

# **Form und Funktion prosodischer Grenzen im Gespräch**

Dissertation  
zur Erlangung des Doktorgrades  
der Philosophischen Fakultät  
der Christian-Albrechts-Universität  
zu Kiel

vorgelegt von  
Benno Peters

Kiel 2006

Erstgutachter: Prof. K. J. Kohler

Zweitgutachter: Prof. J. M. Harrington

Tag der mündlichen Prüfung: 12. 07. 2006

Durch den zweiten Prodekan Prof. N. Nübler

zum Druck genehmigt am: 12. 12. 2006

## **Danksagung**

Prof. Klaus J. Kohler danke ich sehr herzlich, weil er mich als akademischer Lehrer, Projektleiter und Doktorvater über viele Jahre mit fachlichem Wissen, konstruktivem Rat, finanzieller Förderung und viel Geduld begleitet hat.

Ich bedanke mich auch bei allen Kollegen des Instituts für Phonetik und digitale Sprachverarbeitung für die schöne Zeit am IPDS in Kiel. Eine besondere Unterstützung bei dieser Arbeit waren Ernst Dombrowski, Tamara Khromovskikh, Felicitas Kleber, Oliver Niebuhr, Dr. Michel Scheffers und Thomas Wesener durch ihre fachlichen Anregungen und die Hilfe bei der Datenverarbeitung.

Mein Dank gilt weiterhin der Deutschen Forschungsgemeinschaft. Durch die Finanzierung der Projekte *Lautmuster deutscher Spontansprache* (Ko 331/22) und *Prosodische Phrasierung in deutscher Spontansprache* (Ko 331/23) wurde diese Arbeit möglich.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b> .....	1
<b>2</b>	<b>Methoden</b> .....	8
2.1	Datenbasis .....	11
2.1.1	Das <i>Video Task</i> Szenario .....	15
2.1.2	Umfang der Sprachdaten .....	18
2.1.3	Aufbereitung der Sprachdaten .....	19
2.1.3.1	Transliteration .....	20
2.1.3.2	Automatische kanonische Transkription .....	20
2.1.3.3	Segmentelle Etikettierung .....	21
2.1.3.4	Prosodische Etikettierung .....	21
2.1.3.5	Erweiterte Etikettierung in den <i>Video Task</i> Daten .....	25
2.1.3.6	KielDat Datenbank .....	36
2.1.4	Zusammenfassung der Datenbasis .....	36
<b>3</b>	<b>Turninterne prosodische Grenzen in flüssiger Sprache</b> .....	37
3.1	Stand der Forschung .....	37
3.1.1	Grundlegende Beschreibungseinheiten .....	37
3.1.2	Phonologische Beschreibungen .....	39
3.1.3	Phonetische Beschreibungen .....	47
3.1.4	Form-Funktions-Beziehungen .....	51
3.1.4.1	Melodische Muster .....	56
3.1.4.2	Pausen .....	60
3.1.4.3	Segmentelle Dauerstrukturen .....	65
3.2	Symbolbasierte Analyse der Produktionsdaten .....	69
3.2.1	Methoden der automatischen Merkmalsextraktion .....	69

3.2.1.1	Symbolbasierte Extraktion suprasegmenteller Merkmale. . . . .	70
3.2.1.2	Symbolbasierte und numerische Extraktion von Pausen, Atmen und phonatorischen Veränderungen . . . . .	71
3.2.1.3	Numerische Extraktion der Längungsinformation. . . . .	71
3.2.2	Teilkorpora . . . . .	73
3.2.3	Pausen und Atmungsphasen an turninternen prosodischen Grenzen . . . .	75
3.2.4	Pausen und Atmungsphasen im Inneren prosodischer Phrasen. . . . .	77
3.2.5	Melodische Muster an turninternen prosodischen Grenzen . . . . .	77
3.2.6	Melodische Muster im Inneren prosodischer Phrasen . . . . .	79
3.2.7	F0-Reset an turninternen prosodischen Grenzen. . . . .	80
3.2.8	F0-Reset im Inneren prosodischer Phrasen . . . . .	81
3.2.9	Längung an turninternen prosodischen Grenzen. . . . .	82
3.2.10	Längung im Inneren prosodischer Phrasen . . . . .	84
3.2.11	Phonation an turninternen prosodischen Grenzen . . . . .	84
3.2.12	Diskussion zu den Einzelmerkmalen an turninternen prosodischen Grenzen. . . . .	86
3.2.13	Merkmalsbündel an turninternen prosodischen Grenzen. . . . .	87
3.3	Experimentelle Untersuchungen. . . . .	89
3.3.1	Stand der Forschung . . . . .	90
3.3.2	Perzeptionsexperimente zur phonetischen Kohäsion an turninternen prosodischen Grenzen (Experiment 1a, 1b und 1c) . . . . .	100
3.3.2.1	Experiment 1a . . . . .	101
3.3.2.1.1	Hypothesen . . . . .	101
3.3.2.1.2	Stimulusmaterial . . . . .	102
3.3.2.1.3	Durchführung . . . . .	106
3.3.2.1.4	Ergebnisse . . . . .	107
3.3.2.1.5	Diskussion . . . . .	111
3.3.2.2	Experiment 1b. . . . .	112
3.3.2.2.1	Stimulusmaterial und Durchführung . . . . .	112
3.3.2.2.2	Ergebnisse . . . . .	113
3.3.2.2.3	Diskussion zu Experiment 1a und 1b . . . . .	116

3.3.2.4	Experiment 1c.....	117
3.3.2.4.1	Hypothesen.....	118
3.3.2.4.2	Stimulusmaterial.....	118
3.3.2.4.3	Durchführung.....	121
3.3.2.4.4	Ergebnisse.....	122
3.3.2.5	Diskussion der Ergebnisse aus den Experimenten 1a, 1b und 1c.....	125
3.4	Interpretative Analysen.....	127
3.4.1	Stand der Forschung.....	128
3.4.2	Ergebnisse der Interpretation von Korpusbeispielen.....	135
3.4.2.1	Starke tonale Bewegungen.....	138
3.4.2.2	F0-Reset.....	139
3.4.2.3	Längung des finalen Reims.....	139
3.4.2.4	Zusammenfassung.....	140
3.5	Klassifikationssystem I.....	141
3.5.1	Verteilungen der prosodischen Grenzen über die inhaltlichen Klassen für das Klassifikationssystem I.....	142
3.5.2	Die nicht-spezifizierten Merkmale Längung und Reset in System I.....	143
<b>4</b>	<b>Prosodische Grenzen und Sprecherwechsel.....</b>	<b>145</b>
4.1	Stand der Forschung.....	147
4.1.1	Der Turn als Einheit.....	149
4.1.2	Typen von Sprecherwechseln und deren phonetische Beschreibung.....	160
4.1.3	Perzeptionsexperimente.....	163
4.2	Symbolbasierte Analyse der Produktionsdaten.....	167
4.2.1	Einzelmerkmale an turnfinalen prosodischen Grenzen.....	168
4.2.1.1	Pausen und Atmungsphasen.....	168
4.2.1.2	Melodische Muster.....	170
4.2.1.3	Segmentelle Längung.....	174
4.2.1.4	Phonation.....	175
4.2.1.5	Merkmalsbündel an turnfinalen prosodischen Grenzen.....	178

4.2.1.6	Diskussion .....	179
4.3	Perzeptionsexperiment zur Erwartung von Turnübergaben an prosodischen Grenzen (Experiment 2).....	181
4.3.1	Hypothesen.....	184
4.3.2	Stimulusmaterial .....	185
4.3.3	Durchführung .....	189
4.3.4	Ergebnisse .....	190
4.3.5	Zusammenfassung und Diskussion .....	195
4.4	Interpretation von Korpusbeispielen .....	197
4.4.1	Ergebnisse .....	199
4.4.2	Diskussion .....	200
4.5	Vergleich zwischen turnfinalen und turninternen prosodischen Grenzen. .	201
4.6	Zusammenfassung und abschließende Diskussion zu den turnfinalen prosodischen Grenzen .....	204
4.7	Klassifikationssystem II - Turnfinale prosodische Grenzen .....	209
<b>5</b>	<b>Prosodische Grenzen in unflüssiger Sprache .....</b>	<b>211</b>
5.1	Symbolbasierte Analyse der Produktionsdaten .....	214
5.1.1	Melodische Muster .....	217
5.1.2	Pausen und Atmungsphasen .....	218
5.1.3	Längung .....	219
5.1.4	Phonation .....	221
5.1.5	Zusammenfassung .....	224
5.2	Perzeptionsexperimente zur Sprechflüssigkeit (Experiment 3a und 3b) ..	224
5.2.1	Experiment zur Wahrnehmung unflüssiger Sprechweise (Experiment 3a). 225	
5.2.1.1	Hypothesen.....	225
5.2.1.2	Stimulusmaterial .....	226
5.2.1.3	Durchführung .....	227
5.2.1.4	Ergebnisse .....	229
5.2.1.5	Zusammenfassung und Diskussion .....	234

