal, Datenellipsen kennzeichnen 95 % aller Messwerte) in einem Lobanov-normalisiertem Vokalraum aller Versuchspersonen in BeDiaCo_c mit Datenellipsen (zur besseren Sichtbarkeit 80 % aller Referenzvokalmittelwerte) für die Distributionen der Referenzvokale.

6.2.3.4 Phonationsart

Abbildung 6.41 zeigt die Distribution der Phonationsart über die Intonationsphrasenpositionen je Geschlecht. Signifikant unterscheidet sich nur der Gebrauch in initialer Position ($\chi^2 = 9,5$, $df = 1$, $p < 0,01$). Hier verwenden Frauen signifikant mehr glottalisierte Füllpartikeln als erwartet (15 statt 8) und signifikant weniger

modale Füllpartikeln (57 statt 64) – Männer hingegen signifikant weniger glottalisierte Füllpartikeln (3 statt 10) und signifikant mehr modale Füllpartikeln (80 statt 73) als erwartet. Allerdings ist zu beachten, dass die perzeptive Annotation der Phonationsart bei Männern aufgrund ihrer allgemein niedrigen Grundfrequenz ungleich schwieriger ist als bei Frauen. Mangels anderer Erklärung für die Unterschiede kann nur festgestellt werden, dass das Bild der Phonationsart bei Füllpartikeln inkonsistent bleibt.

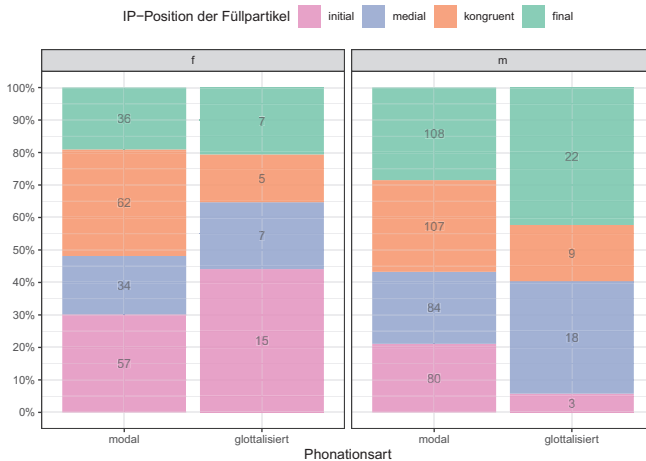


Abbildung 6.41: Absolute und relative Anteile der Vokalphonation je Intonationsphrasenposition je Geschlecht in BeDiaCo_c.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



7 Zusammenfassung und Vergleich

7.1 Hypothese der Formvariabilität

Tabelle 7.1 vergleicht die Ergebnisse von BeDiaCo_c und GECO-FP bezüglich der phonetischen Füllpartikelformen, die nicht durch ihre Mikro-, Meso- und Makrokontexte erklärt werden. Vielmehr zeigt sich, dass sie in großen Teilen allgemeinen linguistischen Prinzipien unterliegen (beispielsweise Formdistribution und Ausgleichskürzung), die im nächsten Kapitel diskutiert werden. Ein Teil der Formvariabilität lässt sich mit sprecherspezifischen Eigenschaften erklären (beispielsweise Grundfrequenz).

In beiden Korpora gibt es keine signifikante Korrelation zwischen Sprechgeschwindigkeit und Füllpartikeln je Minute. Schnelleres Sprechen führt also nicht zu mehr Füllpartikeln. In BeDiaCo_c besteht zwischen der Gesamttoken- und -silbenanzahl einer Versuchsperson sowie ihrer Füllpartikelanzahl eine positive Korrelation, was jedoch nicht sonderlich überraschend ist – wer mehr spricht, hat öfter die Gelegenheit zur Äußerung einer Füllpartikel. Dass diese Korrelation in GECO-FP nicht vorhanden ist, kann ein Effekt der Korpusgröße sein.

In BeDiaCo_c sind weniger glottale Füllpartikeln (4 %) vorhanden als in GECO-FP (20 %). Dieser Unterschied könnte eine Folge der Korpusgröße oder der unterschiedlichen Herkunft der Versuchspersonen sein. Denkbar ist auch, dass glottale Formen bei Männern (die nur in BeDiaCo_c vorkommen) nicht gut erkannt werden und daher nicht annotiert sind. Die Häufigkeit segmentaler Formen folgt sowohl für nicht-glottale als auch für glottale Füllpartikeln in beiden Korpora einer Zipfverteilung.

In BeDiaCo_c sind die häufigsten nicht-glottalen Formen solche ohne vorangehenden Plosiv (V, VN), gefolgt von solchen mit vorangehendem Plosiv, was in GECO-FP genau umgekehrt ist. In beiden Korpora folgen dann GV- und GVN-Formen, die vor dem Vokal längere Sequenzen glottaler Plosive zeigen. Orthographische *ähm*-Formen kommen insgesamt 1,2 mal so häufig vor wie *äh*-Formen, bei Männern etwas weni-

Tabelle 7.1: Vergleich der kontextfreien Eigenschaften von Füllpartikeln in beiden Korpora (✓ = repliziertes Ergebnis, ✗ = nicht repliziertes Ergebnis).

GECO-FP	BeDiaCo _c
Anzahl (s. 5.1)	(s. 6.1.1)
Keine Korrelation zw. Sprechgeschwindigkeit und FP/min	✓
Keine Korrelation zw. Tokenanzahl und FP-Anzahl	✗ Positive Korrelation
Keine Korrelation zw. Silbenanzahl und FP-Anzahl	✗ Positive Korrelation
20 % glottale, 80 % nicht-glottale FP	✗ 4 % glottale, 96 % nicht-glottale FP
Formen (s. 5.2.1)	(s. 6.1.2)
Zipfverteilung segmentaler Formen	✓
Rang absoluter Vorkommen: ?VN, ?V, VN, V, GVN, GV	✗ VN, V, ?VN, ?V, GVN, GV
Orthographisches <i>ähm</i> 1,3 mal häufiger als <i>äh</i>	✓ <i>ähm</i> 1,2 mal häufiger als <i>äh</i> (Frauen 1,3 mal, Männer 1,1 mal)
Dauer (s. 5.1.3)	(s. 6.1.3)
Positive Korrelation zw. Dauer und Segmentanzahl	✓
Ausgleichskürzung: Vokal in V länger als in VN	✓
V 1,5 σ/s , VN 2,3 σ , N 2,5 σ	✓ (N 2,7 σ)
Grundfrequenz (s. 5.1.4)	(s. 6.1.4)
Im Mittel 187–221 Hz	Frauen im Mittel 172–233 Hz, Männer 83–132 Hz
Signifikant niedriger als andere Äußerungen	✓
Glottalisierte Formen (Frauen) < 140 Hz	✗
Vokalqualität (s. 5.1.5)	(s. 6.1.5)
[œ v ə ε] und [œm em əm am]	✓
F ₁ : Höher (tiefer) in VN als in V (+66 Hz)	✓ (+62 Hz)
F ₂ : Kein Effekt	✗ Niedriger (posteriorer) in VN als in V (-111 Hz)
F ₃ : Kein Effekt	✓
Unter 100 ms <i>target undershoot</i> für Vokalhöhe (nur VN)	✓
Phonationsart (s. 5.1.6)	(s. 6.1.6)
Relation glottalisierter Vokale: V 20 %, VN 9 %	✓
Modale Vokale in V und VN sign. länger als glottalisierte	✗ Gilt nur für V

ger (1,1 mal), bei Frauen etwas mehr (1,3 mal). Hier gibt es keinen Unterschied zu GECO-FP. Die häufigsten glottalen Füllpartikeln sind perzeptiv deutlich voneinander abgesetzte Sequenzen glottaler Plosive und Sequenzen glottaler Plosive. Einzelne glottale Plosive, wie sie in GECO-FP vorhanden sind, kommen in BeDiaCo_c kaum vor. Nicht auszuschließen sind Falschnegative, die bei der Annotation nicht auffielen.

Die Dauer einer Füllpartikel korreliert in beiden Korpora positiv mit ihrer Segmentanzahl. Die normalisierte Gesamtdauer in Silben ist in beiden Korpora identisch für V- (1,5 σ/s) und VN-Formen (2,3 σ/s). VN-Formen sind immer signifikant länger als V-Formen (dies repliziert Clark & Fox Tree 2002; Leeuw 2007), nur N-Formen sind in

BeDiaCo_c marginal länger ($2,7\sigma/s$) als in GECO-FP ($2,5\sigma/s$). Vokale in vokalisch-nasalen Formen sind in beiden Korpora signifikant kürzer als in vokalischen Formen (Ausgleichskürzung in geschlossener Silbe, Katz 2012; Maddieson 1985).

Die Grundfrequenz von nicht-glottalen Füllpartikeln ist interindividuell verschieden und variiert (bei Frauen in der Mitte der Füllpartikel gemessen) im Durchschnitt von 172 bis 233 Hz, bei Männern von 83 bis 132 Hz. In beiden Korpora ist die Füllpartikelgrundfrequenz signifikant niedriger als der Durchschnitt aller anderen Äußerungen einer Versuchsperson (dies deckt sich mit den Ergebnissen von Braun & Rosin 2015), was jedoch auch auf viele andere Wörter zutreffen wird.

Die akustische Vokalqualität von V-Formen liegt im Bereich von [œ ɐ ə], die von VN-Formen liegt im Bereich von [œ ɐ ə a]. Insgesamt liegt der Vokal in den VN-Formen von signifikant tiefer (in GECO-FP 66 Hz, in BeDiaCo_c +62 Hz) im Vokalraum als V-Formen; anders als in GECO-FP ist er in BeDiaCo_c auch signifikant posteriorer (-111 Hz) im Vokalraum als der Vokal in V-Formen (was gegenläufige zu den Ergebnisse von Hughes et al. (2016) für das Englische ist). In beiden Korpora gibt es eine positive Korrelation zwischen Vokaldauer und Vokalhöhe im Bereich unter 100 ms, jedoch nur für VN-Formen; in diesem Bereich sind sie tiefer im Vokalraum als über 100 ms. Dies könnte einen *target undershoot* (Lindblom 1963) darstellen, wenn man davon ausgeht, dass das Ziel die ab 100 ms erreichte Höhe ist, und mit der oben beschriebenen Ausgleichskürzung zusammenhängen (je kürzer die gesamte VN-Form, desto kürzer der Vokal in VN-Formen).

Etwa 18 % der V-Füllpartikelvokale in BeDiaCo_c werden glottalisiert (in GECO-FP 20 %), aber nur 8,6 % der Vokale in VN-Formen (zufälligerweise exakt derselbe Wert in beiden Korpora). Modale Vokale sind in BeDiaCo_c anders als in GECO-FP nur in V-Formen signifikant länger als glottalisierte Vokale, nicht in VN-Formen. Die Grundfrequenz glottalisierter Vokale lässt sich für BeDiaCo_c anders als in GECO-FP nicht eindeutig zwei Gruppen über und unter 140 Hz zuordnen.

7.2 Form-Kontext-Hypothese

Im zweiten Teil der Studie werden im Rahmen der Form-Kontext-Hypothese die Analysen für die Mikro- und Makrokontexte aus GECO-FP repliziert und um eine dritte Perspektive – den Mesokontext der Intonationsphrasenposition – ergänzt. Tabelle 7.2 vergleicht die Ergebnisse beider Korpora. Die Form-Kontext-Hypothese

konnte in Teilen bestätigt werden.

Tabelle 7.2: Vergleich der kontextbedingten Eigenschaften von Füllpartikeln in beiden Korpora (✓ = repliziertes Ergebnis, ✗ = nicht repliziertes Ergebnis).

GECO-FP	BeDiaCo _c
Anzahl (s. 5.2.1)	(s. 6.2.1)
V und VN gleichverteilt zwischen Wörtern	✗
VN am häufigsten interpausal	✓
Form unabhängig von Lautkontext	✓
Am häufigsten in NEGATIVEN, UNKLAREN und KOMPLEXEN ANTWORTEN	✓
Dauer (s. 5.2.1.1, 5.2.2.1)	(s. 6.2.1.1, 6.2.2.1, 6.2.3.1)
Vor/zwischen Pausen länger als in anderen Kontexten	✗ gilt nur für Segment-FP-Atmung; Finale Länge; IP-Grenze erklärt längere Dauer
–	Graduelle Länge von VN an IP-Grenzen
Keine Korrelation mit der Länge nachfolgender Pausen	✓
In KOMPL. ANTWORTEN länger als in ERZÄHLUNGEN	✗
Grundfrequenz (s. 5.2.1.2, 5.2.2.2)	(s. 6.2.1.2, 6.2.2.2, 6.2.3.2)
Nach Atmung/Turn und vor Segment höher als intersegmental	✓ → vgl. IP
Tiefer als Umgebung (-2/+2 Token)	✓
Höher in KOMPL. ANTWORTEN nach Atmung/vor Segmenten als nach Segmenten/vor Atmung, in ERZÄHLUNGEN nicht	✗
–	IP-initial/-kongruent höher als IP-medial/-final
Vokalqualität (s. 5.2.1.3, 5.2.2.3)	(s. 6.2.1.3, 6.2.2.3, 6.2.3.3)
F ₁ größer in Turnpause-/Atmung-FP-Segment als intersegm.	✗ Konfundiert: F ₁ größer IP-initial/-kongruent als -medial/-final
Je länger der Vokal, desto größer F ₁	✓
F ₁ größer in nicht-intersegm. KOMPL. ANTWORTEN als in ERZÄHLUNGEN	✗ Größer in POS. ANTWORT als in anderen Dialogzügen
F ₁ -Verlauf in nicht-intersegm. KOMPL. ANTWORTEN tiefer im Vokalraum als in ERZÄHLUNGEN für V u. VN	✗ Nur für erste 39 % von V-Form
F ₂ größer in Turnpause-FP-Segment	✗ Konfundiert: F ₂ größer IP-initial/-kongruent als -medial/-final
Je länger der Vokal desto kleiner F ₂	✓
Kein Effekt für F ₃ i. Abh. v. Kontext	✗ F ₃ größer in Turnpause/Atmung-FP-Segment und Atmung-FP-Pause als intersegmental
Phonationsart (s. 5.2.1.4)	(s. 6.2.1.4, 6.2.3.4)
Glottalisierte Formen nur intersegmental oder in Turnpause-FP-Segment	✗ Auch in anderen Kontexten
–	IP-initial sign. häufiger Glottalisierung bei Frauen als bei Männern

Auf der Mikroebene zeigen V-Formen in BeDiaCo_c Präferenzen für Wort-FP-Wort-Kontexte, in GECO-FP jedoch nicht. In anderen Kontexte ergeben sich für die Distribution von V-, VN- und N-Formen keine Unterschiede zu GECO-FP. Die

Proportion von VN-Formen ist in beiden Korpora am höchsten in Pause-FP-Pause-Kontexten und nimmt in Wort-FP-Pause- und in Pause-FP-Wort-Kontexten ab. In den häufigsten sieben sequenziellen Kontexte Segment-FP-Segment, Atmung-FP-Segment, Segment-FP-Atmung, Segment-FP-Pause, Pause-FP-Segment, Atmung-FP-Pause und Turnpause-FP-Segment gibt es keine Unterschiede in der Distribution zwischen Männern und Frauen, und auch keinen signifikanten Unterschied zwischen BeDiaCo_c und GECO-FP. Dass Frauen mehr VN-Formen als Männer verwenden, kann also nicht auf die sieben häufigsten sequenziellen Kontexte zurückgeführt werden. Des Weiteren kann die Wahl einer Füllpartikelform wie in GECO-FP nicht auf den sie umgebenden Lautkontext zurückgeführt werden. Auf der Makroebene zählen (abgesehen von dem Dialogzug REDEBEREITSCHAFT) in BeDiaCo_c die NEGATIVEN, UNKLAREN und KOMPLEXEN ANTWORTEN zu den Dialogzügen mit den häufigsten Füllpartikeln pro Wort, was die Ergebnisse aus GECO-FP repliziert. Auf der Mesoebene der Intonationsphrase werden nicht signifikant mehr VN-Formen als V-Formen in IP-initialer Position geäußert, was den Ergebnissen von Clark & Fox Tree (2002) zur Position widerspricht.

