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 enthalten „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 Regiolechts. 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 Repostitorium 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 Diapixe *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/f) 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 gehunkt (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	dipl ^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

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

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

Abbildung 4.1: Nicht-glottale Füllpartikel von Versuchsperson f1 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 *dipl* mit *<knarren>*; auf *fp* mit *as* (linker Kontext, Segment), *fg* und *pr* (rechter Kontext, überwiegend stille Pause); auf *segn* mit *o* (Laut nach SAMPA), *GT* (glottale Transitionen antezedenter Segmente in den *fv*-Vokal), *f* (Laut nach SAMPA), *—* (Laut nach SAMPA), *j* (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 unbestimmbare 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 Grenztönen 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 Grenztönen zeigen. Wie in Abbildung 4.6 liegt der Grenztönen 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_fm1_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 Larnygalisierung, 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 initiiierende 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 turnkonstruktionale 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, 180ms⟩ hier im Campus no nie in der Mensa gegessen weil ich eignlich in *beep* bin)_{IP}
 c. A: #(mhm)_{IP}
 d. B: (an dem Campus dort)_{IP}
 e. A: (mhm)_{IP}
 f. B: (deshalb)_{IP} ⟨P, 860ms⟩ (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} ⟨P340ms⟩ (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} ⟨286ms⟩ (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 dia	DZ		DZ						DZ...	
A ip	IP		IP		IP			IP	IP	
A dipl	Wie...		mhm		mhm			mhm	also...	
B dia		DZ								
B ip		IP		IP		IP	IP			
B dipl		äh...		an...		deshalb	das...			

	5h	5i	5j	5j	5k	5l	5m	
A dia	...DZ							
A ip	IP		IP	IP		IP		
A dipl	positiv...		von...	also...	mein...		un...	
B dia		DZ			DZ		DZ	
B ip		IP			IP		IP	
B dipl		ja?			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 nicht mehr in der Mensa weil ich finde das is nicht)_{ip} (also)_{ip}
 <P, 100 ms> (der Geruch un so viele Menschen ich mein des)_{ip} <P, 196 ms> (ess
 ich ja nicht 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}

(BeDiaCoC: 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 BeDiaCoC 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}
 (BeDiaCoC: 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 ähm*)_{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 fünfmünütige 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.

Evaluiert 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 Intonationsphasengrenze 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 *dip1* 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 *vowel*-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 geäußerter 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	[ʏ]
	Bube	[u:]	Beete	[e:]	Böcke	[œ]
	Kippe	[ɪ]	böte	[ø:]	Pocke	[ɔ]
	Puppe	[ʊ]	Bote	[o:]	Tage	[a:]
			bäte	[ɛ:]	packe	[a]
			Kette	[ɛ]		
Final mit [ɐ]	Pieper	[i:]	Güter	[y:]	Bäcker	[ɛ]
	Tupper	[ʊ]	Puder	[u:]	Höcker	[œ]
	Kaper	[a:]	Mütter	[ʏ]	Packer	[a]
	Geber	[e:]	Köder	[ø:]	Kicker	[ɪ]
			Toter	[o:]		
			Täter	[ɛ:]		
			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-Spektren (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 Formantentraktorien. 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 irgendwo 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.