5.2.2	Experiment zur Wahrnehmung unflüssiger Sprechweise an unterschiedlichen syntaktischen Strukturstellen (Experiment 3b) . . . . .	235
5.2.2.1	Hypothesen . . . . .	235
5.2.2.2	Stimulusmaterial und Durchführung . . . . .	236
5.2.2.3	Ergebnisse . . . . .	236
5.2.2.4	Zusammenfassung und Diskussion . . . . .	238
5.3	Zusammenfassung und Diskussion zu den unflüssigen prosodischen Grenzen . . . . .	239
<b>6</b>	<b>Abschließende Diskussion und Ausblick . . . . .</b>	<b>241</b>
<b>7</b>	<b>Literatur . . . . .</b>	<b>250</b>
	<b>Anhang . . . . .</b>	<b>260</b>

# 1 Einleitung

Die vorliegende Arbeit leistet einen Beitrag zur funktionsorientierten Phonetik. Es werden die Beziehungen zwischen phonetischen Formen prosodischer Grenzen und deren sprachlichen Funktionen im Gespräch empirisch untersucht. Ausgangspunkt der Untersuchungen ist, dass längere Äußerungen durch phonetische Mittel syntagmatisch gegliedert werden. Hierbei entstehen Einheiten, die als prosodische Phrasen (häufig auch als Intonationsphrasen) bezeichnet werden. Diese Einheiten zeigen an ihren Grenzen Komplexe phonetischer Merkmale, die zu perzeptiv relevanten Einschnitten führen und gliedernde Funktion haben. Die Gliederungsfunktion prosodischer Grenzen wird hinsichtlich ihrer phonetischen Form und ihrer sprachlichen Bedeutung anhand von Produktions- und Perzeptionsdaten untersucht.

Die Analyse der sprachlichen Funktion phonetischer Parameter erfolgt innerhalb eines komplexen Forschungsansatzes, der verschiedene Informationsebenen einbezieht, die in natürlichen Gesprächen eine Rolle spielen. Prosodische Grenzen werden in ihrer Beziehung zur syntaktischen und semantischen Struktur ebenso untersucht, wie hinsichtlich ihrer interaktionalen Funktionen in der Gesprächssteuerung. Des Weiteren werden Verbindungen zwischen prosodischen Einschnitten und dem Prozess der Sprachplanung gezogen. Hiermit rückt das Gespräch als komplexes und interaktives menschliches Verhalten ins Zentrum der Betrachtung.

Sowohl aus einer historischen Perspektive als auch für den Aufbau dieser Studie, bildet die Gliederung flüssiger und grammatisch wohlgeformter Sprache in prosodische Phrasen den Ausgangspunkt der Untersuchung. Diese prosodischen Phrasen sind durch prosodische Grenzen voneinander abgegrenzt und stehen in enger Beziehung zur grammatischen und inhaltlichen Struktur von Äußerungen. Durch eine Vielzahl von Untersuchungen ist belegt, dass die Position prosodischer Grenzen häufig mit hohen semantischen und syntaktischen Grenzen zusammenfällt. Diese semantisch und syntaktisch motivierte Gliederung von Äußerungen in prosodische Phrasen ist der Phänomenbereich, der üblicherweise als prosodische Phrasierung beschrieben wird und in der intonatorischen Analyse jeder Sprache von zentralem Interesse ist, da die Gliederung in prosodische Phrasen der Formung von Bedeutung und Funktion einer Äußerung dient und so zur deren Verständlichkeit beiträgt.

Es werden mit Bezug auf Sprachproduktion und -perzeption zwei phonetische Mechanismen herausgearbeitet, die für die prosodische Gliederung ausschlaggebend sind. Erstens die phonetische *Kohäsion* (Couper-Kuhlen 1983; Kohler 1983; Selting 1995; Gilles 2005) im Inneren prosodischer Phrasen, die diese als perzeptorische Einheiten erscheinen lässt und zweitens die phonetische *Separierung*, die perzeptiv relevante prosodische Grenzen entstehen lässt und so zu einer Abgrenzung der Einheiten führt. Diese phonetischen Konzepte werden mit der inhaltlichen Struktur von Äußerungen verbunden, indem über Korpusanalysen und Perzeptionsexperimente gezeigt wird, dass inhaltliche Separierung zwischen benachbarten Einheiten durch phonetische Separierung signalisiert wird. Inhaltliche und syntaktische Kohäsion hingegen spiegelt sich in kohäsiven phonetischen Strukturen wider.

Phonetische Merkmale der Separierung können einzeln oder in Bündeln auftreten, wobei die einzelnen Merkmale in ihrer Ausprägung ebenso variieren wie die Merkmalsbündel in ihrer Zusammensetzung. Entsprechend der Ausprägung der Einzelmerkmale und ihrer unterschiedlichen Bündelung ergeben sich Grade der Separierung zwischen aufeinander folgenden prosodischen Phrasen. Diese graduelle Abstufung kann auf einer Kohäsionsskala abgebildet werden, die von starker Separierung bis zu vollständiger Integration aufeinander folgender Sprechabschnitte reicht. Aus den unterschiedlichen Graden von Kohäsion bzw. Separierung ergeben sich prosodische Grenzen verschiedener Stärke. Folgende separierende Merkmale und deren Bündelungen werden untersucht:

- melodische Bewegungen
- segmentelle Dauerrelationen
- Pausen und Atmungsphasen
- F0-Reset
- phonatorische Veränderungen

Neben ihrer Anbindung an inhaltliche und syntaktische Strukturen stehen prosodische Grenzen aber auch in enger funktionaler Beziehung zur interaktionalen Struktur von Gesprächen. Die Merkmalsbündel an prosodischen Grenzen haben einen hohen Signalwert bei der Gliederung von Gesprächen in die Abfolge von Äußerungen der beteiligten Sprecher. So verwenden Sprecher am Ende von Redebeiträgen bestimmte phonetische

Merkmale, um anzuzeigen, dass der Redebeitrag (*Turn*) beendet ist, oder aber andere Merkmale, wenn sie signalisieren wollen, dass ein Redebeitrag fortgesetzt werden soll.

Neben der inhaltlichen Gliederungsfunktion prosodischer Grenzen innerhalb von Turns und ihrer Bedeutung für die Dialogsteuerung gibt es noch einen dritten Bereich, in dem separierende phonetische Merkmalskomplexe eine Rolle spielen: die unflüssige Sprechweise. Gespräche werden während des Sprechens geplant und strukturiert. Während der Planung und Ausführung von Redebeiträgen kommt es oft zu Problemen und Verzögerungen, die sich auf der phonetischen Ebene z.B. durch Pausen oder lokale Veränderungen in der Dauerstruktur zeigen und ebenfalls als prosodische Grenzen beschrieben werden können. Die so entstehenden diskontinuierlichen phonetischen Strukturen in Äußerungen spiegeln den mentalen Planungsvorgang des Sprechenden wider und haben so eine wichtige Funktion im Gespräch, da Hörer hierdurch die sprecherseitigen Planungsprozesse nachvollziehen und darauf reagieren können, z.B. indem sie ein vom Sprecher gesuchtes Wort ergänzen oder indem sie dem Sprecher Zeit zum Nachdenken lassen, anstatt das Rederecht zu übernehmen. Die entstehenden unflüssigen Grenzen decken sich oft nicht mit inhaltlichen oder syntaktischen Grenzen und befinden sich hierdurch z.T. im Inneren der Einheiten, die in der Literatur als prosodische Phrasen (oder Intonationsphrasen) bezeichnet werden; sie zerteilen die Einheiten wohlgeformter Sprache in prosodische Fragmente.

Viele Forschungsansätze thematisieren das Problem prosodischer Fragmente nicht. Die Erklärung hierfür ist einfach. Fast alle Ansätze gehen nicht von einer spontansprachlichen Datenbasis aus, sondern arbeiten - wenn überhaupt mit Sprechdaten – dann mit Lesesprache, oft mit konstruierten Beispielsätzen. In solchen Datengrundlagen ist mit einer weitgehenden Übereinstimmung zwischen inhaltlichen, syntaktischen und prosodischen Strukturen zu rechnen, und das Vorkommen prosodischer Einschnitte, die durch den Sprachplanungsprozess entstehen, ist unwahrscheinlich, da Lesesprache nicht vom Sprecher geplant, sondern abgelesen wird. Die vorliegende Arbeit hingegen basiert auf einer spontansprachlichen Datenbasis und hat zum Ziel, auch solche prosodisch signalisierten Grenzen zu beschreiben, die durch Planungsprobleme verursacht werden und keine kompletten prosodischen Phrasen delimitieren.