Füllpartikeln vor oder zwischen Pausen sind – ohne Berücksichtigung der Intonationsphrasenposition – signifikant länger als in anderen Kontexten (vgl. Abschnitt 6.2.1.1). Dabei ist ihre Dauer unabhängig von der Dauer nachfolgender Pausen. Allerdings konfundiert die Intonationsphrasenposition den Effekt der sequenziellen Kontexte: Nur noch Segment-FP-Pause-Kontexte sind unter Berücksichtigung von IPs signifikant länger als intersegmentale Füllpartikeln. Bei der Einbeziehung von Intonationsphrasen wird die längere Dauer von Füllpartikeln mit rechtsstehenden Pausen (Stille, Atmung oder mit phonetischem Reflex) darüber erklärt, dass sie fast nur in finaler IP-Position stehen (vgl. Abbildung 6.33 a).¹⁴³

In kongruenter und finaler IP-Position sind Füllpartikeln signifikant länger als initial oder medial. Dies widerspricht zwar den Ergebnissen von Swerts (1998), der für niederländische Füllpartikeln längere Dauern in initialer Position beschreibt. Allerdings ist der dort umgekehrt beschriebene Effekt wahrscheinlich eine Folge seiner Umkategorisierung final auftretender Füllpartikeln als initiale Füllpartikeln.¹⁴⁴ Das

¹⁴³Die in BeDiaCo_c hinzugefügte Intonationsphrasenebene erklärt von den drei untersuchten Kontexten außerdem die meiste Varianz der Füllpartikeldauer, nämlich 49 % ($R_c^2 = 0,49$), gegenüber 46 % in sequenziellen Kontexten und 35 % in Dialogzügen.

¹⁴⁴„If an FP occurred at a boundary, it was analyzed as being at the beginning of the next phrase, rather than at the end of the previous phrase.“ (Swerts 1998: 48).

prosodische Phänomen, dass Segmente an Phrasengrenzen in finaler Position länger sind als in medialer oder initialer Position ist als ‚finale Längung‘ einschlägig beschrieben worden (für einen Überblick vgl. Cho 2015; Fletcher 2010; Krivokapić 2014; fürs Deutsche Belz et al. [eingereicht]; Peters et al. 2005). Dialogzüge haben keinen Effekt auf die Füllpartikeldauer.

Die Grundfrequenz von Füllpartikeln ist zwischen Turnpause oder Atmung und einem nachfolgenden Segment wie in GECO-FP signifikant höher als in intersegmentalen Kontexten. Die Intonationsphrasenannotation hat gezeigt, dass diese Kontexte hauptsächlich IP-initial auftreten (vgl. Abbildung 6.33 a). In intersegmentaler Position zwischen zwei Token zur linken und zwei zur rechten zeigen Füllpartikeln eine signifikant tiefere Grundfrequenz als ihre Umgebung, was das Ergebnis aus GECO-FP sowie die Ergebnisse von Batliner et al. (1995) repliziert. Anders als in GECO-FP haben Dialogzüge keinen Effekt auf die Grundfrequenz. Die Analyse der Intonationsphrasenposition zeigt, dass Füllpartikeln in initialer und kongruenter Position eine höhere f_0 als in medialer und finaler Position haben, was der prosodischen Regelmäßigkeit der Grundfrequenzdeklination entspricht (Fuchs et al. 2015; Peters et al. 2005).

Der Einfluss sequenzieller Kontexte auf den Füllpartikelvokal wird für BeDiaCo v. 1 in Teilen bestätigt (vgl. Abschnitt 6.2.1.3). Zwischen Wörter und Pausen ist F_1 im Vergleich zu interpausaler Position erhöht. Der Effekt der sequenziellen Kontexte wird jedoch konfundiert durch die IP-initiale Lage mancher Kontexte. Initial und in kongruenter Position sind die Vokale von Füllpartikeln tiefer und anteriorer, medial und final hingegen zentraler und posteriorer im Vokalraum gelegen. Die Formantmittelwerte sind nicht abhängig von Dialogzügen. Die Auswertung der Formanttrajektorie über die gesamte Vokaldauer konnte nur für V-Formen (nicht für VN-Formen wie in GECO-FP) und auch nur für die vorderen 39% der Trajektorie eine tiefere Lage für nicht-intersegmentale Vokale in KOMPLEXEN ANTWORTEN replizieren.

Für die Phonationsart sind die Ergebnisse sehr unterschiedlich und kritisch zu betrachten. Glottalisierte Vokalformen stehen in BeDiaCo_c anders als in GECO-FP auch in anderen Kontexten als nur intersegmental oder zwischen Turn und Segment. Auf der Mesoebene finden sich IP-initial signifikant mehr glottalisierte Formen bei Frauen als bei Männern. Obwohl auch andere Studien IP-initial häufiger Glottalisierung finden (Dilley et al. 1996), ist dies hier nicht konsistent der Fall. Es ist möglich,

dass die perzeptive Annotation die Ergebnisse verfälscht hat, da beispielsweise eine niedrige Grundfrequenz bei Männern als Glottalisierung wahrgenommen wird (Pierrehumbert & Frisch 1997). Andere Maße wie Jitter, Shimmer, oder die Amplitudendifferenz der ersten beiden Harmonischen sind daher in zukünftigen Studien zu prüfen.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





8 Diskussion

Im ersten Abschnitt dieses Kapitels diskutiere ich die gefundene kontextfreie Variation von Füllpartikeln. Im Anschluss daran werden die kontextbedingten Ergebnisse für die sequenziellen Kontexte und die Dialogstruktur eingeordnet, wonach diskutiert wird, wie Teile der Kontextbedingtheit durch die prosodische Phrasierung erklärt werden können.

8.1 Kontextfreie Formvariation

Der Begriff der ‚kontextfreien Formvariation‘ ist der Versuch, eine Bezeichnung für jene Auftretensformen und -häufigkeiten zu wählen, die (zumindest in dieser Arbeit) nicht nachweisbar über die untersuchten Kontexte erklärt werden können. Die vier häufig in der Literatur beschriebenen nicht-glottalen Formvariationen deutscher Füllpartikeln werden in dieser Arbeit bestätigt, nämlich das Vorkommen von i) V-, VN- und anderen Formen; ii) schwankender Vokalqualität; iii) Glottalplosivepenthese; und iv) Vokallängenvariation (vgl. Tabelle 2.3). Darüber hinaus existieren jedoch sehr viel mehr weitere Formen. Dies legt den Schluss nahe, dass die akustische Realisierung einer Füllpartikel mit einem Formkontinuum beschrieben werden kann, welches in seiner geringsten Ausdehnung ein Segment enthält (einen glottalen Plosiv (G), einen Vokal (V) oder einen Nasal (N); auch andere Konsonanten (C) sind möglich). Durch Hinzufügen weiterer Segmente aus diesem Set können Segmentketten wie GV, GVN, VN, GVNC, GNVN etc. gebildet werden.

Sprecherspezifische Unterschiede sind hauptsächlich frequenzbasiert. So äußern Sprecher m15 in BeDiaCo_c und Sprecherin K in GE_{CO}-FP sehr viel mehr Füllpartikeln als die anderen Versuchspersonen. Es ist davon auszugehen, dass dies idiosynkratische Gewohnheiten abbildet. Sprecherbasierte Vorlieben existieren auch für einzelne Formausprägungen. Geschlechtsspezifisch ist das größere VN-zu-V-Verhältnis bei Frauen und die erhöhte Füllpartikelrate bei Männern. Das höhere VN-zu-V-

Verhältnis insgesamt wird von Wieling et al. (2016) und Fruehwald (2016) als sozio-linguistischer Effekt beschrieben, könnte jedoch auch eine Folge der Zipfverteilung sein (vgl. nächster Absatz) – zudem macht keine der bisherigen Studien zu VN/V-Verhältnissen deutlich, welche Gründe für die Annahme einer Gleichverteilung beider Formen V und VN sprechen sollten (Jessen 2012; Jurafsky et al. 2009; Laserna et al. 2014; McDougall & Duckworth 2018; Wieling et al. 2016). Für die kontinuierlichen Parameter hat sich (außer für die Grundfrequenz, die bei Frauen natürlicherweise höher ist) kein Geschlechterunterschied gezeigt.

Die Häufigkeit verschiedener Formtypen folgt sowohl für nicht-glottale als auch für glottale Füllpartikeln annähernd der Zipfverteilung (Zipf 1949), d. h. die Häufigkeit bestimmter Formtypen nimmt in harmonischer Weise ab (Zeldes 2012: 77). Die Zipf-Verteilung wird als universelles Sprachprinzip besonders als Produktivitätsmerkmal im Kontext von Type-Token-Distributionen (ebd.) und zur Beschreibung von der Häufigkeit und der Länge von Wörtern verwendet (Bentz & Ferrer-i-Cancho 2016). Das Vorkommen glottaler Plosive zu Beginn von Vokalen kann also – obschon diese Formen in beiden Korpora verschieden Häufigkeitsränge einnehmen – phonologisch mit der quasi-obligatorischen Glottalplosivepenthesen erklärt werden, wie sie auch für andere deutsche Wörter gilt, die einen betonten Vokal in Silben mit unbesetztem Onset enthalten (Hall 2011). Die Verteilung der Segmentformen und ihre Häufigkeit scheint somit keine Besonderheit darzustellen, sondern stimmt mit Erfahrungswerten aus der quantitativen Linguistik überein. Füllpartikeln verhalten sich somit wie anderes sprachliches Material auch.

Auch die positive Korrelation zwischen der Gesamtdauer einer Füllpartikel und der Anzahl der in ihr enthaltenen Segmente ist in anderem sprachlichen Material zu beobachten. VN-Formen sind signifikant länger als V-Formen. Die erstmals für das Deutsche über Versuchspersonen normalisierte Dauer in Silben beträgt in beiden Korpora konsistent im Mittel $1,5\sigma$ für V- und $2,3\sigma$ für VN-Formen. Auch die Dauer der einzelnen Segmente innerhalb einer Füllpartikel variiert systematisch. So sind Vokale in VN-Formen kürzer als in V-Formen, was eine Folge der geschlossenen Silbe darstellt und unter dem Begriff der Simplexausgleichskürzung bekannt ist, die auch in anderen Sprachen beobachtet wird (Katz 2012; Maddieson 1985). Damit ist die Beobachtung dieser Formvariation völlig regelhaft.

Die Vokalqualität von Füllpartikeln liegt hauptsächlich im tiefen vorderen und

mittleren zentralen Teil des Vokalraumes und umfasst auch die Areale, in denen Schwa [ə] und halbtiefes Schwa [ɐ] realisiert werden, obwohl diese sonst im Deutschen nicht in betonten Silben vorkommen und nicht betonbar sind (Barry 1995). Allerdings werden auch einsilbige Funktionswörter im Deutschen mit Schwa produziert, wenn sie nicht in satzbetonter Position stehen. Phonologisch wird dafür argumentiert, dass [ɐ] /ər/ zugrunde liegt, während [ə] ein nicht-phonologisches, myofunktionales Relaxationsziel darstellt (ebd.: 234). In dieser Hinsicht scheinen Füllpartikeln also eine Ausnahme darzustellen, da keinerlei Evidenz vorliegt, dass einer Form [ʊm] eine phonologische Form /əm/ zugrunde liegt.

Die Qualität des Vokals ist in VN-Formen systematisch tiefer als in V-Formen. Im Rahmen der artikulatorischen Phonologie (Browman & Goldstein 1992) diskutiere ich mögliche Ursachen für diese Beobachtung.

i) Ein Verdachtsmoment liegt auf der koartikulatorisch mit der artikulatorischen Vokalgeste überlappenden Geste der Velumabsenkung zur anschließenden Produktion des Nasals. Dass diese zeitliche Überlappung beider Gesten zu einer Erhöhung des ersten Formanten führt, ist allerdings in bisherigen Studien nicht nachgewiesen worden (vgl. Styler 2017). Die Erklärung über *target undershoots* (Lindblom 1963), die bei kurzen Vokalen dazu führen, dass das stipulierte Ziel des Vokals nicht erreicht werden kann, scheidet aus, da die Dauer in den Modellen berücksichtigt wurde.¹⁴⁵

ii) Das Verdachtsmoment liegt dennoch weiterhin auf dem folgenden bilabialen Nasal und seinem koartikulatorischen Einfluss, wenn es auch jetzt aus einer anderen Perspektive beleuchtet werden muss. Der bilabiale Verschluss „must be achievable with context-sensitive contributions of the jaw, the upper lip, and the lower lip during everyday coarticulated speech“ (Fowler & Saltzman 1993: 179). Koartikulatorisch plausibel wäre ein antizipatorischer Effekt des bilabialen Verschlusses, so dass der Vokal in einer zentralen Position geäußert wird. Das Gegenteil wird beobachtet: Akustisch erreicht der Vokal in VN-Formen höhere F_1 -Werte (wird also tiefer) als in V-Formen. Nun ist offen, ob dieser erhöhte F_1 eine Folge davon ist, dass der Kiefer weiter geöffnet wird als bei V-Vokalen. Ein Grund für diese Annahme bestünde in einer Sonoritätserhöhung des Vokals, um diesen deutlicher vom folgenden Nasal abzugrenzen. So wird (zumindest für akzentuierte Silben) der Mund weiter geöffnet, um den Widerstand in der oralen Kavität zu verringern und eine

¹⁴⁵Unter 100 ms kommt es zwar zu *target undershoots*, diese erklären aber nicht den *allgemeinen* Unterschied in der Vokalqualität von V- und VN-Formen.

größere Energieausstrahlung an den Lippen zu erlauben (Harrington et al. 2000: 43), möglicherweise damit der Vokal nicht zu sehr reduziert werden kann und so noch von der N-Form zu unterscheiden ist. Allerdings könnte der tiefere F_1 auch daher rühren, dass VN-Formen ein von der V-Form distinktes akustisch-artikulatorischen Ziels besitzen, was dem Effekt erhöhter Vokalsonorität jedoch nicht widerspricht.