Das grundlegende Problem von Form-Funktions-Zuordnungen in spontansprachlichen Korpora ist die große Variabilität phonetischer Strukturen und die Überschneidung

der phonetischen Exponenten verschiedener funktionaler Kategorien. Dies gilt auch für die Merkmale prosodischer Grenzen. So können z.B. Pausen sowohl an intendierten prosodischen Grenzen zur Markierung inhaltlicher Einschnitte vorkommen als auch bei Planungsproblemen mitten in einer syntaktischen Einheit. Zur Bestimmung der funktionalen Relevanz phonetisch unterschiedlicher prosodischer Grenzen wird deshalb vor allem die Bündelung von Merkmalen herangezogen. Die Analyse eines umfangreichen spontansprachlichen Korpus soll Gemeinsamkeiten und Unterschiede flüssiger und unflüssiger prosodischer Grenzen innerhalb und am Ende von Redebeiträgen aufzeigen und so den phonetischen Signalisierungswert unterschiedlicher Merkmalsbündel innerhalb von Gesprächen prüfen. Hierdurch wird ein direkter Bezug zu Hörer- und Sprecherverhalten hergestellt, denn für den reibungslosen Ablauf von Gesprächen ist es sehr wichtig, dass Hörer richtig einschätzen, zu welchem der drei Phänomenbereiche eine wahrgenommene, vom Sprecher produzierte Separierung zuzuordnen ist. Handelt es sich um eine inhaltlich gliedernde äußerungsinterne prosodische Grenze, wird der Sprecher mit seiner Rede fortfahren und das vorher Gesagte weiter ausführen. Die Grenze muss vom Hörer also als inhaltsbezogenes Gliederungssignal interpretiert werden. Liegt eine Produktionsstörung vor, so kommt es unter Umständen zu einer Korrektur des Gesagten, oder der Sprecher braucht etwas Zeit, um die Fortsetzung der Äußerung zu planen. Eine solche Grenze muss als planungsbezogenes Gliederungssignal aufgefasst werden. Wenn die Separierung das Ende eines Gesprächsbeitrags signalisieren soll, muss der Hörer dies erkennen, die Grenze also als gesprächsbezogenes Gliederungssignal deuten und die Sprecherrolle einnehmen. Die grundlegende Annahme dieser Arbeit ist, dass jeder der drei Phänomenbereiche ein eigenes phonetisch kodierte Signalisierungssystem aufweist, das es Hörern ermöglicht, die Funktion einer wahrgenommenen Separierung zu erkennen und hieraus eine Interpretation hinsichtlich des erwarteten Fortgangs des Gesprächs abzuleiten.

Das Ziel der Arbeit ist also die Analyse der phonetischen Merkmalskomplexe der drei Bereiche, deren Abgrenzung voneinander und die perzeptiv orientierte, empirisch basierte Beschreibung und Klassifizierung phonetischer Separierung und ihrer sprachlichen Funktionen. Diese Schritte bieten Ansätze zur Modellierung des Phänomenbereichs der prosodischen Grenzen.

Natürlich spielt für Hörer bei der Interpretation von Äußerungen die prosodische Information ihre Rolle immer im Gesamtkontext der Äußerung. Sprecher und Hörer verarbeiten parallel auch lexikalische, syntaktische sowie mimische und gestische Information. Hier soll vor allem der Beitrag der Phonetik systematisch erfasst werden. Um dieses Ziel zu erreichen, wird eine Kombination von verschiedenen methodischen Herangehensweisen angewendet, die sich gegenseitig ergänzen und sowohl Sprachproduktionsdaten also auch Perzeptionsdaten einbeziehen. Den Kern der empirischen Analyse bildet ein Korpus segmentell und prosodisch etikettierter spontansprachlicher Dialoge, das hinsichtlich der Merkmalsbündelungen an wahrnehmbaren prosodischen Grenzen statistisch ausgewertet wird. Die im Korpus gefundenen Merkmalskomplexe werden dann in systematischer Variation in Stimulusmaterial für Perzeptionsexperimente eingearbeitet und von Versuchspersonen hinsichtlich ihrer separierenden Wirkung und ihrer Funktion im Dialog beurteilt. Hinzu kommt die funktionale Interpretation vieler Datenausschnitte aus dem Korpus. So werden verschiedene separierende Merkmalsbündel auf einer perzeptorischen Skala von Kohäsion zu Separierung abgebildet und in Verbindung zur semantisch-pragmatischen und syntaktischen Struktur von Äußerungen gesetzt bzw. als Korrelate von Produktionsstörungen oder der Dialogsteuerung eingeordnet.

Die Arbeit hat folgenden Aufbau: Als erstes folgt ein Kapitel zu den verwendeten Methoden, das eine Beschreibung der Datenbasis enthält. Diese Beschreibung ist relativ umfangreich. Zum einen, weil die Erhebung und Aufbereitung der Sprachdaten z.T. mit einem neuen, bisher nicht vollständig beschriebenen Verfahren vorgenommen wurde, und zum anderen, weil die komplexe stufenweise Aufbereitung der Datenbasis für die Art der nachfolgenden Analysen mitbestimmend ist.

Die anschließenden zwei zentralen Kapitel befassen sich mit den beiden Phänomenbereichen, in deren Kontext prosodische Grenzen hauptsächlich untersucht werden: turninterne prosodische Grenzen in flüssiger Sprache und prosodische Grenzen im Zusammenhang mit Sprecherwechseln. In jedem der Kapitel wird zuerst in einem kurzen Literaturüberblick der Stand der Forschung zusammengefasst. Anschließend werden die zu prüfenden Hypothesen formuliert. Der empirische Teil jedes Kapitels gliedert sich nach den verwendeten Untersuchungsmethoden wiederum in drei Teile. Als erstes wird über eine automatische Datenbankabfrage der Datenbasis die Extraktion von Merkmalsbündeln an prosodischen Grenzen vorgenommen (deskriptiver statistischer Teil). Dann werden zu

jedem thematischen Komplex Perzeptionsexperimente durchgeführt und ausgewertet, die die perzeptorische Relevanz phonetischer Merkmalsbündel und deren sprachliche Funktion im jeweiligen Phänomenbereich untersuchen (experimenteller Teil). Die phonetischen Parameter, die in den Experimenten überprüft werden, orientieren sich an der im Korpus beobachteten Variation. Darauf folgt jeweils ein Abschnitt, in dem Korpusbeispiele verschiedener phonetischer Ausprägung funktional im Dialogkontext interpretiert werden (interpretativer Teil). Zuletzt wird für beide Phänomenbereiche eine subklassifizierende Systematik für die Form-Funktions-Beziehung prosodischer Grenzen im Dialog vorgeschlagen.

Die prosodischen Grenzen in unflüssiger Sprache sind Thema des darauffolgenden Kapitels. Hier werden über Korpusabfragen und Perzeptionsexperimente die grundlegenden produktions- und perzeptionsbezogenen Eigenschaften unflüssiger prosodischer Grenzen dargestellt.

Folgende Annahmen werden im Laufe der Arbeit nachgegangen, wobei die einzelnen grundsätzlichen Überlegungen später in feinere experimentell oder korpusbasiert überprüfbare Hypothesen untergliedert werden.

- Wahrnehmbare prosodische Grenzen manifestieren sich durch Bündel phonetischer Merkmale. Prosodische Grenzen, die nur durch ein einzelnes phonetisches Merkmal hervorgerufen werden, kommen in natürlicher Sprache nur selten vor.
- Es können drei funktional verschiedene Bereiche differenziert werden: flüssige turninterne Grenzen, turnfinale Grenzen und Grenzen in unflüssiger Sprache. Die phonetische Struktur der prosodischen Grenzen der drei Bereiche ist unterschiedlich.
- Die unterschiedlich funktionalisierten Merkmalsbündel prosodischer Grenzen werden von Hörern dekodiert und unterstützen zum einen das inhaltliche Verständnis des Gehörten und zum anderen die interaktive Strukturierung des Gesprächs.
- Innerhalb jedes der drei funktionalen Bereiche können anhand phonetischer Parameter funktionale Subklassen gebildet werden. Im Bereich der flüssigen turninternen Grenzen spiegelt ein Kontinuum von phonetischer Kohäsion zu phonetischer Separierung ein inhaltliches Kontinuum von Kohäsion zu Separierung wider. Im Bereich der turnfinalen prosodischen Grenzen kann anhand der phonetischen und akustischen Struktur eine Differenzierung zwischen drei unterschiedlichen Typen von Sprecherwechseln belegt

werden: gezielte Turnübergaben, akzeptierte Turnübernahmen und umstrittene Turnübernahmen. Bei den prosodischen Grenzen in unflüssiger Sprache signalisieren unterschiedlich ausgeprägte Merkmalskombinationen Sprechflussstörungen verschiedener Art und Stärke.

- In jedem Bereich gibt es klar funktionalisierte prototypische phonetische Strukturen, die die Kernbereiche der Zuordnung zwischen phonetischer Form und sprachlicher Funktion bilden. Durch sprechertypische und zufällige Variation entsteht aber eine große phonetische Vielfalt prosodischer Grenzen, die sich in den Randbereichen überlagert und die keine eindeutige Form-Funktions-Zuordnung zulässt.
- Neben den phonetischen Merkmalen prosodischer Grenzen, die auf der Sprecher- und Hörerseite für die Strukturierung von Gesprächen ausgenutzt werden, spielen auch syntaktische Merkmale, sowie die semantische und pragmatische Gesamtstruktur von Äußerungen im Gesprächskontext eine wichtige Rolle.

Auf der beiliegenden CD-ROM findet sich eine Sammlung von aufbereiteten Sprachsignalen aus den untersuchten Dialogen. Die ausgewählten Signalabschnitte dienen als Beispiele für die verschiedenen untersuchten Phänomene und repräsentieren meist typische phonetische Realisierungen bestimmter funktionaler Strukturen, zeigen aber auch problematische Fälle, die nicht durch exakt abgrenzbare Systematisierungen erfasst werden können. So soll es dem Leser ermöglicht werden, einen direkten Einblick in die phonetische Struktur der untersuchten Daten zu gewinnen und hierdurch die in dieser Arbeit herausgearbeiteten Klassifikationen nachzuvollziehen. Auf der CD befinden sich außerdem Sammlungen von Stimuli aus den durchgeführten Perzeptionsexperimenten, so dass ein möglichst vollständiger Einblick in die experimentellen Arbeiten gewährleistet wird. Im Text der vorliegenden Arbeit sind Verweise zu den jeweils relevanten Beispielen enthalten. Diese Verweise sind durch das Symbol  markiert.

Die Sprachsignale und Abbildungen auf der CD sind in ein HTML-Dokument eingebunden, dass in einem Browser durch das Laden der Datei `Anfang.htm` geöffnet werden kann. Um die Audiobeispiele über den Browser abzuhören, muss dieser für das Abspielen von WAV-Dateien konfiguriert sein.

## 2 Methoden

Dieser Abschnitt führt kurz in die grundlegenden methodischen Anforderungen der Arbeit und die gewählten methodischen Verfahren ein, die zur Anwendung kommen. Detaillierte Beschreibungen finden sich in den Methodenteilen der nachfolgenden Kapitel. Der Beschreibung der mit der Datenbasis verbundenen Problemstellungen und deren Lösungsvorschlägen wird in diesem Abschnitt allerdings relativ viel Raum gewidmet, da sowohl im Bereich der Datenerhebung als auch der annotierenden Datenaufbereitung neue Verfahren entwickelt und erstmals angewendet wurden.

Die Zielsetzung der Arbeit führt zu drei Arbeitsbereichen, für die methodische Verfahren entwickelt werden:

- (1) Phonetische Korrelate prosodischer Grenzen werden über Häufigkeitsverteilungen beschrieben. Einige der untersuchten Merkmale können direkt aus der prosodischen Etikettierung übernommen werden, andere werden durch Messverfahren aus der segmentellen Etikettierung abgeleitet. Die Häufigkeitsverteilungen der Merkmale können herangezogen werden, um drei Grenztypen auf phonetischer Basis voneinander zu differenzieren: turninterne flüssige Grenzen, turninterne unflüssige Grenzen und turnfinale Grenzen. Weiterhin werden innerhalb dieser Grenztypen Subklassen gebildet. Die Analysen basieren auf Produktionsdaten aus spontansprachlichen Gesprächen und beziehen verschiedene phonetisch annotierte Korpora ein (siehe 2.1).
- (2) Beziehungen zwischen der phonetischen Realisierung prosodischer Grenzen und sprachlichen Funktionen werden anhand einer interpretativen Korpusanalyse untersucht.
- (3) Perzeptorische Aspekte phonetischer Separierung und deren semantisch-pragmatische und interaktionale Funktion werden experimentell untersucht.

Diese Ziele sind sehr unterschiedlich und nur über die Kombination verschiedener Methoden erreichbar. Für den ersten Bereich wird das unter Abschnitt 2.1 beschriebene Korpus verschiedener spontansprachlicher Dialoge herangezogen. Die prosodische Etikettierung der untersuchten Dialoge umfasst etwa 9000 Kennzeichnungen prosodischer Grenzen. Für die statistische Beschreibung der phonetischen Merkmale prosodischer Grenzen wird anhand der Etikettierung an jeder etikettierten Grenze durch automatische Datenbankabfragen ein Bündel der vorhandenen phonetischen Grenzmerkmale abgeleitet. Die phonetischen Parameter prosodischer Grenzen, die aus der Etikettierung gewonnen werden können, haben unterschiedlichen Status: einige Merkmale (z.B. melodische Konturen oder F0-Reset) stammen aus der prosodischen Etikettierungsebene und gehen auf perzeptiv orientierte Beurteilungen durch trainierte Etikettierer zurück. Auf andere Parameter (z.B. Pausen und Atmungsphasen) kann direkt über die segmentelle Etikettierung zugegriffen werden. Die Analyse der finalen Längung vor prosodischen Grenzen hingegen beruht auf einem automatisierten mathematischen Verfahren, das anhand der segmentellen Etikettierung für jede prosodische Grenze den Grad der finalen Längung ableitet.

Die Korpusuntersuchung liefert zunächst einen Überblick über die Art und Häufigkeit einzelner phonetischer Merkmale sowie deren Bündelungen an prosodischen Grenzen. Es können aber anhand der automatisch abfragbaren Information auch schon strukturelle und funktionale Subklassifizierungen vorgenommen werden, da die Etikettierung Informationen über die Position einer prosodischen Grenze im Dialog (turnintern vs. turnfinal) sowie über das Vorkommen von Produktionsstörungen (syntaktische Abbrüche, Zögerungslängung und Häsitationspartikeln sind gekennzeichnet) enthält.

Die Zuordnung bestimmter Bündelungen zur pragmatischen, semantischen und interaktionalen Struktur der Dialoge erfordert ein Hinausgehen über automatische Korpusabfragen, da auf diesem Weg nur wenig Information über die Äußerungs- und Dialogstruktur gewonnen werden kann. Es ist vielmehr notwendig, eine große Zahl von Beispielen aus dem Korpus herauszugreifen und in Bezug auf sprachliche Funktionen im Dialogkontext zu interpretieren. So können verschiedene Typen inhaltlicher und struktureller Grenzen in der Interpretation erfasst und phonetischen Ausprägungen prosodischer Grenzen zugeordnet werden.

Die Kombination von Korpusabfrage und Interpretation liefert noch keine Erkenntnisse über die Einordnung verschiedener Merkmalskombinationen auf perzeptorischen Skalen zwischen Kohäsion und Separierung. Deshalb werden für bestimmte Merkmalskombinationen, die sich im Korpus zeigen, Perzeptionsexperimente zur Wahrnehmung und Funktion dieser phonetischen Muster durchgeführt. Diese Experimente liefern Erkenntnisse über die Perzeption von Pausen verschiedener Länge und ihre perzeptorische Interaktion mit anderen Merkmalen oder beziehen sich auf den perzeptorischen Zusammenhang zwischen melodischen Mustern und segmenteller Längung an prosodischen Grenzen und bei Sprecherwechsel im Dialog.

Die Entwicklung der Perzeptionsexperimente steht in enger Verbindung mit den Erkenntnissen aus der Korpusanalyse (automatische Abfragen und Interpretation), so dass sich die Vorgehensweisen und die Ergebnisse der drei methodischen Stränge gegenseitig ergänzen. Die Forschungsergebnisse der methodischen Stränge sollen am Ende ein Gesamtbild des Phänomenbereichs ergeben.

Da die Experimente auf die Aufdeckung von Form-Funktions-Zusammenhängen ausgerichtet sind, ist es nicht sinnvoll, die Ergebnisse lediglich über die gefundenen Signifikanzen (z.B. auf dem 5%-Niveau) zu beschreiben, denn das gewählte Messverfahren ist sehr sensitiv, und schon bei kleinen Differenzen zwischen Mittelwerten ergeben sich signifikante Unterschiede. Es ist jedoch zu beachten, dass ein kleiner Unterschied in der Beurteilung phonetisch unterschiedlicher Stimuli nicht unbedingt mit funktional unterschiedlichen Klassen einhergeht. Häufig sind über die Stimulusserien kontinuierlich steigende oder fallende Urteile zu beobachten, die sich von Stimulus zu Stimulus signifikant unterscheiden. Dennoch ist bei kleinen messbaren Unterschieden nicht automatisch eine funktionale Relevanz gegeben. Hieraus ergibt sich folgendes Vorgehen.

In der Auswertung der intervallskalierten Daten aus den Perzeptionsexperimenten kommt eine Kombination deskriptiver und prüfstatistischer Verfahren zum Einsatz, die eine übersichtliche Präsentation der experimentellen Ergebnisse ermöglicht und Interpretierbarkeit gewährleistet. In deskriptiven Statistiken (meist Darstellungen von Mittelwerten) können klar und übersichtlich die Auswirkungen der Parametermanipulationen auf die Urteile der Versuchspersonen herausgearbeitet werden. Diese Art der Darstellung ist besonders geeignet, verschiedene Effektstärken aufzuzeigen und starke Effekte von schwachen Effekten zu trennen. Ergänzend wird für jedes Experiment eine

mehrfaktorielle Varianzanalyse (ANOVA) mit Messwiederholung berechnet, um zu einer prüfstatistischen Absicherung der Beobachtungen zu gelangen. In das prüfstatistische Verfahren gehen aufgrund der Struktur der Experimente lediglich Messwiederholungsfaktoren ein. Notwendige Freiheitsgradkorrekturen werden nach Greenhouse-Geisser (1959) vorgenommen. Bei einem Niveau unter 5% Fehlerwahrscheinlichkeit wird der Terminus 'signifikant' verwendet, bei einem Niveau unter 1% der Terminus 'hochsignifikant'.