Da (i) keine hinreichende Erklärung bietet, ist die letztliche Konsequenz aus den Überlegungen aus (ii), dass in V- und VN-Formen zwei verschiedene Vokalziele erreicht werden. Dies würde zur Hypothese passen, dass „filled pauses [...] have targets of their own“ (Gick et al. 2004: 231) – und zwar in der Spezifizierung, dass dies nicht nur für Füllpartikeln im Allgemeinen gilt, sondern dass VN-Formen aufgrund ihres folgenden bilabialen Nasals andere Vokaltargets haben als V-Formen. Offen bleibt, ob dieses Verhalten auch in lexikalischen Wörtern zu beobachten ist und somit eine komplementär verteilte Allophonie vorliegt.

Einschränkend gilt natürlich, dass die vorliegende Studie nur akustische Daten betrachtet, und die Beziehung zwischen artikulatorischer Geste und akustischer Messung keine 1-zu-1-, sondern vielmehr eine n-zu-m-Beziehung ist. Ein Geschlechtereffekt scheidet als Erklärung aus – Frauen haben zwar im Allgemeinen einen größeren akustischen Vokalraum als Männer (Weirich & Simpson 2014) und daher mehr Möglichkeit zur Variation, doch der Effekt ist bei beiden Geschlechtern zu beobachten.

Zusammenfassend gibt es also starke Hinweise auf verschiedene Vokalziele in V- und VN-Formen, sowie – wenn man einen Schritt zurücktritt – für die Existenz eines Formkontinuums zwischen einem glottalen Plosiv und multisegmentalen Lautclustern. Welche Segmente und Segmentzusammenschlüsse sich letztlich als Füllpartikeln definieren lassen, ist sowohl eine Frage ihres Gebrauchs in einer Einzelsprache als auch ihres funktionalen Status. Da in dieser Arbeit die Funktion komplett ausgeblendet wird, bleibt zu fragen: Rechtfertigt die vorliegende Analyse akustischer Variation nun die phänomenale Einordnung der produzierten Entitäten als Partikeln – und damit als Wörter? Ähnlichkeiten zu Wörtern ergeben sich daraus, dass sowohl glottale als auch nicht-glottale Füllpartikeln annähernd der Zipf-Verteilung folgen, wie sie auch für andere Wörter des Deutschen beobachtet wird. Auch zeigen VN-Formen Effekte, wie sie allgemein in Silben zu beobachten sind, die aus einem Vokal und einem Konsonanten bestehen. Nicht-glottale Formen können zudem als eigenständige Intonationsphrase gebraucht werden. Somit kann zumindest für die

nicht-glottalen Formen ein phonologischer Wortstatus angenommen werden. Für die glottalen Formen gilt aus phonetischer Perspektive immerhin, dass sie eigenständig segmentierbar sind. Wenn man also den Status eines phonologischen Wortes für nicht-glottale Füllpartikeln annehmen kann, so sollten sie auch in eine grammatische Beschreibung integriert werden. Terminologisch ist der Begriff der *Füllpartikeln* dem der *gefüllten Pause* vorzuziehen, da letzterer eher auf eine prosodische Betrachtung zielt, Füllpartikeln aber Eigenschaften phonologischer Wörter zeigen und zudem nicht flektierbar sind, was ein syntaktisches Kriterium ist. Mithilfe dieser Eigenschaft wird also ein erster Versuch unternommen, diesen Entitäten eine Wortart zuzuweisen, was jedoch nur für *nicht-glottalen* Füllpartikeln gelingt. Offen bleibt, welche Kriterien die glottalen Füllpartikeln aufweisen müssten (Länge, Salienz, Betonbarkeit), um ihnen eine Wortart zuzuweisen. Auch die genauere syntaktische Distribution und die funktionalen Eigenschaften glottaler und nicht-glottaler Füllpartikeln sind ein offenes Desiderat. Im nächsten Abschnitt 8.2.2 werden wir abschließend nochmals sehen, dass sich nicht-glottale Füllpartikeln auch kontextuell ähnlich zu Wörtern verhalten, da sie sie regelhaft prosodisch integriert werden.

8.2 Kontextbedingte Formvariation

8.2.1 Sequenzieller Kontext und Dialogstruktur

Die Hypothese der kontextabhängigen Formvariabilität wurde anhand der drei granular verschiedenen Faktoren sequenzieller Kontext, Dialogstruktur und Intonationsphrase untersucht. Für die Faktoren des sequenziellen Kontextes und der Dialogstruktur sind die Ergebnisse nicht eindeutig. Dies ist auch eine Folge davon, dass sie eine beträchtliche Zahl von Faktorausprägungen enthalten, was besonders bei den sequenziellen Kontexten dazu führt, dass ein Teil dieser Ausprägungen aufgrund ihrer zu geringen Häufigkeit nur unvollständig ausgewertet werden kann.

Die Ergebnisse für die sequenziellen Kontexte legen dennoch nahe, dass die verwendete Füllpartikelform in Teilen kontextsensibel ist, also von ihrem direkten adjazenten Kontext abhängt. Unstrittig ist, dass in interpausaler Position am häufigsten VN-Formen und in intersegmentaler Position am häufigsten V-Formen vorkommen (ähnlich zu Jessen 2012; Leeuw 2007). Die konkrete lautliche Realisierung eines Segments hat jedoch keinen Einfluss auf die nachfolgende Füllpartikel. Die Dauer einer

Füllpartikel ist (unabhängig von ihrer Form) in interpausaler Position signifikant länger als in intersegmentaler Position (auch hier ähnlich zu Jessen 2012). Zwischen der Länge der nachfolgenden Pause und der Länge der Füllpartikeln konnte keine signifikante Korrelation nachgewiesen werden – damit widersprechen die Daten anderen Dialogstudien (Clark & Fox Tree 2002; Rose 2015; Smith & Clark 1993). Es stellt sich die Frage, ob dies eine Folge der verwendeten Methode ist; so schätzen Clark & Fox Tree die Pausenlänge und Rose findet einen reliablen Unterschied nur für die gesamte Länge von V+Pause vs. VN+Pause. Andererseits liegt möglicherweise ein sprachspezifischer Unterschied zwischen Englisch und Deutsch vor, da auch O’Connell & Kowal (2005) in dialogischen (Interview-)Daten keine Evidenz für eine Korrelation zwischen der Länge einer Füllpartikel und der Länge ihrer nachfolgenden Pause finden.

Eine zusätzliche Schwierigkeit besteht darin, dass die sequenziellen Kontexte in die prosodische Hierarchie eingebunden sind – die Abfolge Pause-FP-Wort kann beispielsweise sowohl am Anfang als auch in der Mitte einer Intonationsphrase stehen. Eine mögliche Folge davon ist, dass die asymmetrischen Kontexte Pause-FP-Wort und Wort-FP-Pause uneinheitliche Muster zeigen. Die uneinheitlichen Ergebnisse zur Korrelation von Füllpartikellänge und Länge der nachfolgenden Pause könnten also auch damit erklärt werden, dass die obigen Studien verschiedene untere Schwellen für Pausen definieren – dann würde die Korrelation aufgrund der verzerrten Datenauswertung just die Korrelation zwischen Pausenlänge und Stärke einer prosodischen Grenze wiedergeben (vgl. Fletcher 2010). Hier besteht weiterer Forschungsbedarf.

Für Grundfrequenz und die Vokalhöhe und -lage werden zwar Effekte in Abhängigkeit von sequenziellen Mustern gefunden, diese lassen sich aber – kategorienreduzierend im Ockham’schen Sinne – besser durch die prosodische Hierarchie erklären (s. Abschnitt 8.2.2). Unabhängig von der dort erklärten allgemeinen Grundfrequenzdeklination konnte jedoch für Füllpartikeln in Fünfgrammen eine lokale Absenkung der Grundfrequenz mit nachfolgendem Anstieg beobachtet werden (dies repliziert Batliner et al. 1995; Klug 2013).

Die Ergebnisse zur Dialogstruktur sind nicht schlüssig. Zwar enthalten beantwortende Dialogzüge in beiden Korpora den größten Anteil an Füllpartikeln (bis auf POSITIVE ANTWORTEN in GECO-FP), was nahelegt, dass Füllpartikeln tat-

sächlich eher in Antworten als in anderen Dialogzügen stehen. Abgesehen von dieser frequenzbasierten Analyse sind allerdings keine konsistenten Auswirkungen auf die Füllpartikelform zu beobachten. Unterschiedliche Längeneffekte in KOMPLEXEN ANTWORTEN und ERZÄHLUNGEN konnten nicht repliziert werden, genauso wenig wie Grundfrequenzunterschiede oder eine Differenzierung der Vokalqualität. Zwar sind die Annotationsschemata der Dialogstruktur und ihre Segmentierung in beiden Korpora leicht abweichend, dies alleine erklärt die abweichenden Ergebnisse jedoch nicht, da trotz der Unterschiede in der zugewiesenen Länge von beispielsweise einer KOMPLEXEN ANTWORT eine dort enthaltene Füllpartikel eben genau dies auch in Annotationen mit abweichender Segmentierung bleibt, nämlich Teil einer KOMPLEXEN ANTWORT. Offen bleibt also, ob die Dialogstruktur als Konstituente höherer Ordnung, die mehrere Intonationsphrasen oder Äußerungen enthalten kann, auf phonetische Parameter von Füllpartikeln einwirken kann. Die vorliegenden Ergebnisse können dies nicht hinreichend bestätigen – für eine Dialogfunktion, die aus der unterschiedlichen lautlichen Realisierung von Füllpartikeln abgeleitet wird, liegt keine schlüssige Evidenz vor.

Andere Gründe für die nichtreplizierten Ergebnisse können im sprachlichen Hintergrund der Versuchspersonen liegen. Während die Versuchspersonen in GECO-FP überwiegend einen süddeutschen sprachlichen Hintergrund besitzen, entstammt die sprachliche Prägung der in BeDiaCo_c enthaltenen Versuchspersonen überwiegend aus der Mitte und dem Norden Deutschlands. Es müsste also untersucht werden, ob die vorliegenden Ergebnisse mit dialektal kontrollierten Versuchspersonen (entweder aus Baden-Württemberg oder aus der Mitte und dem Norden Deutschlands) repliziert werden können.

8.2.2 Einfluss prosodischer Phrasierung

Ein zentrales Ergebnis dieser Arbeit ist die Erkenntnis, dass die Form einer Füllpartikel mithilfe ihrer Position in einer prosodischen Phrase (nämlich der Intonationsphrase) erklärt werden kann. Prosodie kommt zwei wichtige Funktionen zu: Phrasierung beziehungsweise „grouping function“ (Chunking, Einteilung in größere Einheiten) und Prominenzmarkierung (Akzent, Hervorhebung) (Keating 2006). Prosodische Struktur wird häufig mithilfe weiterer hierarchischen Ebenen über der Silbendomäne modelliert. Silben gruppieren sich demnach zu phonologischen Wör-

tern, diese zu intermediären Phrasen und diese wiederum zu Intonationsphrasen (Beckman & Pierrehumbert 1986). Nach Swerts (1998) markieren Füllpartikeln prosodische Phrasen. Hier argumentiere ich dafür, die Perspektive in ihr Gegenteil zu kehren: Prosodische Phrasen markieren die phonetische Form von Füllpartikeln. Im Folgenden diskutiere ich diese Markierung bezüglich a) Vokalqualität, b) Dauer und c) Grundfrequenz.

a) In domäneninitialer Position kommt es insbesondere für Intonationsphrasen häufig zur Stärkung des Konsonanten, was beispielsweise artikulatorisch mit mehr palatolingualem Kontakt einhergeht (für Literaturzusammenfassungen vgl. Cho 2015, 2016). Zu akustischen Korrelaten vokalinitialer Stärkung gibt es jedoch nur wenig Forschung. Im Französischen werden die tiefen und untermittelhohen Vokale [a œ ø] IP-initial mit höherem F_1 , die vorderen Vokale [e ε] mit höherem F_2 geäußert, die Dauer der Vokale aber bleibt davon unberührt (Georgeton & Fougeron 2014: 9). Da sich auch die Vokale in Füllpartikeln in der hier vorliegenden Studie in untermittelhoher und eher vorderer Position befinden, liegt der Schluss nahe, dass es sich bei den beobachteten höheren F_1 - und F_2 -Werten in initialen und kongruenten IP-Positionen um eine initiale Stärkung dieser Vokale in phraseninitialer Position handelt. Diese Stärkung korreliert mit höheren Formanten und könnte bedeuten, dass der Kiefer offener ist und die Vokallage eher anterior realisiert wird. Sie wirkt sich in gleichem Maße auf beide Formen – V und VN – aus. Wenn die Vokale in Füllpartikeln also tatsächlich ein festgelegtes artikulatorisches und akustisches Ziel oder einen Zielbereich haben sollten (vgl. Gick et al. 2004), so ist für beide Formen davon auszugehen, dass hier zumindest ein akustischer *target overshoot* vorliegt, das Ziel also übersteuert wird. Dieser Effekt ist unabhängig von der oben diskutierten kontextfreien Vokalvariation in V- und VN-Formen.

b) In domänenfinaler Position werden Segmente gelängt (für einen Überblick vgl. Cho 2015; Fletcher 2010; Krivokapić 2014; für das Deutsche Belz et al. [eingereicht]; Kohler 1983; Peters et al. 2005). Für die Gesamtdauer der Füllpartikel konnte dieser Effekt bestätigt werden. Weiterhin gilt, dass Segmente direkt an der Phrasengrenze stärker gelängt werden als Segmente, die weiter entfernt stehen (Byrd et al. 2006), was in der Artikulatorischen Phonologie (Browman & Goldstein 1992) mithilfe einer prosodischen Geste modelliert wird (π -Geste), die an prosodischen Grenzen koaktiv ist (Byrd & Saltzman 2003) und an hierarchisch höhere Grenzen einen größeren Ef-

fekt aufweist. Tatsächlich wird der Nasal in VN-Formen in IP-finaler Position stärker gedehnt als der vorangehende Vokal, der von der Grenze weiter entfernt ist. Wenn VN-Formen eine eigenständige IP bilden, ist der Vokal signifikant länger als in finaler Position. Dies könnte eine Folge davon sein, dass in solcher kongruenter Position die an beiden Enden wirkende π -Geste und der Phrasenakzent additiv zusammenwirken. Damit wird hier erstmals gezeigt, dass Füllpartikeln den Eigenschaften domänenfinaler Längung unterliegt. Darüber hinaus lässt sich diese Hypothese für Füllpartikeln in allen Sprachen aufstellen, bei denen finale Längung zu beobachten ist.

c) Über die ganze Intonationsphrase hinweg sinkt die Grundfrequenz langsam ab und wird an Phrasengrenzen wieder zurückgesetzt (Fuchs et al. 2015). Tatsächlich konnte in dieser Studie gezeigt werden, dass Füllpartikeln in initialer Position mit höherer und in finaler Position mit niedrigerer Grundfrequenz geäußert werden, was dem allgemeinen Prinzip der Grundfrequenzdeklination über die Intonationsphrase hinweg entspricht. Daneben ist die Grundfrequenz von Füllpartikeln in intersegmentaler Position geringer als ihre direkte Umgebung und repliziert damit die Studie von Shriberg & Lickley (1993) im Englischen.