## 2.1 Datenbasis

Diese Untersuchung ist eine Fortführung der Arbeiten zur prosodischen Gestaltung deutscher Spontansprache im Rahmen der DFG-Projekte *Lautmuster deutscher Spontansprache* und *Prosodische Phrasierung in deutscher Spontansprache*, die am *Institut für Phonetik und digitale Sprachverarbeitung* der Universität Kiel unter der Leitung von Prof. K.J. Kohler bearbeitet wurden (Ko331/22-1,2 und Ko331/23-1). Die Datenbasis der DFG-Untersuchung *Lautmuster deutscher Spontansprache* umfasst vor allem Dialoge, die im Rahmen von VERBMOBIL (URL(1)) in einem Terminabspracheszenario erhoben wurden (Kohler et al. 1995). Die Daten sind als *The Kiel Corpus of Spontaneous Speech Vol. I* und *Vol. II* veröffentlicht (IPDS 1995, 1996).

Die im Projekt *Prosodische Phrasierung in deutscher Spontansprache* vorgenommene Erweiterung der Fragestellung auf die Ausprägung und die Funktionen prosodischer Grenzen im Bereich der Dialogsteuerung und des unflüssigen Sprechens setzt eine Datengrundlage voraus, die ein höheres Maß an Natürlichkeit bietet. Dies betrifft vor allem die Dialogsteuerung, die in den Daten aus dem Terminabspracheszenario vollkommen unnatürlich ist, da die Sprechenden nur wechselseitig sprechen können, solange ein Knopf gedrückt wird, der den Aufnahmekanal freischaltet. So entstand die Notwendigkeit zur Aufnahme neuer Daten. Im folgenden werden einige grundsätzliche Überlegungen zur Sprachdatenerhebung im Rahmen phonetischer Untersuchungen angestellt, die in der Konzeption und Umsetzung eines neuen Aufnahmeszenarios münden, das Sprachdaten liefert, die den Anforderungen dieser Arbeit gerecht werden.

In der phonetischen Forschung spielen spontansprachliche Korpora eine zunehmend wichtige Rolle, denn Sprachkorpora bieten die Möglichkeit, phonetische Hypothesen verschiedenster Art durch systematische Datenbankabfragen zu überprüfen. Bei der Erhebung von Sprachdaten gibt es grundsätzlich zwei Herangehensweisen. Sie können zielgerichtet für eine bestimmte Problemstellung erhoben werden oder im Rahmen der Grundlagenforschung für die Bearbeitung verschiedener Fragestellungen ausgelegt sein.

In allen unter Laborbedingungen erhobenen "spontansprachlichen" Daten tritt das methodische Problem des 'Sei-spontan-Paradoxons' auf (Watzlawick 1983). Dies besagt, dass eine Aufforderung zu spontanem Agieren prinzipiell paradox ist, da Spontaneität eben gerade ein Verhalten ist, das ohne gezielte Aufforderung im Rahmen einer Situation entsteht. Die Sprachdaten, die in der gezielten Elizitation erhoben werden, sind also nicht im eigentlichen Sinne spontan und sollten wohl eher als 'Quasi-Spontansprache' bezeichnet werden. Ich werde trotzdem im weiteren den Begriff 'Spontansprache' als Sammelbegriff für Sprache verwenden, die nicht einer orthographischen Vorlage entstammt. Denn obwohl sich solche elizitierten Gespräche aus einer Laborsituation ergeben, entstehen trotzdem oft Unterhaltungen die durch ungezwungenes Sprechen bestimmt sind und keiner direkten Vorgabe entstammen (wie z.B. beim Ablesen von Texten).

Natürlich gibt es auch die Möglichkeit, Gespräche mit einem versteckten Mikrofon aufzunehmen und nachträglich von den Sprechern das Einverständnis zur Verwendung der Daten einzuholen (Simpson 1991). Unter der Bedingung der heimlichen Aufnahme ist aber kaum eine Aufnahmequalität zu erzielen, die den Anforderungen einer instrumentellen Analyse gerecht wird; da solche Mitschnitte nicht in schallbehandelten Räumen gemacht werden können, gibt es viele Nebengeräusche und durch Bewegungen der Sprecher wird die Position zum Mikrofon ständig variiert, was zu Schwankungen in der Intensität des Signals führt. Auch eine Kanaltrennung zwischen den Sprechern ist generell ausgeschlossen, da sich die Sprecher im Rahmen eines natürlichen Gesprächs normalerweise im selben Raum befinden.

Ein weiteres Dilemma der Datenerhebung, nämlich das Wechselspiel zwischen Kontrollierbarkeit und Natürlichkeit der Daten, formuliert Lehiste (1963:353):

*It appears that the linguist who wants to use experimental methods in his investigation is forced to trade a certain amount of natural*

*ness in his experimental design.[...]The two requirements of naturalness and rigorousness appear to occupy the opposite ends of a continuous scale of gradual mutual exclusion. At the one extreme we find complete naturalness coupled with complete lack of control; at the other end, complete control over the experimental design at the expense of naturalness and linguistic relevance.*

Dieser gegenseitige Ausschluss von Natürlichkeit und Kontrolle ist insofern ein Problem, als sich eine phonetische Korpusuntersuchung auf beobachtbare Phänomene in den Daten bezieht, die in ausreichender Häufigkeit auftreten müssen, z.B. auf die Intonation in syntaktisch unmarkierten Fragekonstruktionen. Es würde in diesem Fall also die Anforderung an das Korpus gestellt werden, dass eine Mindestanzahl syntaktisch unmarkierter Fragen in den Daten auftritt und die beobachtbaren Fälle miteinander vergleichbar sind. Wenn aber ein Korpus untersucht wird, das ohne jede Kontrolle über die sprachlichen Äußerungen aufgenommen wurde, ist die Wahrscheinlichkeit gering, dass eine ausreichende Anzahl von *tokens* für eine systematische Analyse des Phänomens in den Daten vorkommt. Je komplexer die untersuchte Konstruktion ist, desto unwahrscheinlicher wird deren Auftreten. Wenn aber ein sehr rigides Elizitationsverfahren (im Extremfall das Ablesen von Sätzen) zum Einsatz kommt, muss eingewendet werden, dass die erhobenen Daten nicht aus einem natürlichen Gesprächskontext stammen und somit höchstens einzelne Aspekte der sprachlichen Wirklichkeit repräsentieren können. Insofern ist bei jeder Datenerhebung ein Kompromiss zwischen Kontrolle und Natürlichkeit einzugehen.

Die meisten qualitativ hochwertigen Daten, die bisher auf dem Gebiet der Spontansprache erhoben und ausgewertet wurden, sind im Rahmen von *Map Task* (Brown et al. 1983; Swerts & Collier 1992) oder im Terminabspracheszenario (Pätzold & Simpson 1994) entstanden. Bei *Map Task* beschreibt ein Sprecher (instruction-giver) dem Dialogpartner (instruction-follower) eine Route, die auf einer stilisierten Landkarte eingezeichnet ist. Der Dialogpartner soll die beschriebene Route dann auf einer anderen Karte einzeichnen. Hierbei kommt es zu Schwierigkeiten, die von den Sprechern gelöst werden müssen, da die Landkarten der Dialogpartner sehr ähnlich, aber nicht identisch sind. Im Terminabspracheszenario geht es darum, anhand vorgegebener Terminkalender eine Reihe von

Terminen zu verabreden. Auch hierbei stoßen die Dialogpartner auf zu lösende Probleme, da in den Kalenderblättern bereits einige Termine eingetragen sind, die mit denen des Partners kollidieren. Beide Szenarien haben den Vorteil, dass das Auftreten bestimmter Wörter relativ gut gesteuert werden kann. Im Terminabspracheszenario treten Wochentage, Monatsnamen und Zahlen auf, in *Map Task* einige Gegenstände und Ortsbezeichnungen, die auf den Landkarten verzeichnet sind. Die Gesamtanzahl verschiedener Wörter, die elizitiert werden, ist dabei vergleichsweise gering. Einige sonst häufige lexikalische Einheiten (z.B. die Personalpronomina *sie* und *er*) kommen kaum vor, da die Sprecher fast nie über weitere Personen reden. Weiterhin erzeugen die Szenarien in der Regel nur wenig emotionale Beteiligung bei den Versuchspersonen und führen z.T. zu einer gekünstelten oder gelangweilten Sprechweise. Durch die Aufgabenstellung des *Map Task*-Szenario entsteht weiterhin eine klare Rollenverteilung (*instruction-giver* und *instruction-follower*), die u.a. die Auftretenshäufigkeiten prosodischer Muster für die verschiedenen Rollen beeinflusst (Fletcher et al. 2002). Das soll nicht bedeuten, dass diese Arten der Elizitation von Spontansprache ungeeignet wären, nur müssen die Charakteristika der elizitierten Sprache, die aus dem Versuchsaufbau resultieren, Eingang in die Bewertung der Ergebnisse finden und dürfen nicht generalisiert werden. In diesem Zusammenhang soll hier ein neues Dialog-Szenario vorgestellt werden, das Sprachdaten liefert, die sich in einigen Punkten erheblich von bisher erhobenen Daten unterscheiden: das *Video Task* Szenario (Peters 2001). Folgende grundsätzliche Fragen sind bei der Entwicklung des Datenerhebungsverfahrens berücksichtigt worden:

- Sollen Monologe, Dialoge oder Gespräche in Gruppen elizitiert werden?
- Wie werden die Versuchspersonen veranlasst, frei und ungezwungen zu sprechen?
- Welche Stilebene (von förmlich bis intim) wird das verwendete Szenario hervorrufen?
- Welche Rollenverteilung wird durch das Szenario bedingt?
- In welcher emotionalen Situation befinden sich die Versuchspersonen?
- Welche lexikalischen Items sind zu erwarten?
- Welche prosodische Gestaltung ist zu erwarten?
- Wie lang sollen die Dialoge sein?
- Wie kann eine angemessene Qualität der Aufnahme erzielt werden?
- Welche Fragestellungen sollen und können an den erhobenen Daten bearbeitet werden?

Jedes Szenario wird in Bezug auf die oben genannten Punkte spezielle Schwerpunkte setzen und in der Lage sein, einige der Parameter gut zu kontrollieren. Eine umfassende Kontrolle aller Parameter ist durch kein Szenario zu gewährleisten. Bei der Erhebung von Spontansprache wird es immer einen Kompromiss zwischen dem Wunsch nach umfangreicher Kontrolle, wie man sie aus gelesener Laborsprache kennt, und der Natürlichkeit von unbeobachteter Spontansprache geben.

### **2.1.1 Das *Video Task* Szenario**

Die Grundidee von *Video Task* ist, zwei Versuchspersonen ähnliches, aber nicht identisches Filmmaterial zu präsentieren und sie im Anschluss an die Präsentation die Unterschiede und Gemeinsamkeiten des Gesehenen in einem Gespräch erörtern zu lassen. Hierbei ist es möglich, verschiedenste Filmmaterialien so aufzuarbeiten, dass sie für das Szenario verwendet werden können.

In den bislang durchgeführten Datenerhebungen dienten als Ausgangsmaterial zwei manipulierte Versionen der bekannten deutschen Fernsehserie *LINDENSTRASSE* (URL (2)), die aus einer Reihe bereits ausgestrahlter Folgen zusammengeschnitten wurden. Das Szenario wird deshalb auch als *Daily Soap* Szenario bezeichnet, die erhobenen Daten als *LINDENSTRASSE*-Korpus. Die Videosequenzen sind etwa 15 Minuten lang und weichen in folgenden Punkten teilweise voneinander ab:

- Auswahl und Abfolge der Szenen
- Vollständigkeit von Szenen
- Auftreten von Bild- und Tonstörungen

Im Vorfeld der Datenerhebung werden Versuchspersonen gesucht, die mit der Handlung und den Charakteren der *LINDENSTRASSE* zumindest einigermaßen vertraut sind. Das ist relativ einfach, denn die Serie läuft seit mehr als zehn Jahren im deutschen Fernsehen und erfreut sich großer Beliebtheit, wobei viele der Stammzuschauer die Serie mit Humor und einer gewissen Ironie verfolgen. Es wird darauf geachtet, dass die Dialogpartner miteinander befreundet sind. Vor der Durchführung der Aufnahme wird den Dialogpartnern in getrennten Räumen je eine der beiden 15-minütigen Fassungen zweimal

hintereinander gezeigt. Danach werden sie über Kopfhörer und Mikrofone miteinander verbunden und bekommen die Aufgabe, sich über die Unterschiede zwischen den 'Folgen' der Serie zu unterhalten. Es besteht kein Sichtkontakt. Auf die Benutzung von schriftlichen Notizen wird verzichtet. Die Vorteile, die dieses Szenario in Bezug auf die oben genannten Bedingungen bietet, sind folgende:

- Das Szenario eignet sich prinzipiell sowohl zur Aufnahme von Dialogen, als auch um Gespräche in einer größeren Gruppe zu elizitieren. Die vorliegende Arbeit konzentriert sich aber auf den Dialog zwischen zwei Sprechern. Dies hat zwei Gründe. Erstens ist diese Gesprächsform besonders im privaten Bereich die häufigste Sprechsituation, und zweitens ist die sprachliche Interaktion in größeren Gesprächsgruppen noch komplexer und schwieriger zu erfassen als im Gespräch zwischen zwei Personen.
- Die Aufgabe, die gestellt wird, ist interessant. Das erleichtert es den Versuchspersonen, sich vollständig auf die inhaltliche Fragestellung zu konzentrieren und dadurch weniger an die Laborsituation zu denken. Der Verzicht auf schriftliche Notizen entspricht eher einer Alltagssituation als eine Situation, in der alles schriftlich festgehalten werden muss. Auch dies steigert die Natürlichkeit der Kommunikation. In Interviews, die im Anschluss an die Dialoge mit den Versuchspersonen geführt wurden, hat sich gezeigt, dass diese in der Regel davon ausgingen, dass es bei dem Experiment tatsächlich um eine inhaltliche Fragestellung bezüglich der LINDENSTRASSE geht. Dass das eigentliche Ziel des Experiments die Aufzeichnung von Sprachsignalen ist, wurde von den Versuchspersonen in der Regel nicht erkannt.
- Die Vertrautheit der Dialogpartner mit den Charakteren der Serie sorgt für eine persönliche Atmosphäre des Gesprächs. Sich über die Personen aus einer lange verfolgten und relativ realitätsnahen Fernsehserie zu unterhalten, kommt einem Gespräch über reale Freunde oder Bekannte ziemlich nahe, und gerade dieser Gesprächstyp ist in der alltäglichen Kommunikation sehr verbreitet. Dieser Aspekt wird dadurch verstärkt, dass die Dialogpartner miteinander befreundet sind und es ihnen somit vertraut ist, sich ungezwungen zu unterhalten. Hinzu kommt, dass die Aufgabe den Versuchspersonen sehr viel Freiraum in der Gestaltung des Dialogs lässt. Die bisher erhobenen Dialoge sind entspannt, freundschaftlich, humorvoll und wirken sehr natürlich.

- Die Aufgabenstellung von *Video Task* erzeugt keine eindeutige Rollenverteilung. Die Aufgabe erfordert, dass beide Dialogpartner sowohl Fragen stellen als auch selber über das Gesehene berichten. In den bisher im *Video Task* Szenario aufgenommenen Dialogen sind die Gesprächsanteile beider Sprecher etwa gleich.
- Durch die Auswahl des Videomaterials kann das Auftreten emotional gefärbter Sprache hervorgerufen werden. Das hier benutzte Material führt eher zu einer belustigten Grundeinstellung. Durch eine Auswahl von Szenen mit bedrohlichen oder abstoßenden Inhalten kann voraussichtlich auch eine völlig andere emotionale Basis geschaffen werden, die sich dann auch im Gespräch manifestiert. Eine andere Möglichkeit, die Grundstimmung zu verändern, wäre, den Dialogpartnern ein zeitliches Limit für die Lösung der Aufgabe zu setzen, um deren sprachliches Verhalten unter Zeitdruck zu beobachten. Auch kann bei der Auswahl der Dialogpartner der Grad der Vertrautheit zwischen den Versuchspersonen variiert werden, um z.B. eine förmlichere Atmosphäre zu schaffen.
- Das Szenario gewährleistet das Auftreten etlicher lexikalischer Einheiten, die in *Map Task* oder den Terminabsprachen nicht vorkommen. Bedingt durch das freie Sprechen und die offene Aufgabenstellung verwenden die Versuchspersonen sehr viele unterschiedliche Wörter. Eine gewisse Kontrolle über das Auftreten lexikalischer Einheiten ist dadurch gegeben, dass bestimmte Wörter und Namen, die im Filmmaterial vorkommen, von den Versuchspersonen aufgegriffen werden.
- Die humorvolle Einstellung der meisten *LINDENSTRASSEN*-Zuschauer zur Serie und die komplexe Aufgabenstellung führen zu einer lebendigen und vielseitigen prosodischen Gestaltung. Dass die Aufgabe nur in einem wechselseitigen Abgleich zwischen den gesehenen Filmsequenzen zu lösen ist, führt zu häufigen Sprecherwechseln und zum Einsatz verschiedenster Mechanismen der Gesprächssteuerung.
- Durch den Umfang des präsentierten Videomaterials und die Anzahl der auftretenden Abweichungen kann die Dauer der Dialoge gesteuert werden. Die schallgedämpfte Aufnahmeumgebung gewährleistet eine gute technische Qualität. Die räumliche Trennung der Versuchspersonen ermöglicht eine Kanaltrennung der Sprachsignale.
- Die Sprachdaten aus dem *Video Task* Szenario sind nicht auf eine bestimmte nachfolgende Untersuchung zugeschnitten. Das Ziel der Datenerhebung war ein möglichst alltäglicher, lockerer Dialog zwischen befreundeten Personen. Insofern können die

Daten zur Basis verschiedenster Untersuchungen gemacht werden, die sich auf diesen Typ von Sprache beziehen. Die Kanaltrennung erlaubt hierbei das getrennte Abhören der Kanäle und den Einsatz instrumenteller akustischer Analysen auch in überlappenden Signalbereichen.