Die Realisierung von Füllpartikeln folgt also prosodischen Prinzipien. Obwohl spontane gesprochene Sprache weniger kontrolliert ist als Lesesprache, hat die vorliegende Studie somit gezeigt, dass durch die Annotation und Auswertung des Kontextes Regelmäßigkeiten aufgezeigt werden können. Abbildung 8.1 fasst den Einfluss der prosodischen Phrasierung auf die akustische Variation von Füllpartikeln in einem qualitativen Modell übersichtlich zusammen.

In Abbildung 8.1 ist veranschaulicht, wie sich initiale, mediale und finale Intonationsphrasenpositionen auf die akustische Form einer Füllpartikel auswirken. Das Modell ist stark vereinfacht und berücksichtigt keine Füllpartikeln, die eine eigenständige Intonationsphrase bilden. Eine längere Dauer (in Millisekunden) wird in finaler Position erwartet (finale Längung). Die Grundfrequenz (f_0) ist in initialer Position erhöht, sinkt medial langsam ab und ist in finaler Position im Allgemeinen geringer als in initialer und medialer Position (allgemeine Grundfrequenzdeklination). Die Vokalqualität ist in initialer Position für F_1 und F_2 erhöht (initiale Stärkung), d. h. der Vokal liegt tiefer und anteriorer im Vokalraum als in medialer und finaler Position.

Die Ergebnisse zeigen, dass zumindest ein Teil der Formvariation von Füllparti-

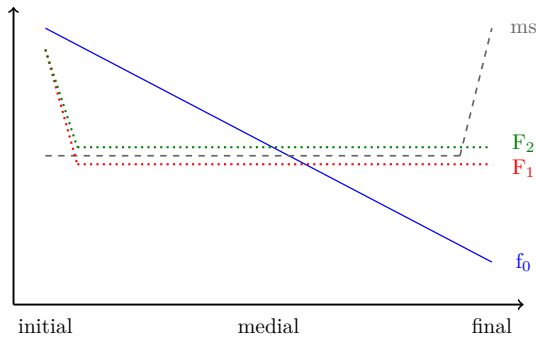


Abbildung 8.1: Qualitatives Modell der Parameter Dauer (ms), Grundfrequenz (f_0) und Formanten (F_1 , F_2) für Füllpartikelvokale und ihre Position in der Intonationsphrase.

keln auf ihren prosodischen Kontext zurückgeführt werden kann. Füllpartikeln sind in dieser Hinsicht mindestens genauso variabel, aber auch genauso vorhersehbar in ihrer akustischen Realisierung wie andere deutsche Wörter. Unabhängig von ihrer pragmatischen oder interaktionalen Funktion markieren sie mit ihren akustischen Eigenschaften somit in erster Linie exakt die prosodische Funktion, die ihnen über die prosodische Hierarchie an der Stelle ihres Auftretens zugewiesen wird.

8.3 Fazit und Ausblick

Diese Arbeit leistet einerseits einen ausführlichen Beitrag zur gebrauchsbasierten Erfassung der akustischen Füllpartikelformen als auch zur Beschreibung von kontextuellen Füllpartikelverwendungsmustern im Deutschen, ohne ihnen *a priori* einen funktionalen Status zuzuweisen. Die akustische Variabilität von Füllpartikeln in deutschen spontansprachlichen aufgabenfreien Dialogen ist sowohl auf kontextfreie Variation (besonders für die Unterscheidung zwischen V- und VN-Formen) als auch auf kontextbedingte Variation zurückzuführen. Die kontextbedingten Eigenschaften von Füllpartikeln erklären sich wiederum in Teilen über ihre sequenziellen Kontexte, während für die Dialogstruktur kein konsistenter Einfluss nachgewiesen wird. Besonders deutlich tritt der Einfluss des Kontextes für prosodische Kontexte mit Bezug auf Intonationsphrasen zutage und erklärt dort temporale (finale Längung), spatiale (initiale Stärkung) und suprasegmentale (f_0 -Deklinaton) Variationen. Es ist

somit davon auszugehen, dass Füllpartikeln jene phonetischen Eigenschaften tragen, welche die prosodische Phrasierung ihnen an der Position ihres Auftretens zuweist. Die phonetische Formenvielfalt und -variation von Füllpartikeln in deutschen spontansprachlichen Dialogen ist demnach nicht arbiträr, sondern in Teilen von ihrem Kontext, insbesondere aber von ihrer prosodischen Position abhängig.

Für zukünftige Studien ergeben sich auf dieser Basis interessante Forschungsdesiderate. So ist offen, ob alle gefundenen Formunterschiede perzeptiv eindeutig wahrnehmbar sind oder ob sich einige als salienter als andere erweisen, was wiederum die Frage aufwirft, ob damit funktionale Unterschiede angezeigt werden (können). Unabhängig davon können nun die hier ermittelten Form-Kontext-Muster auf ihre funktionalen Eigenschaften im Dialog hin untersucht werden. Bezüglich der unterschiedlichen Vokalziele für V- und VN-Formen bleibt zu klären, ob die bilabiale Schließgeste zum Nasal hin bei zentralen Vokalen eine artikulatorische Absenkung des Vokals zur Folge hat. Ein Vergleich mit elizitierbaren lexikalischen oder gelesenen Nicht-Wörtern könnte hier weiterhelfen. Eine weiteres offenes Forschungsdesiderat besteht im Vergleich der vorliegenden Ergebnisse mit Daten aus anderen gesprochen sprachlichen Registern, beispielsweise aufgabenbasierten Dialogen, um möglicherweise vorliegende registerspezifische Effekte modellieren zu können.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Literatur

- Acton, Eric K. (2011): On Gender Differences in the Distribution of um and uh. *University of Pennsylvania Working Papers in Linguistics* 17 (2). Art. 2.
- Adank, Patti; Roel Smits & Roeland van Hout (2004): A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116 (5). 3099–3107. DOI: 10.1121/1.1795335.
- Akaike, Hirotugu (1974): A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* AC-19 (6). 716–723.
- Albin, Aaron (2014): PraatR. An architecture for controlling the phonetics software "Praat" with the R programming language. *Journal of the Acoustical Society of America* 135 (4). 2198. DOI: 10.1121/1.4877175.
- Anderson, Anne H.; Miles Bader; Ellen G. Bard; Elizabeth Boyle; Gwyneth Doherty; Simon Garrod; Stephen Isard; Jacqueline Kowtko; Jan McAllister; Jim Miller; Catherine Sotillo; Henry S. Thompson & Regina Weinert (1991): The HCRC Map Task Corpus. *Language and Speech* 34 (4). 351–366.
- Arciuli, Joanne; David Mallard & Gina Villar (2010): "Um, I can tell you're lying". Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics* 31 (3). 397–411. DOI: 10.1017/S0142716410000044.
- Arnold, Jennifer E.; Maria Fagnano & Michael K. Tanenhaus (2003): Disfluencies Signal Thee, Um, New Information. *Journal of Psycholinguistic Research* 32 (1). 25–36. DOI: 10.1023/A:1021980931292.
- Arnold, Jennifer E. & Michael K. Tanenhaus (2011): Disfluency effects in comprehension: how new information can become accessible. In: Gibson, Edward & Neal J. Pearlmuter (Hrsg.): *The processing and acquisition of reference*. (= Bradford book). Cambridge, Mass: MIT Press. 197–218.
- Arnold, Jennifer E.; Michael K. Tanenhaus; Rebecca J. Altmann & Maria Fagnano (2004): The Old and Thee, uh, New. *Psychological Science* 15 (9). 578–582.

- Audacity Team (2017): *Audacity®: Free Audio Editor and Recorder. Computer application*. Version 2.2.1.
- Bada, Erdoğan & Bilal Genç (2008): Pausing preceding and following to in to-infinitives: A study with implications to reading and speaking skills in ELT. *Journal of Pragmatics* 40 (11). 1939–1949. DOI: 10.1016/j.pragma.2008.03.010.
- Baker, Rachel & Valerie Hazan (2011): DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods* 43 (3). 761–770. DOI: 10.3758/s13428-011-0075-y.
- Barry, William J. (1995): Schwa vs. Schwa + /r/ in German. *Phonetica* 52. 228–235.
- Barton, Kamil (2018): *MuMIn. Multi-Model Inference*. Version 1.42.1. URL: <https://CRAN.R-project.org/package=MumIn> (besucht am 12.03.2019).
- Bates, Douglas; Martin Mächler; Ben Bolker & Steve Walker (2015): Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67 (1). 1–48. DOI: 10.18637/jss.v067.i01.
- Bates, Douglas; Martin Maechler & Ben Bolker (2014): *lme4: Linear mixed-effects models using Eigen and S4*. Version 1.1-7. URL: <http://cran.r-project.org/web/packages/lme4/lme4.pdf> (besucht am 28.09.2013).
- Batliner, Anton; Andreas Kießling; Susanne Burger & 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 (besucht am 15.07.2019).
- Baumann, Stefan & Bodo Winter (2018): What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38. DOI: 10.1016/j.wocn.2018.05.004.
- Beattie, Geoffrey W. & B. L. Butterworth (1979): Contextual Probability of Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech* 22 (3). 201–211.
- Beattie, Geoffrey W. & Heather Shovelton (2002): Lexical access in talk: A critical consideration of transitional probability and word frequency as possible determinants of pauses in spontaneous speech. *Semiotica* 2002 (141). 49–71. DOI: 10.1515/semi.2002.066.
- Beckman, Mary E. & Janet B. Pierrehumbert (1986): Intonational structure in Japanese and English. *Phonology* 3. 255. DOI: 10.1017/S095267570000066X.

- Bell, Allan (1984): Language Style as Audience Design. English. *Language in Society* 13 (2). 145–204. URL: <http://www.jstor.org/stable/4167516>.
- Belz, Malte (2013): Disfluencies und Reparaturen bei Muttersprachlern und Lernern – eine kontrastive Analyse. Masterarbeit. Humboldt-Universität zu Berlin. DOI: 10.18452/14206.
- Belz, Malte (2019a): *Ein Schema zur phonetischen Annotation von Füllpartikeln und ihres Kontextes. Version. 1.0*. Technischer Bericht. Humboldt-Universität zu Berlin. DOI: 10.18452/20163.
- Belz, Malte (2019b): *GECO-FP*. Humboldt-Universität zu Berlin. DOI: 10.18452/20794.
- Belz, Malte (2019c): *Pseudonymisierung von Sprachchunks in Audiodaten mit Praat und R (Pseudonymization of speech chunks in audio data using Praat and R)*. Humboldt-Universität zu Berlin. DOI: 10.18452/20145.
- Belz, Malte & Myriam Klapi (2013): Pauses following Fillers in L1 and L2 German Map Task Dialogues. In: Eklund, Robert (Hrsg.): *Proceedings of DiSS 2013. The 6th Workshop on Disfluency in Spontaneous Speech*. 9–12.
- Belz, Malte & Christine Mooshammer (2020): *Berlin Dialogue Corpus (BeDiaCo). Version 1*. Humboldt-Universität zu Berlin. URL: <https://rs.cms.hu-berlin.de/phon> (besucht am 25.05.2020).
- Belz, Malte; Oksana Rasskazova; Jelena Krivokapić & Christine Mooshammer ([eingereicht]): Interaction between Phrasal Structure and Segment Identity in German: An Acoustic and Articulatory Study.
- Belz, Malte & Uwe D. Reichel (2015): Pitch characteristics of filled pauses. In: Eklund, Robert (Hrsg.): *Proceedings of Disfluency in Spontaneous Speech (DiSS). The 7th Workshop on Disfluency in Spontaneous Speech*. 1–4.
- Belz, Malte; Simon Sauer; Anke Lüdeling & Christine Mooshammer (2017): Fluently disfluent? Pauses and repairs of advanced learners and native speakers of German. *International Journal of Learner Corpus Research* 3 (2). 118–148. DOI: 10.1075/ijlcr.3.2.02bel.
- Belz, Malte & Jürgen Trouvain (2019): Are 'silent' pauses always silent? In: *19. International Congress of Phonetic Sciences (ICPhS)*. URL: https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_90.pdf (besucht am 21.08.2019).