Zusammenfassend können die Sprachdaten aus dem *Video Task* Szenario als phonetisch reiche Daten charakterisiert werden. Das Szenario bindet das Interesse der Versuchspersonen, lenkt von der Laborsituation ab und fördert hierdurch ungezwungenes sprachliches Verhalten. Dies führt zu komplexen Gesprächssteuerungsmechanismen zwischen den Dialogpartnern, einer großen Vielfalt lexikalischer Items und einer vielseitigen prosodischen Gestaltung.

### 2.1.2 Umfang der Sprachdaten

Es wurden 6 Dialoge zwischen vier weiblichen und zwei männlichen Paaren mit einer Dauer von je 10 bis 15 Minuten aufgenommen. Die Dialoge enthalten zusammen 13.500 fortlaufende Wörter und haben eine Gesamtdauer von 80 Minuten. Alle Sprecherinnen und Sprecher stammen aus Schleswig-Holstein und sprechen Standarddeutsch norddeutscher Prägung.

**Tabelle 1:** Übersicht über die sechs *Video Task* Dialoge (LINDENSTRASSE)

Dialogkennung	Sprecherkennungen	Wörter	Dauer
101a	TEV (w, 23, Studentin) CJE (w, 28, Studentin)	2682	9:37
102a	AHA (w, 22, Schülerin) TLU (w, 20, Schülerin)	1955	13:02
103a	APE (w, 27, Studentin) KPA (w, 32, Designerin)	1884	10:03
104a	KTH (w, 23, Studentin) CHE (w, 23, Studentin)	1983	9:23
105a	UHO (m, 33, Techniker) JKO (m, 31, Kaufmann)	2911	15: 27
106a	MPI (m, 35, Schiffbauer) TRA (m, 31, Techniker)	2209	11:26

### 2.1.3 Aufbereitung der Sprachdaten

Die Aufbereitung sprachlicher Daten für nachfolgende Untersuchungen kann unterschiedlicher Art sein. Im einfachsten Fall wird nur eine orthographische Verschriftung vorgenommen. Die Aufbereitung der *Video Task*-Daten umfasst verschiedene Ebenen und ist sehr aufwendig. Im Kern hat das Aufbereitungsverfahren das Ziel, die Verbindungen zwischen Text, Signaleigenschaften und metasprachlichen Kategorien (z.B. Phonemen oder phonologischen Melodiemustern) herzustellen, um später die funktionalen und akustischen Korrelate dieser Kategorien analysieren zu können.

Die erhobenen Sprachdaten wurden orthographienah verschriftet, manuell segmentiert und prosodisch etikettiert. Zuletzt wurde manuell eine Kodierung der unterschiedlichen Turnübergänge eingefügt. Diese Schritte, die im folgenden kurz beschrieben sind, wurden mit Hilfe der am IPdS entwickelten Rechnerumgebung *xassp* (IPDS 1997) durchgeführt. Die Konventionen für die Verschriftung und die Segmentierung sind im Rahmen des VERBMOBIL-Projekts erarbeitet worden (URL (1); Kohler et al. 1995; Kohler et al. 1997).

Alle Etikettierungen wurden durch trainierte Bearbeiter vorgenommen (in der Regel studentische Hilfskräfte aus dem Studiengang Phonetik und digitale Sprachverarbeitung). Die Etikettierungen wurden bei Unsicherheiten durch mehrere Bearbeiter beurteilt und zuletzt einer automatischen Labelsyntaxüberprüfung unterzogen. Die prosodische Etikettierung der Dialoge wurde außerdem von einem weiteren trainierten Bearbeiter komplett überprüft. Dieses Vorgehen ist grundsätzlich verschieden von automatischen Segmentierungen und prosodischen Etikettierungen, die auf rein algorithmischen, akustisch determinierten Funktionen beruhen und schnell große Mengen von Sprachdaten aufbereiten können, aber auch sehr fehleranfällig sind und oft nicht die perzeptiv relevanten Kategorien, sondern lediglich akustische Unterschiede zwischen Abschnitten des Signals erfassen.

### 2.1.3.1 Transliteration

Die orthographienahe Verschriftung (*Transliteration*) wird im 7-Bit ASCII-Format kodiert. Folgende Objekte, die für die vorliegende Arbeit von Bedeutung sind, werden symbolisiert (Kohler et al. 1995):

- lexikalische Einträge
- nichtverbale Elemente (z.B. Häsitationspartikeln, Lachen, Husten usw.)
- Pausen und Atmen
- starke segmentelle Längung
- Interpunktion nach Duden
- syntaktische Abbrüche

Beispiel für die Transkription (Referenz: l021a010)

```
<Schmatzen> <A> <{hm> <P> als/- <P> doch , ja , die  
<:<Lachen> kenn' ich:> <Lachen> . <A> ach so . <A> ja , ich  
hab' nur gesehen , wie er <{hm> grade von seiner  
Tauchlehrerin abgeholt wurde <P> mit seiner Tauchbrille  
<:<Lachen> auf:> <Lachen> . <P> <Schlucken> mhm . <P>
```

### 2.1.3.2 Automatische kanonische Transkription

Die Transliteration dient als Input für den Graphem-Phonem-Konverter des *RULSYS* Text-to-Speech-Systems für das Deutsche (Kohler 1992, 1997). Der Konverter generiert eine *default*-Phonemfolge für die lexikalischen Einträge der Transliteration. Die phonetisch-phonematische Transkription wird im Transkriptionsalphabet *modified SAMPA* (Barry & Fourcin 1992; Wells et al. 1989; Kohler et al. 1995) vorgenommen, das zur ASCII-Kodierung der Phonemsysteme von sechs Sprachen aus der Europäischen Union erstellt wurde. Diese kanonische Transkription dient als Basis für die manuelle segmentelle Etikettierung der Sprachsignale.

Außer den lexikalischen Einträgen werden auch alle anderen Elemente der Transliteration automatisch umgesetzt, auch lexikalische Haupt- und Nebenakzente und Kompositumsgrenzen werden automatisch markiert. Die Zugehörigkeit zur Klasse der Funktionswörter wird gekennzeichnet. Satzzeichen werden direkt aus der Transliteration übernommen. Das Endergebnis der Konvertierung ist eine Datei mit einer Liste der durch *RULSYS* generierten segmentellen und sonstigen Etiketten.

### **2.1.2.3 Segmentelle Etikettierung**

Ziel der segmentellen Etikettierung ist die Zuordnung der Elemente der durch *RULSYS* generierten Labeldatei zur zeitlichen Abfolge des Sprachsignals. Diese Verknüpfung zwischen den abstrakten phonetischen und linguistischen Kategorien der Etiketten mit dem Signal gewährleistet den späteren Zugriff auf die Signalbereiche, die diesen Kategorien zugeordnet sind. Der Bearbeiter hat für die Segmentierung das Oszillogramm sowie ein unter *xassp* berechnetes Sonagramm auf dem Bildschirm und das Sprachsignal über einen Kopfhörer zur Verfügung und ordnet die Etiketten bestimmten Zeitpunkten des Signals zu.

Speziell in der Spontansprache kommt es häufig zu starken Reduktionserscheinungen wie der Elision von Segmenten oder dem Verschmelzen mehrerer Segmente zu einem Abschnitt im Signal, der die artikulatorischen Parameter verschiedener Segmente enthält und nicht weiter segmentierbar ist. Häufige Beispiele hierfür sind Nasalierungen und Glottalisierungen in der Umgebung von Nasalen bzw. von Glottalverschlüssen. Für die Markierung solcher phonetischen Erscheinungen stehen spezielle Etiketten zur Verfügung.

### **2.1.3.4 Prosodische Etikettierung**

Nach Abschluss der segmentellen Etikettierung wird unter Verwendung des Etikettierungssystems PROLAB die auf *KIM* basierende prosodische Etikettierung vorgenommen (Kohler 1991a, 1995b; Kohler et al. 1995; Peters 1999; Peters & Kohler 2004).

Die Ziele von *KIM* und dessen Annotationssystem PROLAB sind die Erstellung einer prosodischen Phonologie bzw. die prosodische Annotation von Datenkorpora der

Lese- und Spontansprache. Beide - phonologische Modellierung und Annotation - operieren auf perzeptorischer Basis. D.h. für die Kategorienbildung sind die für den Hörer signifikant differenzierten prosodischen Muster entscheidend. Deshalb hat die kontrastive auditive Evaluation prosodischer Äußerungsphänomene durch den Bearbeiter sowie durch Versuchspersonen in formalen Perzeptionsexperimenten einen zentralen Stellenwert. F0-Analysen und linguistische Systemüberlegungen haben dabei nur Hilfsfunktionen.