- Belz, Malte; Alina Zöllner & Megumi Terada (2020): *Dokumentation und Annotationsrichtlinien für das Korpus BeDiaCo Version 1*. Humboldt-Universität zu Berlin. DOI: 10.18452/21361.
- Bentz, Christian & Ramon Ferrer-i-Cancho (2016): Zipf's law of abbreviation as a language universal. In: Bentz, Christian; Gerhard Jäger & Igor Yanovich (Hrsg.): *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistic*. DOI: 10.15496/publikation-10057.
- Betz, Simon; Robert Eklund & Petra Wagner (2017): Prolongation in German. In: Eklund, Robert & Ralph L. Rose (Hrsg.): *Proceedings of DiSS 2017. The 8th Workshop on Disfluency in Spontaneous Speech*. (= TMH-QPSR). 13–16.
- Biber, Douglas & Susan Conrad (2009): *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Blackmer, Elizabeth R. & Janet L. Mitton (1991): Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition* 39 (3). 173–194.
- Blomgren, Michael; Yang Chen; Manwa L. Ng & Harvey R. Gilbert (1998): Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America* 103 (5). 2649–2658. DOI: 10.1121/1.422785.
- Bloomfield, Leonard (1927): Literate and Illiterate Speech. *American Speech* 2 (10). 432–439.
- Bloomfield, Leonard (1933): *Language*. London & Aylesbury: Compton.
- Boersma, Paul (2001): Praat, a system for doing phonetics by computer. *Glott International* 5 (9). 341–345.
- Bokelmann, Johanna; Marie-Therese Ellert; Nico Friesenhan; Glen Generlich; Yelyzaveta Hiebert; Tatjana Malon; Romy Sachs; Aleksandra Swiech; Monika Walak; Isabell Zander; Tamara Rathcke & Christine Mooshammer (2016): Opa vs Oper: Neutralization of /v/ and unstressed /a/ contrast in a perception and production study. In: Draxler, Christoph & Felicitas Kleber (Hrsg.): *12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. 125–128.
- Bombien, Lasse; Raphael Winkelmann & Michel Scheffers (2018): *wrassp. An R wrapper to the ASSP Library*. R package version 0.1.8.
- Boomer, Donald S. & Allen T. Dittmann (1962): Hesitation pauses and juncture pauses in speech. *Language and Speech* 5 (4). 215–220.

- Bortfeld, Heather; Silvia D. Leon; Jonathan E. Bloom; Michael F. Schober & Susan E. Brennan (2001): Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech* 44 (2). 123–147. DOI: 10.1177/00238309010440020101.
- Bosker, Hans Rutger; Hugo Quené; Ted Sanders & Nivia H. De Jong (2014): Native ‘um’s elicit prediction of low-frequency referents, but non-native ‘um’s do not. *Journal of Memory and Language* 75. 104–116. DOI: 10.1016/j.jml.2014.05.004.
- Brand, Christiane & Sandra Götz (2013): Fluency versus accuracy in advanced spoken learner language. A multi-method approach. In: Gilquin, Gaëtanelle & Sylvie de Cock (Hrsg.): *Errors and disfluencies in spoken corpora*. (= Benjamins Current Topics). Amsterdam & Philadelphia: John Benjamins. 117–137.
- Braun, Angelika & Annabelle Rosin (2015): On the speaker-specificity of hesitation markers. In: The Scottish Consortium for ICPhS 2015 (Hrsg.): *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK*. Glasgow: University of Glasgow.
- Brennan, Susan E. & Maurice Williams (1995): The Feeling of Another’s Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language* 34. 383–398.
- Broen, Patricia & Gerald M. Siegel (1972): Variations in normal speech disfluencies. *Language and Speech* 15 (3). 219–231.
- Browman, Catherine P. & Louis Goldstein (1992): Articulatory phonology. An overview. *Phonetica* 49 (3-4). 155–180. DOI: 10.1159/000261913.
- Byrd, Dani; Jelena Krivokapić & Sungbok Lee (2006): How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America* 120 (3). 1589–1599. DOI: 10.1121/1.2217135.
- Byrd, Dani & Elliot Saltzman (2003): The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180. DOI: 10.1016/S0095-4470(02)00085-2.
- Candea, Maria; Ioana Vasilescu & Martine Adda-Decker (2005): Inter- and intra-language acoustic analysis of autonomous fillers. In: Véronis, Jean & Estelle Campione (Hrsg.): *Proceedings of DiSS’05, Disfluency in Spontaneous Speech Workshop*. ISCA Archive. 47–52.

- Carletta, Jean; Amy Isard; Stephen Isard; Jacqueline Kowtko; Gwyneth Doherty-Sneddon & Anne H. Anderson (1997): The reliability of a dialogue structure coding scheme. *Computational Linguistics* 23 (1). 13–31.
- Carletta, Jean; Amy Isard; Stephen Isard; Jacqueline Kowtko; Alison Newlands; Gwyneth Doherty-Sneddon & Anne H. Anderson (1996): *HCRC Dialogue Structure Coding Manual*. URL: www.hcrc.ed.ac.uk/publications/tr-82.ps.gz.
- Carlson, Lauri (1985): *Dialogue Games. An Approach to Discourse Analysis*. Dordrecht: Springer Netherlands.
- Cassidy, Steve & Jonathan Harrington (2001): Multi-level annotation in the Emu speech database management system. *Speech Communication* 33 (1-2). 61–77. DOI: 10.1016/S0167-6393(00)00069-8.
- Cho, Taehong (2008): Prosodic strengthening in transboundary V-to-V lingual movement in American English. *Phonetica* 65 (1-2). 45–61. DOI: 10.1159/000130015.
- Cho, Taehong (2015): Language Effects on Timing at the Segmental and Suprasegmental Levels. In: Redford, Melissa A. (Hrsg.): *The Handbook of Speech Production*. Hoboken, NJ: John Wiley & Sons, Inc. 505–529.
- Cho, Taehong (2016): Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass* 10 (3). 120–141. DOI: 10.1111/lnc3.12178.
- Chomsky, Noam (1965): *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Christenfeld, Nicholas (1995): Does it hurt to say um? *Journal of Nonverbal Behavior* 19 (3). 171–186. DOI: 10.1007/BF02175503.
- Christenfeld, Nicholas (1996): Effects of a Metronome on the Filled Pauses of Fluent Speakers. *Journal of Speech and Hearing Research* 39 (6). 1232–1238.
- Christenfeld, Nicholas & Beth Creager (1996): Anxiety, alcohol, aphasia, and Ums. *Journal of Personality and Social Psychology* 70 (3). 451–460.
- Clark, Herbert H. & Jean E. Fox Tree (2002): Using *uh* and *um* in spontaneous speaking. *Cognition* 84 (1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Cohen, Jacob (1960): A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1). 37–46.
- Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences*. 2. Aufl. Hoboken: Taylor and Francis.

- Collard, Philip; Martin Corley; Lucy J. MacGregor & David I. Donaldson (2008): Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34 (3). 696–702. DOI: 10.1037/0278-7393.34.3.696.
- Cook, Mark (1971): The incidence of filled pauses in relation to part of speech. *Language and Speech* 14 (2). 135–139. DOI: 10.1177/002383097101400203.
- Corley, Martin; Robert J. Hartsuiker & Matjaz Perc (2011): Why *Um* Helps Auditory Word Recognition: The Temporal Delay Hypothesis. *PLoS ONE* 6 (5). e19792. DOI: 10.1371/journal.pone.0019792.
- Corley, Martin; Lucy J. MacGregor & David I. Donaldson (2007): It's the way that you, er, say it: hesitations in speech affect language comprehension. *Cognition* 105 (3). 658–668. DOI: 10.1016/j.cognition.2006.10.010.
- Corley, Martin & Oliver W. Stewart (2008): Hesitation Disfluencies in Spontaneous Speech: The Meaning of *um*. *Language and Linguistics Compass* 2 (4). 589–602. DOI: 10.1111/j.1749-818X.2008.00068.x.
- Dahl, David B. (2016): *xtable*. *Export Tables to LaTeX or HTML*. R package version 1.8-2. URL: <https://CRAN.R-project.org/package=xtable>.
- Dahl, Deborah A.; Madeleine Bates; Michael Brown; William Fisher; Kate Hunnicke-Smith; David Pallett; Christine Pao; Alexander Rudnicky & Elizabeth Shriberg (1994): Expanding the scope of the ATIS task: the ATIS-3 corpus. In: *Proceedings of the workshop on Human Language Technology*. 43–48. DOI: 10.3115/1075812.1075823.
- Davidson, Lisa (2018): The Effects of Pitch, Gender, and Prosodic Context on the Identification of Creaky Voice. *Phonetica*. DOI: 10.1159/000490948.
- Dietrich, Rainer (2007): *Psycholinguistik*. 2., aktualisierte und erw. Aufl. (= Sammlung Metzler, Bd. 342). Stuttgart: J.B. Metzler.
- Dilley, Laura; Stefanie Shattuck-Hufnagel & Mari Ostendorf (1996): Glottalization of Word-Initial Vowels as a Function of Prosodic Structure. *Journal of Phonetics* 24 (4). 423–444. DOI: 10.1006/jpho.1996.0023.
- Du Bois, John W.; Wallace L. Chafe; Charles Meyer & Sandra A. Thompson (2000): *Santa Barbara corpus of spoken American English, Part 1*. Linguistic Data Consortium. URL: <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus> (besucht am 17.06.2019).

- Dudenredaktion (2001): *Duden. Deutsches Universalwörterbuch*. 4. Aufl. Mannheim: Dudenverlag.
- Duez, Danielle (1993): Acoustic Correlates of Subjective Pauses. *Journal of Psycholinguistic Research* 22 (1). 21–39. (Besucht am 07.06.2019).
- Duez, Danielle (2001): Acoustico-phonetic Characteristics of Filled Pauses in Spontaneous French Speech: Preliminary Results. In: *ITRW on Disfluency in Spontaneous Speech (DiSS'01)*. 41–44.
- Eberhard, Kathleen; Hannele Buffy Mary Nicholson; Sandra Kübler; Susan Gundersen & Matthias Scheutz (2010): The Indiana "Cooperative Remote Search Task" (CReST) Corpus. In: *Proceedings of LREC*. 3024–3030.
- Ehlich, Konrad (1986): *Interjektionen*. (= Linguistische Arbeiten, Bd. 111). Tübingen: Max Niemeyer Verlag.
- Ehlich, Konrad & Jochen Rehbein (1976): Halbinterpretative Arbeitstranskriptionen (HIAT). *Linguistische Berichte* (45). 21–41.
- Ehlich, Konrad & Jochen Rehbein (1979): Erweiterete halbinterpretative Arbeitstranskriptionen (HIAT 2): Intonation. *Linguistische Berichte* (59). 51–75.
- Eklund, Robert (2004): Disfluency in Swedish human-human and human-machine travel booking dialogues. Department of Computer and Information Science. Dissertation. Linköping & Sweden: Linköpings Universitet.
- Faure, Marc (1980): Results of a contrastive study of hesitation phenomena in French and German. In: Dechert, Hans W. & Manfred Raupach (Hrsg.): *Temporal variables in speech. Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton. 287–290.
- Fedriani, Chiara & Andrea Sansò, Hrsg. (2017): *Pragmatic Markers, Discourse Markers and Modal Particles. New perspectives*. Bd. 186. (= Studies in Language Companion Series, Bd. 186). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Ferreira, Fernanda (1993): Creation of Prosody During Sentence Production. *Psychological Review* 100 (2). 233–253. DOI: 10.1037/0033-295X.100.2.233.
- Ferreira, Fernanda & Karl G. D. Bailey (2004): Disfluencies and human language comprehension. *Trends in Cognitive Sciences* 8 (5). 231–237. DOI: 10.1016/j.tics.2004.03.011.

- Fiehler, Reinhard; Birgit Barden; Mechthild Elstermann & Barbara Kraft (2004): *Eigenschaften gesprochener Sprache*. (= Studien zur deutschen Sprache, Bd. 30). Tübingen: Narr.
- Finlayson, Ian R. & Martin Corley (2012): Disfluency in dialogue: an intentional signal from the speaker? *Psychonomic Bulletin & Review* 19 (5). 921–928. DOI: 10.3758/s13423-012-0279-x.
- Fischer, Kerstin (1999): Die Ikonizität der Pause: Zwischen kognitiver Last und kommunikativer Funktion. In: Wachsmuth, Ipke & Bernhard Jung (Hrsg.): *Proceedings der KogWis99*. Gesellschaft für Kognitionswissenschaft. 250–255.
- Fischer, Kerstin (2000): *From Cognitive Semantics to Lexical Pragmatics. The Functional Polysemy of Discourse Particles*. Berlin, New York: Mouton de Gruyter.
- Fletcher, Janet (2010): The Prosody of Speech: Timing and Rhythm. In: Hardcastle, William J.; John Laver & Fiona E. Gibbon (Hrsg.): *The handbook of phonetic sciences*. 2. Aufl. (= Blackwell handbooks in linguistics). Chichester, U.K. & Malden, MA: Wiley-Blackwell. 523–602.
- Flynn, Nicholas (2011): Comparing Vowel Formant Normalisation Procedures. *York Papers in Linguistics Series 2*. 1–28.
- Fors, Kristina Lundholm (2015): Production and Perception of Pauses in Speech. Dissertation. University of Gothenburg.
- Foulkes, Paul; Gareth Carrol & Samantha Hughes (2004): Sociolinguistic and acoustic variability in filled pauses. In: *Annual conference of the International Association for Forensic Phonetics and Acoustics*. URL: <https://docs.google.com/file/d/0Bw5I8R4GWPHNd0hxQzAteTZYS2M/edit> (besucht am 09.10.2018).
- Fowler, Carol A. & Elliot Saltzman (1993): Coordination and coarticulation in speech production. *Language and Speech* 36 (2, 3). 171–195. DOI: 10.1177/002383099303600304.
- Fox, Barbara A. & Robert Jasperson (1995): A Syntactic Exploration of Repair in English Conversation. In: Davis, Philip W. (Hrsg.): *Alternative linguistics. Descriptive and theoretical modes*. (= Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory). Amsterdam, Philadelphia: J. Benjamins. 77–134.
- Fox, John & Sanford Weisberg (2011): *An R Companion to Applied Regression*. Second. Thousand Oaks CA: Sage.

- Fox Tree, Jean E. (1995): The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language* 34 (6). 709–738. DOI: 10.1006/jmla.1995.1032.
- Fox Tree, Jean E. (2001): Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition* 29 (2). 320–326. DOI: 10.3758/BF03194926.
- Fraundorf, Scott H. & Duane G. Watson (2011): The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language* 65 (2). 161–175. DOI: 10.1016/j.jml.2011.03.004.
- Fraundorf, Scott H. & Duane G. Watson (2013): Alice's adventures in *um*-derland: Psycholinguistic sources of variation in disfluency production. *Language and Cognitive Processes* 29 (9). 1083–1096. DOI: 10.1080/01690965.2013.832785.
- Fruehwald, Josef (2016): Filled pause choice as a sociolinguistic variable. In: *University of Pennsylvania Working Papers in Linguistics: Selected Papers from New Ways of Analyzing Variation*. 2. Aufl. 41–49. (Besucht am 27.06.2018).
- Fuchs, Susanne; Caterina Petrone; Amélie Rochet-Capellan; Uwe D. Reichel & Laura L. Koenig (2015): Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics* 52. 35–45. DOI: 10.1016/j.wocn.2015.04.002.
- Gamer, Matthias; Jim Lemon & Ian Fellows Puspendra Singh (2012): *irr. Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84. URL: <https://CRAN.R-project.org/package=irr>.
- Georgeton, Laurianne & Cécile Fougeron (2014): Domain-initial strengthening on French vowels and phonological contrasts. Evidence from lip articulation and spectral variation. *Journal of Phonetics* 44. 83–95. DOI: 10.1016/j.wocn.2014.02.006.
- Gick, Bryan; Ian Wilson; Karsten Koch & Clare Cook (2004): Language-Specific Articulatory Settings: Evidence from Inter-Utterance Rest Position. *Phonetica* 61 (61). 220–233. DOI: 10.1159/000084159.
- Gil, David (2013): Para-Linguistic Usages of Clicks. In: Matthew S. Dryer & Martin Haspelmath (Hrsg.): *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Kapitel 142. URL: <http://wals.info/chapter/142> (besucht am 17.01.2017).

- Ginzburg, Jonathan; Raquel Fernández & David Schlangen (2014): Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics* 7. DOI: 10.3765/sp.7.9.
- Glück, Helmut (2005): *Metzler Lexikon Sprache*. 3., neubearbeitete Aufl. Stuttgart: Metzler.
- Godfrey, John & Edward Holliman (1993): *Switchboard-1 Release 2 LDC97S62*. Linguistic Data Consortium. URL: <https://catalog.ldc.upenn.edu/LDC97S62>.
- Goldman-Eisler, Frieda (1958): The predictability of words in context and the length of pauses in speech. *Language and Speech* 1 (3). 226–231. DOI: 10.1177/002383095800100308.
- Goldman-Eisler, Frieda (1961): A comparative study of two hesitation phenomena. *Language and Speech* 4 (1). 18–26. DOI: 10.1177/002383096100400102.
- Goodwin, Charles (1981): *Conversational Organization. Interaction between Speakers and Hearers*. (= Language, Thought, And Culture: Advances in the Study of Cognition). New York: Academic Press.
- Gordon, Matthew & Peter Ladefoged (2001): Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29 (4). 383–406. DOI: 10.1006/jpho.2001.0147.
- Gósy, Mária; Dorottya Gyarmathy & András Beke (2017): Phonetic analysis of filled pauses based on a Hungarian-English learner corpus. *International Journal of Learner Corpus Research* 3 (2). 149–174. DOI: 10.1075/ijlcr.3.2.03gos.
- Götz, Sandra (2013): *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/sc1.53.
- Grice, Martine; Stefan Baumann & Ralf Benz Müller (2005): German Intonation in Autosegmental-Metrical Phonology. In: Jun, Sun-Ah (Hrsg.): *Prosodic Typology*. Oxford University Press. 1–33. DOI: 10.1093/acprof:oso/9780199249633.003.0003.
- Grosz, Barbara J.; Scott Weinstein & Aravind K. Joshi (1995): Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics* 21 (2). 203–225.
- Günthner, Susanne (2017): Diskursmarker in der Interaktion - Formen und Funktionen univerbierter *guck mal-* und *weiß du-*Konstruktionen. In: Blühdorn, Hardarik; Arnulf Deppermann; Henrike Helmer & Thomas Spranz-Fogasy (Hrsg.):

- Diskursmarker im Deutschen. Reflexionen und Analysen.* Göttingen: Verlag für Gesprächsforschung. 103–130.
- Hall, Tracy Alan (2011): *Phonologie. Eine Einführung.* Berlin/New York: de Gruyter.
- Harrington, Jonathan (2010): Acoustic Phonetics. In: Hardcastle, William J.; John Laver & Fiona E. Gibbon (Hrsg.): *The handbook of phonetic sciences.* 2. Aufl. (= Blackwell handbooks in linguistics). Chichester, U.K. & Malden, MA: Wiley-Blackwell. 81–129.
- Harrington, Jonathan; Janet Fletcher & Mary E. Beckman (2000): Manner and place conflicts in the articulation of accent in Australian English. In: Broe, Michael B. & Janet B. Pierrehumbert (Hrsg.): *Papers in Laboratory Phonology V. Acquisition and the lexicon.* Cambridge: Cambridge University Press. 40–51.
- Harrison, Xavier A.; Lynda Donaldson; Maria Eugenia Correa-Cano; Julian Evans; David N. Fisher; Cecily E. D. Goodwin; Beth S. Robinson; David J. Hodgson & Richard Inger (2018): A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6. e4794. DOI: 10.7717/peerj.4794.
- Hau, Rita (2006): *PONS Wörterbuch für Schule und Studium. Latein.* 3., neu bearb. Aufl., 1. Nachdr. Stuttgart: Klett.
- Hawkins, P. R. (1971): The syntactic location of hesitation pauses. *Language and Speech* 14 (3). 277–288. DOI: 10.1177/002383097101400308.
- Hay, Jennifer; Paul Warren & Katie Drager (2006): Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34 (4). 458–484. DOI: 10.1016/j.wocn.2005.10.001.
- Heeman, Peter A. & James F. Allen (1999): Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog. *Computational Linguistics* 25 (4). 527–571. DOI: 10.3115/976909.979650.
- Heinz, Bettina (2003): Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics* 35 (7). 1113–1142. DOI: 10.1016/S0378-2166(02)00190-X.
- Hieke, Adolf E.; Sabine Kowal & Daniel C. O'Connell (1983): The Trouble with "Articulatory" Pauses. *Language and Speech* 26 (3). 203–214. DOI: 10.1177/002383098302600302.

- Himmelman, Nikolaus P.; Meytal Sandler; Jan Strunk & Volker Unterladstetter (2018): On the universality of intonational phrases: a cross linguistic interrater study. *Phonology* 35 (2). 207–245. DOI: 10.1017/S0952675718000039.
- Honikman, Beatrice (1964): Articulatory settings. In: Abercrombie, Daniel; Dennis Butler Fry; MacCarthy, Peter Arthur Desmond; N. C. Scott & Trim, John L. M. (Hrsg.): *In Honour of Daniel Jones*. London: Longman. 73–84.
- Horne, Merle (2006): The Filler EH in Swedish. *Lund Working Papers in Linguistics* (52). 65–86.
- Horne, Merle (2009): Creaky Fillers and Speaker Attitude: Data from Swedish. In: Barth-Weingarten, Dagmar; Nicole Dehé & Anne Wichmann (Hrsg.): *Where Prosody Meets Pragmatics*. (= Studies in Pragmatics). Bingley, UK: Emerald. 277–288.
- Horváth, Viktória (2010): Filled pauses in Hungarian: their phonetic form and function. *Acta Linguistica Hungarica* 57 (2-3). 288–306. DOI: 10.1556/ALing.57.2010.2-3.6.
- Hudson Kam, Carla L. & Nicole A. Edwards (2008): The use of uh and um by 3- and 4-year-old native English-speaking children: Not quite right but not completely wrong. *First Language* 28 (3). 313–327. DOI: 10.1177/0142723708091149.
- Hughes, Vincent; Sophie Wood & Paul Foulkes (2016): Strength of forensic voice comparison evidence from the acoustics of filled pauses. *The International Journal of Speech, Language and the Law* 23 (1). 99–132.
- Imo, Wolfgang (2016): Dialogizität – eine Einführung. *Zeitschrift für germanistische Linguistik* 44 (3). 337–356. DOI: 10.1515/zgl-2016-0019.
- Iseli, Markus; Yen-Liang Shue & Abeer Alwan (2007): Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America* 121 (4). 2283–2295. DOI: 10.1121/1.2697522.
- Jäger, Karl-Heinz, Hrsg. (1979): *Texte gesprochener deutscher Standardsprache IV*. Bd. II/4. (= Heutiges Deutsch, Bd. II/4). München: Max Hueber Verlag.
- Jefferson, Gail (1974): Error correction as an interactional resource. *Language in Society* 3 (02). 181–199. DOI: 10.1017/S0047404500004334.
- Jessen, Michael (2012): *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*. (= LINCOM Studies in Phonetics, Bd. 9). München: LINCOM.

- Johnson, Keith; Peter Ladefoged & Mona Lindau (1993): Individual differences in vowel production. *The Journal of the Acoustical Society of America* 94 (2). 701–714. DOI: 10.1121/1.406887.
- Jucker, Andreas H. (2015a): Pragmatics of fiction: Literary uses of *uh* and *um*. *Journal of Pragmatics* 86. 63–67. DOI: 10.1016/j.pragma.2015.05.012.
- Jucker, Andreas H. (2015b): *Uh* and *um* as planners in the *Corpus of Historical American English*. In: Taavitsainen, Irma; Merja Kytö; Claudia Claridge & Jeremy Smith (Hrsg.): *Developments in English. Expanding electronic evidence.* (= Studies in English language). Cambridge: Cambridge University Press. 162–177. DOI: 10.1017/CB09781139833882.013.
- Jurafsky, Daniel; Rajesh Ranganath & Dan McFarland (2009): Extracting social meaning: identifying interactional style in spoken conversation. In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.* 638–646. URL: <https://dl.acm.org/citation.cfm?id=1620754.1620847> (besucht am 17.06.2019).
- Kamp, Hans; Josef van Genabith & Uwe Reyle (2011): Discourse Representation Theory. In: Gabbay, Dov M. & Franz Guenther (Hrsg.): *Handbook of philosophical logic.* 2. Aufl. Dordrecht: Springer. 125–394. DOI: 10.1007/978-94-007-0485-5_3.
- Karpiński, Maciej (2013): Acoustic Features of Filled Pauses in Polish Task-Oriented Dialogues. *Archives of Acoustics* 38 (1). DOI: 10.2478/aoa-2013-0008.
- Kasl, Stanislav V. & George F. Mahl (1965): Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology* 1 (5). 425–433. DOI: 10.1037/h0021918.
- Katz, Jonah (2012): Compression effects in English. *Journal of Phonetics* 40 (3). 390–402. DOI: 10.1016/j.wocn.2012.02.004.
- Keating, Patricia A. (2006): Phonetic Encoding of Prosodic Structure. In: Harrington, Jonathan & Marija Tabain (Hrsg.): *Speech production. Models, Phonetic Processes, and Techniques.* (= Macquarie monographs in cognitive science). New York: Psychology Press. 167–186.

- Keseling, Gisbert (1989): Die Partikel *ÄH*. Ein paraverbales Element im Sprachsystem? In: Weydt, Harald (Hrsg.): *Sprechen mit Partikeln*. Berlin, New York: Walter de Gruyter. 575–591.
- Kisler, Thomas; Uwe Reichel & Florian Schiel (2017): Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347. DOI: 10.1016/j.cs1.2017.01.005.
- Kjellmer, Göran (2003): Hesitation. In Defence of ER and ERM. *English Studies* 84 (2). 170–198. DOI: 10.1076/enst.84.2.170.14903.
- Klein, Wolfgang (1985): Gesprochene Sprache - geschriebene Sprache. *Zeitschrift für Literaturwissenschaft und Linguistik* 15 (59). 9–35.
- Klug, Katharina (2013): „Ähm“ – Sind Häsitations-Partikeln sprecherspezifisch? Untersuchung der Parameter Grundfrequenz und Vokalqualität. In: Anders, Lutz Christian; Ines Bose; Ursula Hirschfeld & Baldur Neuber (Hrsg.): *Aktuelle Forschungsthemen der Sprechwissenschaft 3. Phonetik, Rhetorik, Sprechkunst, Sprach- und Stimmstörungen*. (= Hallesche Schriften zur Sprechwissenschaft und Phonetik (HSSP)). Frankfurt am Main: Peter Lang. 65–94.
- Kohler, Klaus J. (1974): Koartikulation und Steuerung im Deutschen. In: Engel, Ulrich & Paul Grebe (Hrsg.): *Sprachsystem und Sprachgebrauch. Festschrift für Hugo Moser*. Teil 1. Pädagogischer Verlag Schwann. 172–192.
- Kohler, Klaus J. (1983): Prosodic boundary signals in German. *Phonetica* (40). 89–134. DOI: 10.1159/000261685.
- Kohler, Klaus J. (1999a): *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*.
- Kohler, Klaus J. (1999b): German. In: *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge Univ. Press. 86–89.
- Kohler, Klaus J.; Benno Peters & Thomas Wesener (2005): Phonetic Exponents of Disfluency in German Spontaneous Speech. In: *Prosodic Structures in German Spontaneous Speech*. (= Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel). 185–201.
- Kowtko, Jacqueline & Patti J. Price (1989): Data collection and analysis in the air travel planning domain. In: *Proceedings of the DARPA Speech and Natural*

- Language Workshop*. 119–125. URL: <https://www.aclweb.org/anthology/H89-2017> (besucht am 17. 06. 2019).
- Krifka, Manfred & Renate Musan, Hrsg. (2012): *The expression of information structure*. (= The expression of cognitive categories). Berlin & Boston: Mouton de Gruyter.
- Krivokapić, Jelena (2007): Prosodic Planning. Effects of Phrasal Length and Complexity on Pause Duration. *Journal of Phonetics* 35 (2). 162–179. DOI: 10.1016/j.wocn.2006.04.001.
- Krivokapić, Jelena (2014): Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical transactions of the Royal Society. Series B, Biological sciences* 369 (1658). DOI: 10.1098/rstb.2013.0397.
- Kübler, Sandra & Heike Zinsmeister (2015): *Corpus linguistics and linguistically annotated corpora*. London [u.a.]: Bloomsbury.
- Kügler, Frank & Stefan Baumann (2019): *Annotationsrichtlinien DIMA*. URL: <http://dima.uni-koeln.de/wp-content/uploads/2019/02/DIMA-Annotationsrichtlinien.V4.0.pdf> (besucht am 06. 05. 2019).
- Kügler, Frank; Bernadett Smoliboeki; Denis Arnold; Stefan Baumann; Bettina Braun; Martine Grice; Stefanie Jannedy; Jan Michalsky; Oliver Niebuhr; Jörg Peters; Simon Ritter; Christine Tanja Röhr; Antje Schweitzer et al. (2015): DIMA – Annotation guidelines for German intonation. In: The Scottish Consortium for ICPhS 2015 (Hrsg.): *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK*. Glasgow: University of Glasgow.
- Künzel, Hermann J. (1987): *Sprechererkennung. Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.
- Kuznetsova, Alexandra; Per B. Brockhoff & Rune H. B. Christensen (2017): lmerTest Package. Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82 (13). 1–26. DOI: 10.18637/jss.v082.i13.
- Labov, William (1972): *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Ladd, D. Robert (1996): *Intonational phonology*. (= Cambridge studies in linguistics, Bd. 79). Cambridge & New York: Cambridge University Press.

- Ladefoged, Peter & Keith Johnson (2011): *A course in phonetics*. 6. Aufl. Boston, MA: Wadsworth/Cengage Learning.
- Lake, Johanna K.; Karin R. Humphreys & Shannon Cardy (2011): Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review* 18 (1). 135–140. DOI: 10.3758/s13423-010-0037-x.
- Lallgee, Mansur G. & Mark Cook (1969): An experimental investigation of the function of filled pauses in speech. *Language and Speech* 12 (1). 24–28.
- Lalljee, Mansur & Mark Cook (1974): Filled Pauses and Floor-Holding. The Final Test? *Semiotica* 12 (3). DOI: 10.1515/semi.1974.12.3.219.
- Landis, J. Richard & Gary G. Koch (1977): The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1). 159. DOI: 10.2307/2529310.
- Laserna, Charlyn M.; Yi-Tai Seih & James W. Pennebaker (2014): Um . . . Who Like Says You Know: Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology* 33 (3). 328–338. DOI: 10.1177/0261927X14526993.
- Laver, John (1994): *Principles of phonetics*. (= Cambridge textbooks in linguistics). Cambridge & New York, NY: Cambridge University Press.
- Lay, Clarry H. & Allan Paivio (1969): The effects of task difficulty and anxiety on hesitations in speech. *Canadian journal of behavioural science* 1 (1). 25–37.
- Leeuw, Esther de (2007): Hesitation Markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19 (2). 85–114. DOI: 10.1017/S1470542707000049.
- Lehiste, Ilse & Gordon E. Peterson (1961): Transitions, Glides, and Diphthongs. *The Journal of the Acoustical Society of America* 33 (3). 268–277. DOI: 10.1121/1.1908638.
- Leifeld, Philip (2013): texreg. Conversion of Statistical Model Output in R to HTML Tables. *Journal of Statistical Software* 55 (8). 1–24. DOI: 10.18637/jss.v055.i08.
- Lenth, Russell (2019): *emmeans. Estimated Marginal Means, aka Least-Squares Means*. Version 1.4. URL: <https://CRAN.R-project.org/package=emmeans> (besucht am 10.09.2019).
- Levelt, Willem J. M. (1983): Monitoring and self-repair in speech. *Cognition* 14 (1). 41–104.

- Levelt, Willem J. M. (1989): *Speaking. From intention to articulation*. Cambridge & Mass. [u.a.]: MIT Press.
- Levinson, Stephen C. (2000): *Pragmatik*. Neu übersetzt von Martina Wiese. Tübingen: Niemeyer.
- Lickley, Robin J. (2001): Dialogue moves and disfluency rates. In: *ITRW on Disfluency in Spontaneous Speech (DiSS'01)*. 93–96.
- Lickley, Robin J. (2015): Fluency and Disfluency. In: Redford, Melissa A. (Hrsg.): *The Handbook of Speech Production*. Hoboken, NJ: John Wiley & Sons, Inc. 445–469. DOI: 10.1002/9781118584156.ch20.
- Lickley, Robin J. & Ellen G. Bard (1998): When Can Listeners Detect Disfluency in Spontaneous Speech? *Language and Speech* 41 (2). 203–226. DOI: 10.1177/002383099804100204.
- Lindblom, Björn (1963): Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773–1781. DOI: 10.1121/1.1918816.
- Linell, Per (1982): *The written language bias in linguistics*. (= Studies in Communication, Bd. 2). Linköping: University of Linköping.
- Linell, Per (2005): *The Written Language Bias in Linguistics. Its nature, origins and transformations*. (= Routledge Advances in Communication and Linguistic Theory). New York: Routledge.
- Lobanov, Boris M. (1971): Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49 (2B). 606–608.
- Local, John & Gareth Walker (2005): Methodological imperatives for investigating the phonetic organization and phonological structures of spontaneous speech. *Phonetica* 62 (2-4). 120–130. DOI: 10.1159/000090093.
- Lounsbury, Floyd G. (1954): Transitional Probability, Linguistic Structure, and Systems of Habit-family Hierarchies. In: Osgood, Charles E. & Thomas A. Sebeok (Hrsg.): *Psycholinguistics. A survey of theory and research problems*. Baltimore: Waverly Press. 93–101.
- Lüdeling, Anke (2011): Corpora in Linguistics: Sampling and Annotation. In: Grandin, Karl (Hrsg.): *Going Digital. Evolutionary and Revolutionary Aspects of Digitization*. (= Nobel Symposium). New York: Science History Publications. 220–243.

- Lüdeling, Anke (2017): Variationistische Korpusstudien. In: Konopka, Marek & Angelika Wöllstein (Hrsg.): *Grammatische Variation. Empirische Zugänge und theoretische Modellierung.* (= Jahrbuch des Instituts für Deutsche Sprache). Berlin & Boston: de Gruyter. 129–144.
- MacGregor, Lucy J. (2008): Disfluencies affect language comprehension: evidence from event-related potentials and recognition memory. School of Philosophy, Psychology and Language Sciences. Dissertation. Edinburgh: University of Edinburgh.
- Maclay, Howard & Charles E. Osgood (1959): Hesitation Phenomena in Spontaneous English Speech. *Word* 5. 19–44. DOI: 10.1080/00437956.1959.11659682.
- Maddieson, Ian (1985): Phonetic cues to syllabification. In: Fromkin, Victoria A. (Hrsg.): *Phonetic linguistics. Essays in honor of Peter Ladefoged.* Orlando u.a.: Academic Press. 203–221.
- Maekawa, Kikuo & Hiroki Mori (2015): Voice-quality analysis of Japanese filled pauses: a preliminary report. In: Eklund, Robert (Hrsg.): *Proceedings of Disfluency in Spontaneous Speech (DiSS). The 7th Workshop on Disfluency in Spontaneous Speech.* 61–64.
- Mahl, George F. (1956): Disturbances and silences in the patient's speech in psychotherapy. *The Journal of abnormal and social psychology* 53 (1). 1–15. DOI: 10.1037/h0047552.
- Mann, William C. & Sandra A. Thompson (1988): Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8 (3). 243–281. DOI: 10.1515/text.1.1988.8.3.243.
- Martin, James G. & Winifred Strange (1968): The perception of hesitation in spontaneous speech. *Perception & Psychophysics* 3 (6). 427–438. DOI: 10.3758/BF03205750.
- McDougall, Kirsty & Martin Duckworth (2017): Profiling fluency. An analysis of individual variation in disfluencies in adult males. *Speech Communication* 95. 16–27. DOI: 10.1016/j.specom.2017.10.001.
- McDougall, Kirsty & Martin Duckworth (2018): Individual patterns of disfluency across speaking styles: a forensic phonetic investigation of Standard Southern British English. *The International Journal of Speech, Language and the Law* 25 (2). 205–230.

- Michalke, Meik (2017): *syllly.de. Language Support for 'syllly' Package: German.* (Version 0.1-2). URL: <https://github.com/unDocUmeantIt/syllly> (besucht am 28.08.2019).
- Moniz, Helena; Fernando Batista; Ana Isabel Mata & Isabel Trancoso (2014): Speaking style effects in the production of disfluencies. *Speech Communication* 65. 20–35. DOI: 10.1016/j.specom.2014.05.004.
- Mooshammer, Christine (2010): Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *Journal of the Acoustical Society of America* 127 (2). 1047–1058. DOI: 10.1121/1.3277160.
- Musan, Renate (2010): *Informationsstruktur.* (= Kurze Einführungen in die germanistische Linguistik, Bd. 9). Heidelberg: Winter.
- Nakagawa, Shinichi & Holger Schielzeth (2013): A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2). 133–142. DOI: 10.1111/j.2041-210x.2012.00261.x.
- Nespor, Marina & Irene Vogel (1986): *Prosodic phonology.* Dordrecht: Foris Publication.
- Nicholson, Hannele Buffy Mary; Kathleen Eberhard & Matthias Scheutz (2010): "Um... I don't see any": The Function of Filled Pauses and Repairs. In: *Proceedings of DiSS-LPSS Joint Workshop 2010, The 5th Workshop on Disfluency in Spontaneous Speech and The 2nd International Symposium on Linguistic Patterns in Spontaneous Speech.* 89–92.
- Norrick, Neal R. (2009): Interjections as pragmatic markers. *Journal of Pragmatics* 41 (5). 866–891. DOI: 10.1016/j.pragma.2008.08.005.
- Nycz, Jennifer & Lauren Hall-Lew (2013): Best practices in measuring vowel merger. In: *Proceedings of Meetings on Acoustics 20, 060008.* Acoustical Society of America. 1–19. DOI: 10.1121/1.4894063.
- O'Connell, Daniel C. & Sabine Kowal (2004): The History of Research on the Filled Pause as Evidence of The Written Language Bias in Linguistics (Linell, 1982). *Journal of Psycholinguistic Research* 33 (6). 459–474. DOI: 10.1007/s10936-004-2666-6.
- O'Connell, Daniel C. & Sabine Kowal (2005): Uh and Um Revisited: Are They Interjections for Signaling Delay? *Journal of Psycholinguistic Research* 34 (6). 555–576. DOI: 10.1007/s10936-005-9164-3.

- O'Shaughnessy, Douglas (1992): Recognition of hesitations in spontaneous speech. In: *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92, 1992 IEEE International Conference on*. 521–524.
- Odebrecht, Carolin (2018): MKM – ein Metamodell für Korpusmetadaten. Dokumentation und Wiederverwendung historischer Korpora. Dissertation. Berlin: Humboldt-Universität zu Berlin. DOI: 10.18452/19407.
- Oviatt, Sharon (1995): Predicting spoken disfluencies during human–computer interaction. *Computer, Speech, and Language* 9 (1). 19–35. DOI: 10.1006/csla.1995.0002.
- Pätzold, Matthias & Adrian P. Simpson (1995): An acoustic analysis of hesitation particles in German. In: Kjell, Elenius & Peter Branderud (Hrsg.): *Proceedings of the XIIIth International Congress of Phonetic Science (ICPhS 95)*. 512–515.
- Pätzold, Matthias & Adrian P. Simpson (1997): Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. In: *The Kiel Corpus of Read/Spontaneous Speech - Acoustic data base, processing tool and analysis results*. 215–247.
- Peters, Benno (2003): Multiple Cues for Phonetic Phrase Boundaries in German Spontaneous Speech. In: *15th International Congress of Phonetic Sciences (ICPhS-15)*. 1795–1798. URL: http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1795.pdf (besucht am 12.03.2018).
- Peters, Benno (2005a): The database. The Kiel Corpus of Spontaneous Speech. In: *AIPUK 35a*. Kiel. 1–6.
- Peters, Benno (2005b): Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. In: *AIPUK 35a*. Kiel. 203–345.
- Peters, Benno; Matthias Hoffmann & Laura-Marie Andresen (2016): Sprachdatenerhebung und Kontextvariation. In: Draxler, Christoph & Felicitas Kleber (Hrsg.): *12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. 136–139.
- Peters, Benno; Klaus J. Kohler & Thomas Wesener (2005): Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In: *AIPUK 35a*. Kiel. 143–184.
- Pierrehumbert, Janet B. & Stefan Frisch (1997): Synthesizing Allophonic Glottalization. In: Santen, Jan P. H.; Joseph P. Olive; Richard W. Sproat & Julia Hirschberg (Hrsg.): *Progress in Speech Synthesis*. New York, NY: Springer. 9–26.

- Plag, Ingo; Julia Homann & Gero Kunter (2017): Homophony and morphology. The acoustics of word-final S in English. *Journal of Linguistics* 53 (01). 181–216. DOI: 10.1017/S0022226715000183.
- Poerner, Nina & Florian Schiel (2018): A Web Service for Pre-segmenting Very Long Transcribed Speech Recordings. In: *Proceedings of LREC*. 2860–2865.
- Pompino-Marschall, Bernd (2009): *Einführung in die Phonetik*. 3. Aufl. (= De Gruyter Studienbuch). Berlin, New York: Walter de Gruyter.
- Pompino-Marschall, Bernd & Marzena Żygis (2010): Glottal Marking of Vowel-Initial Words in German. *ZAS Papers in Linguistics* (52). 1–17.
- R Core Team (2018): *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Ramanarayanan, Vikram; Erik Bresch; Dani Byrd; Louis Goldstein & Shrikanth S. Narayanan (2009): Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *The Journal of the Acoustical Society of America* 126 (5). EL160–164. DOI: 10.1121/1.3213452.
- Rasoloson, Janie Noëlle (1994): *Interjektionen im Kontrast. Am Beispiel der deutschen, madagassischen, englischen und französischen Sprache*. (= Arbeiten zur Sprachanalyse, Bd. 22). Frankfurt am Main: Peter Lang.
- Rasskazova, Oksana; Christine Mooshammer & Susanne Fuchs (2018): Articulatory settings during inter-speech pauses. In: Belz, Malte; Christine Mooshammer; Susanne Fuchs; Stefanie Jannedy; Oksana Rasskazova & Marzena Żygis (Hrsg.): *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13)*. 161–164. DOI: 10.18452/18805.
- Rathcke, Tamara & Christine Mooshammer (to appear): *Grandpa or Opera? An unstressed vowel merger in German*. *Journal of the International Phonetic Association*.
- Recasens, Daniel (1999): Lingual coarticulation. In: Hardcastle, William J. & Nigel Hewlett (Hrsg.): *Coarticulation. Theory, Data and Techniques*. (= Cambridge Studies in Speech Science and Communication). Cambridge University Press. 80–104.
- Redi, Laura & Stefanie Shattuck-Hufnagel (2001): Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29 (4). 407–429. DOI: 10.1006/jpho.2001.0145.

- Reitbrecht, Sandra (2017): *Häsitationsphänomene in der Fremdsprache Deutsch und ihre Bedeutung für die Sprechwirkung*. (= Schriften zur Sprechwissenschaft und Phonetik, Bd. 10). Berlin: Frank & Timme.
- Rendle-Short, Johanna (2004): Showing structure: using *um* in the academic seminar. *Pragmatics* 14 (4). 479–498. DOI: 10.1075/prag.14.4.04ren.
- Rochester, Sherry R. (1973): The Significance of Pauses in Spontaneous Speech. *Journal of Psycholinguistic Research* 2 (1). 51–81.
- Rodgers, Jonathan E. J. (1999): Three influences on glottalization in read and spontaneous German speech. In: *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*. 173–280. URL: http://www.ipds.uni-kiel.de/kjk/pub_exx/jr1999_1/aipuk34_rodgers.pdf (besucht am 28.08.2019).
- Roettger, Timo B. (2019): Researcher degrees of freedom in phonetic research. *Laboratory Phonology* 10 (1). 4. DOI: 10.5334/labphon.147.
- Rose, Ralph L. (2015): Um and uh as differential delay markers: the role of contextual factors. In: Eklund, Robert (Hrsg.): *Proceedings of Disfluency in Spontaneous Speech (DiSS). The 7th Workshop on Disfluency in Spontaneous Speech*. 73–76.
- Sacks, Harvey; Emanuel A. Schegloff & Gail Jefferson (1974): A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4). 696–735. DOI: 10.2307/412243.
- Sauer, Simon & Anke Lüdeling (2016): Flexible multi-layer spoken dialogue corpora. *International Journal of Corpus Linguistics* 21 (3). 419–438. DOI: 10.1075/ijcl.21.3.06sau.
- Schachter, Stanley; Nicholas Christenfeld; Bernard Ravina & Frances Bilous (1991): Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology* 60 (3). 362–367. DOI: 10.1037/0022-3514.60.3.362.
- Schegloff, Emanuel A. (1982): Discourse as an interactional achievement: Some uses of ‘uh huh’ and other things that come between sentences. In: Tannen, Deborah (Hrsg.): *Analyzing Discourse: Text and Talk*. Washington D.C.: Georgetown University Press. 71–93.
- Schegloff, Emanuel A. (2010): Some Other “Uh(m)”s. *Discourse Processes* 47 (2). 130–174. DOI: 10.1080/01638530903223380.

- Schegloff, Emanuel A.; Gail Jefferson & Harvey Sacks (1977): The preference for self-correction in the organization of repair in conversation. *Language* 53 (2). 361–382. DOI: 10.2307/413107.
- Schmidt, Jürgen Erich (2001): Bausteine der Intonation? *Germanistische Linguistik* 157-158. 9–32.
- Schnadt, Michael J. & Martin Corley (2006): The Influence of Lexical, Conceptual and Planning Based Factors on Disfluency Production. In: *Proceedings of the twenty-eighth meeting of the Cognitive Science Society*. 750–755. URL: <http://csjarchive.cogsci.rpi.edu/Proceedings/2006/docs/p750.pdf> (besucht am 10.01.2017).
- Schneider, Ulrike (2014): Frequency, Chunks and Hesitations. A Usage-Based Analysis of Chunking in English. Dissertation. Freiburg: Albert-Ludwigs-Universität Freiburg. DOI: 10.6094/978-3-928969-57-4.
- Schönle, Paul-Walter & Bastian Conrad (1985): Hesitation vowels: a motor speech respiration hypothesis. *Neuroscience Letters* 55. 293–296. DOI: 10.1016/0304-3940(85)90451-3.
- Schweitzer, Antje & Natalie Lewandowski (2013): Convergence of Articulation Rate in Spontaneous Speech. In: *Proceedings of Interspeech*. 525–529.
- Schwitalla, Johannes (2002): Kleine Wörter. Partikeln im Gespräch. In: Dittmann, Jürgen & Claudia Schmidt (Hrsg.): *Über Wörter. Grundkurs Linguistik*. Freiburg im Breisgau: Rombach. 259–283.
- Schwitalla, Johannes (2012): *Gesprochenes Deutsch: eine Einführung*. 4. Aufl. (= Grundlagen der Germanistik, Bd. 33). Berlin: Erich Schmidt Verlag.
- Segalowitz, Norman (2010): *Cognitive bases of second language fluency*. (= Cognitive science and second language acquisition series). New York: Routledge.
- Séguin, Marie-Claude (2008): Analyse formantique des pauses remplies chez les adolescents unilingues en outaouais québécois: une étude pilote. In: *Actes du congrès annuel de l'Association canadienne de linguistique*.
- Selting, Margret (1996): On the interplay of syntax and prosody in the constitution of turn-constructural units and turns in conversation. *Pragmatics* 6 (3). 371–388. DOI: 10.1075/prag.6.3.06sel.
- Selting, Margret (2000): The construction of units in conversational talk. *Language in Society* 29 (4). 477–517. DOI: 10.1017/S0047404500004012.

- Selting, Margret (2007): 'Grammatik des gesprochenen Deutsch' im Rahmen der Interaktionalen Linguistik. In: Ágel, Vilmos & Mathilde Hennig (Hrsg.): *Zugänge zur Grammatik der gesprochenen Sprache*. (= Reihe Germanistische Linguistik). Tübingen: Max Niemeyer Verlag. 99–136.
- Selting, Margret; Peter Auer; Dagmar Barth-Weingarten; Jörg Bergmann; Pia Bergmann; Karin Birkner; Elizabeth Couper-Kuhlen; Arnulf Deppermann; Peter Gilles; Susanne Günthner; Martin Hartung; Friederike Kern; Christine Mertzlufft et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10. 353–402.
- Seyfeddinipur, Mandana; Sotaro Kita & Peter Indefrey (2008): How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition* 108 (3). 837–842. DOI: 10.1016/j.cognition.2008.05.004.
- Shan, Guogen & Shawn Gerstenberger (2017): Fisher's exact approach for post hoc analysis of a chi-squared test. *PLoS ONE* 12 (12). e0188709. DOI: 10.1371/journal.pone.0188709.
- Shattuck-Hufnagel, Stefanie & Alice E. Turk (1996): A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research* 25 (2). 193–247. DOI: 10.1007/BF01708572.
- Shriberg, Elizabeth E. (1994): Preliminaries to a Theory of Speech Disfluencies. Unveröffentlichte Dissertation. Berkeley: University of California.
- Shriberg, Elizabeth E. (2001): To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31 (1). 153–169. DOI: 10.1017/S0025100301001128.
- Shriberg, Elizabeth E. & Robin J. Lickley (1993): Intonation of clause-internal filled pauses. *Phonetica* 50 (3). 172–179. DOI: 10.1159/000261937.
- Siegel, Gerald M.; Joanne Lenske & Patricia Broen (1969): Suppression of normal speech disfluencies through response cost. *Journal of Applied Behavior Analysis* 2 (4). 265–276. DOI: 10.1901/jaba.1969.2-265.
- Simpson, Adrian P. (2013): Spontaneous Speech. In: Jones, Mark J. & Rachael-Anne Knight (Hrsg.): *The Bloomsbury Companion to Phonetics*. (= Bloomsbury Companions). London & New York: Bloomsbury. 155–169.
- Smith, Vicki L. & Herbert H. Clark (1993): On the Course of Answering Questions. *Journal of Memory and Language* 32 (1). 25–38. DOI: 10.1006/jmla.1993.1002.

- Stenström, Anna-Brita (1990): Pauses in monologue and dialogue. In: Svartvik, Jan (Hrsg.): *The London-Lund Corpus of Spoken English. Description and Research.* (= Lund Studies in English). Lund: Lund University Press. 211–252.
- Stepanova, Svetlana (2007): Some features of filled hesitation pauses in spontaneous Russian. In: *Proceedings of ICPHS XVI.* 1325–1328.
- Stevens, Kenneth N. (1997): Articulatory-Acoustic-Auditory Relationships. In: Hardcastle, William J. (Hrsg.): *The handbook of phonetic sciences.* (= Blackwell handbooks in linguistics). Oxford, UK: Blackwell. 472–507.
- Stirling, Lesley; Janet Fletcher; Ilana Mushin & Roger Wales (2000): Representational issues in annotation: Using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication* 33. 113–134. DOI: 10.1016/S0167-6393(00)00072-8.
- Stutterheim, Christiane von (1997): *Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte.* (= Reihe Germanistische Linguistik, Bd. 184). Tübingen: Max Niemeyer Verlag.
- Styler, Will (2017): On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America* 142 (4). 2469–2482. DOI: 10.1121/1.5008854.
- Svartvik, Jan, Hrsg. (1990): *The London-Lund Corpus of Spoken English. Description and Research.* Bd. 82. (= Lund Studies in English, Bd. 82). Lund: Lund University Press.
- Swerts, Marc (1998): Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30. 485–496.
- Swerts, Marc & Ronald Geluykens (1994): Prosody as a marker of information flow in spoken discourse. *Language and Speech* 37 (1). 21–43. DOI: 10.1177/002383099403700102.
- Szmrecsanyi, Benedikt (2010): The English genitive alternation in a cognitive sociolinguistics perspective. In: Geeraerts, Dirk; Gitte Kristiansen & Yves Peirsman (Hrsg.): *Advances in Cognitive Sociolinguistics.* Berlin, New York: De Gruyter Mouton. 139–166.
- Tannenbaum, Percy H.; Frederick Williams & Carolyn S. Hillier (1965): Word Predictability in the Environments of Hesitations. *Journal of verbal learning and verbal behavior* 4 (2). 134–140. DOI: 10.1016/S0022-5371(65)80097-4.

- Todt, Dietmar (1981): Zum Auftreten von Fülllauten in spontan gesprochenen Berichten. *Nova acta Leopoldina* 54 (245). 597–611.
- Tottie, Gunnel (2011): *Uh* and *um* as sociolinguistic markers in British English. *International Journal of Corpus Linguistics* 16. 173–196. DOI: 10.1075/ijcl.16.2.02tot.
- Tottie, Gunnel (2013): *Uh* and *Um* as sociolinguistic markers in British English. In: Gilquin, Gaëtanelle & Sylvie de Cock (Hrsg.): *Errors and disfluencies in spoken corpora*. (= Benjamins Current Topics). Amsterdam & Philadelphia: John Benjamins. 33–57.
- Tottie, Gunnel (2015): Turn management and the fillers *uh* and *uhm*. In: Aijmer, Karin & Christoph Rühlemann (Hrsg.): *Corpus Pragmatics. A Handbook*. Cambridge, U.K.: Cambridge University Press. 381–407.
- Trouvain, Jürgen (2015): On clicks in German. In: Leemann, Adrian; Marie-José Kolly; Stephan Schmid & Volker Dellwo (Hrsg.): *Trends in phonetics and phonology. Studies from German-speaking Europe*. Bern: Peter Lang. 21–34.
- Trouvain, Jürgen & Zofia Malisz (2016): Inter-Speech Clicks in an Interspeech Keynote. In: *Interspeech*. 1397–1401. DOI: 10.21437/Interspeech.2016-1064.
- Tschäpe, Nadja; Jürgen Trouvain; Dominik Bauer & Michael Jessen (2005): Idiosyncratic patterns of filled pauses. In: *14th Annual Conference of the International Association for Forensic Phonetics and Acoustics*. 25.
- Vaissière, Jacqueline (1983): Language-independent prosodic features. In: Cutler, Anne & D. Robert Ladd (Hrsg.): *Prosody: Models and Measurements*. (= Springer Series in Language and Communication). Berlin, Heidelberg: Springer. 53–66.
- van Rij, Jacolien; Martijn Wieling; R. Harald Baayen & Hedderik van Rijn (2017): *itsadug. Interpreting Time Series and Autocorrelated Data Using GAMMs*. Version 2.3.
- Vasilescu, Ioana; Maria Candea & Martine Adda-Decker (2005): Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages. In: *Interspeech*.
- Walker, Esther J.; Evan F. Risko & Alan Kingstone (2014): Fillers as Signals: Evidence From a Question–Answering Paradigm. *Discourse Processes* 51 (3). 264–286. DOI: 10.1080/0163853X.2013.862478.

- Ward, Nigel (2006): Non-lexical conversational sounds in American English. *Pragmatics & Cognition* 14 (1). 129–182. DOI: 10.1075/pc.14.1.08war.
- Watanabe, Michiko; Keikichi Hirose; Yasuharu Den & Nobuaki Minematsu (2008): Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication* 50. 81–94. DOI: 10.1016/j.specom.2007.06.002.
- Weirich, Melanie & Adrian P. Simpson (2014): Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics* 43. 1–10. DOI: 10.1016/j.wocn.2014.01.001.
- Wichmann, Anne (2008): Speech corpora and spoken corpora. In: Lüdeling, Anke & Merja Kytö (Hrsg.): *Corpus Linguistics. An International Handbook*. (= Handbücher zur Sprach- und Kommunikationswissenschaft). Berlin: Walter de Gruyter. 187–207.
- Wickham, Hadley (2009): *ggplot2. Elegant Graphics for Data Analysis*. New York: Springer. URL: <https://ggplot2.tidyverse.org/> (besucht am 28.08.2019).
- Wieling, Martijn (2018): Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* (70). 86–116. DOI: 10.1016/j.wocn.2018.03.002.
- Wieling, Martijn; Jack Grieve; Gosse Bouma; Josef Fruehwald; John Coleman & Mark Liberman (2016): Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change* 6 (2). 199–234. DOI: 10.1163/22105832-00602001.
- Wiese, Heike (2010): Kiezdeutsch. *APuZ – Aus Politik und Zeitgeschichte (Themenband: "Sprache")* 8. 33–38.
- Willkop, Eva-Maria (1988): *Gliederungspartikeln im Dialog*. (= Studien Deutsch, Bd. 5). München: Iudicium.
- Wingate, Marcel E. (1984): Fluency, disfluency, dysfluency, and stuttering. *Journal of Fluency Disorders* 9 (2). 163–168. DOI: 10.1016/0094-730X(84)90033-0.
- Winkelmann, Raphael (2015): *PraatToFormants2AsspDataObj.R*. URL: <https://gist.github.com/raphywink/2512752a1efa56951f04> (besucht am 07.03.2017).

- Winkelmann, Raphael; Jonathan Harrington & Klaus Jänsch (2017): EMU-SDMS. Advanced speech database management and analysis in R. *Computer Speech & Language*. 392–410. DOI: 10.1016/j.cs1.2017.01.002.
- Winkelmann, Raphael; Klaus Jaensch; Steve Cassidy & Jonathan Harrington (2016): *emuR: Main Package of the EMU Speech Database Management System*. Version 0.2.1 (R package).
- Wood, Simon N. (2011): Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73 (1). 3–36. DOI: 10.1111/j.1467-9868.2010.00749.x.
- Wood, Simon N. (2017): *Generalized Additive Models. An Introduction with R*. 2. Aufl. Chapman and Hall/CRC.
- Wood, Simon N. (2019): *mgcv. Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. Version 1.8.28. URL: <https://cran.r-project.org/package=mgcv> (besucht am 23.09.2019).
- Xie, Yihui (2019): *knitr. A General-Purpose Package for Dynamic Report Generation in R*. Version 1.23.
- Zeileis, Achim & Gabor Grothendieck (2005): zoo. S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software* 14 (6). 1–27. DOI: 10.18637/jss.v014.i06.
- Zeldes, Amir (2012): *Productivity in argument selection. From morphology to syntax*. (= Trends in linguistics. Studies and monographs, Bd. 260). Berlin: De Gruyter Mouton.
- Zipf, George Kingsley (1949): *Human behavior and the principle of least effort. An introduction to human ecology*. Cambridge, Mass: Addison-Wesley.