Die perzeptorische Orientierung zeigt sich auch bei der prosodischen Etikettierung durch trainierte Bearbeiter. Die Etikettierung wird in erster Linie auf auditiver Basis durchgeführt, wobei die Bearbeiter die gehörten Melodiemuster in den Daten den in einer vorangehenden Trainingsphase bewusst gemachten *KIM*-Kategorien zuordnen. Die graphische Darstellung von F0-Verläufen wird nur unterstützend herangezogen und bestimmt nicht die Kategorisierung. Diese weitgehende Beschränkung auf den Höreindruck sorgt zum einen dafür, dass mikroprosodische Störeinflüsse und algorithmisch bedingte Signalanalysefehler kaum Eingang in die Klassifikation der melodischen Muster finden (anders als bei stark messorientierten Etikettierungsverfahren). Zum anderen können durch dieses Verfahren auch die vielfältigen "allotonischen" Varianten der Muster den jeweiligen Basiskategorien zugeordnet werden.

Unterstützend wird bei der Kategorisierung auch die Technik der Analyse-durch-Synthese eingesetzt. Durch die Manipulation und Resynthese der Melodieverläufe werden leicht abgewandelte Varianten des beobachteten Musters generiert. So kann durch kontrastive auditive Evaluierung in Zweifelsfällen besser entschieden werden, welcher Kategorie ein vorgefundener Verlauf zuzuordnen ist.

Es sollen hier nun kurz die Kategorien des *Kieler Intonationsmodells (KIM)* eingeführt werden, die für die vorliegende Arbeit von Bedeutung sind. Für umfassende Darstellungen siehe Kohler 1991a; Peters 1999; Peters & Kohler 2004.

Im Rahmen der Etikettierung werden die Positionen prosodischer Grenzen mit dem Label  $PG_n$  markiert. Das  $n$  im Label  $PG_n$  dient als Platzhalter für eine spätere phonetisch und funktional orientierte Subklassifikation in verschiedene Typen prosodischer Grenzen (z.B.  $PG_1$ ,  $PG_2$  etc., vgl. 3.2.13). Eine solche Subklassifikation kann anhand der phonetischen Merkmale, die in der prosodischen und segmentellen Etikettierung festgehalten sind, durch geeignete Programme automatisch vorgenommen werden. Dieser Arbeitsschritt ist aber bisher nicht durchgeführt worden.

Prosodische Grenzen sind durch Bündel phonetischer Merkmale definiert (z.B. Pausen und Atmen, segmentelle Längung, Brüche im tonalen Verlauf) die zur Wahrnehmung von Zäsuren beim Hörer führen. Die Stärke der Merkmale und ihre Gewichtung variiert von Fall zu Fall und führt zu unterschiedlichen Graden der Kohäsion bzw. Separierung an der jeweiligen Grenze. Verschieden ausgeprägte prosodische Grenzen wurden aber in der bisherigen manuellen Etikettierung nicht unterschiedlich symbolisiert, da hierfür kein Kategoriensystem existierte. Die einzige Ausnahme bildet die Unterscheidung zwischen prosodischen Grenzen mit und ohne nachfolgendem F0-Reset (Neueinsatz der Deklinationslinie).

Jede lexikalische Einheit wird hinsichtlich der Merkmale Satzakkentuierung und melodischer Verlauf klassifiziert. Hierbei wird jedem Wort eine Akzentstufe zwischen 0 und 3 zugewiesen. Akzentstufe 0 entspricht einer deakzentuierten Realisierung, 1 einer partiellen Deakzentuierung, 2 ist die *default*-Kategorie für primäre Satzakkente, und 3 bezeichnet emphatisch verstärkte Akzente. Wenn eine lexikalische Einheit akzentuiert ist, wird der melodische Verlauf an der Satzakkentposition einer von drei Kategorien zugeordnet:

- Gipfelkontur mit (steigend-)fallendem Konturverlauf
- Talkontur mit steigendem Konturverlauf
- ebene Kontur mit glattem Konturverlauf

Die Klassen der Gipfel- und Talkonturen werden je nach der Synchronisation des melodischen Verlaufs mit dem Nukleus der satzakkentuierten Silbe weiter subklassifiziert. Die Gipfelkonturen in früh, mittel und spät, die Talkonturen in früh und spät. Ein früher Gipfel hat sein F0-Maximum in der Regel vor der akzentuierten Silbe, der mittlere Gipfel innerhalb des Silbenkerns der akzentuierten Silbe und der späte Gipfel danach. In Perzeptionsexperimenten wurde nachgewiesen, dass diese drei Synchronisationsmuster verschiedene semantisch-pragmatische Bedeutungen kodieren (Kohler 1991b). Der frühe Gipfel signalisiert bekannte Information und die Abgeschlossenheit von Aussagen, mittlerer und später Gipfel signalisieren neue Information, wobei der späte Gipfel den Eindruck von Überraschtheit erzeugen kann.

Auch die frühen vs. späten Talkonturen werden in Abhängigkeit der Kontursynchronisation subklassifiziert und kodieren unterschiedliche Bedeutungen. Das frühe Tal hat seinen Tiefpunkt vor dem Silbenkern der akzentuierten Silbe und beginnt bereits auf der akzentuierten Silbe mit dem Anstieg. Das späte Tal hat den Tiefpunkt im Silbenkern und beginnt folglich später mit dem Anstieg. Frühe und späte Täler kommen häufig vor in Verbindung mit weiterweisenden Äußerungsteilen, oder um Fragen intonatorisch zu markieren. In Verbindung mit Fragen kodiert das frühe Tal eine neutrale/sachbezogene oder sogar distanzierte Einstellung des Fragenden. Das späte Tal hingegen erweckt den Eindruck der Involviertheit und kann auch höflicher klingen als das frühe Tal (Kohler 1995a, 2005b).

Ein weiteres Phänomen, das erfasst wird, sind auditiv prominente Silben ohne melodische Bewegungen, die die Prominenz verursachen. Die F0-Kontur verläuft in diesen Fällen eben und die auditive Hervorhebung der Silbe wird durch Dauer- und/oder Intensitätszunahme verursacht.

Prosodische Phrasen enthalten meist mehr als nur einen Satzakkzent, und es ist wichtig, die Verbindung (Konkatenation) zwischen den Gipfel-, Tal-, und ebenen Konturen zu erfassen, da auch durch unterschiedliche Konkatenationsmuster verschiedene inhaltliche Zusammenhänge repräsentiert werden können. Für Gipfelkonturen gibt es zwei häufige Konkatenationsmuster innerhalb der prosodischen Phrase: die Gipfelsequenz, in der zwei benachbarte Gipfelkonturen mit einer Grundfrequenzabsenkung zwischen den Maxima realisiert werden und dadurch als getrennte Gipfel nebeneinander stehen. Dem gegenüber steht das Hutmuster (t'Hart, Collier & Cohen 1990), das ein Verschmelzungsprodukt zweier benachbarter Gipfel ist, die ohne intervenierende Einbuchtung zwischen den F0-Maxima produziert werden. Gipfelsequenzen erzeugen zwei enge Foki und zeigen so oft eine inhaltliche Trennung zwischen den benachbarten Satzakkzentuierten Wörtern an. Hutmuster hingegen fassen die akzentuierten Wörter unter einem breiten Fokus zusammen (Peters et al. 2005a). Aus diesem Grund wird in *KIM* zwischen der Gipfelkonkatenation mit und ohne Einbuchtung unterschieden. Ähnliches gilt auch für Talkonturen. Weiterhin wird auch markiert, ob Gipfelkonturen innerhalb der Phrase einem Downstep folgen und ob es an prosodischen Grenzen zu einem Reset in dieser Downstepbewegung kommt (Neueinsatz der Deklinationslinie).

Der letzte Komplex erfasster melodischer Einheiten sind die F0-Verläufe mit denen Gipfel, Täler und ebene Konturen am Ende prosodischer Phrasen auslaufen. Unterschieden werden 12 Konturen:

- Talkonturen mit leicht bis mittelhoch vs. hoch steigenden Verläufen,
- Gipfelkonturen mit ebenem, leicht fallendem oder tief fallendem Verlauf nach dem Gipfelmaximum,
- ein ebener, leicht fallender oder tief fallender Verlauf, auf den jeweils ein sehr leichtes Ansteigen im Bereich von 5 bis 25 Hz folgt,
- vier Melodiemuster mit fallend-steigender melodischer Bewegung (kombinierte Konturen),
- auf ebene Konturen an Satzakkzentpositionen kann phrasenfinal ein ebenes Auslaufen, ein leichtes Fallen oder ein starkes Fallen folgen. Diese Konturen können wie bei den Gipfeln mit oder ohne final sehr leicht steigende Bewegung vorkommen, zeichnen sich aber durch ebene melodische Muster im Bereich der satzakzentuierten Silbe aus.

In der prosodischen Etikettierung werden auch andere suprasegmentelle Phänomene wie Wechsel der Sprechgeschwindigkeit oder der globalen Sprechlautstärke markiert. Folgende Phänomene werden erfasst:

- Registerwechsel
- Abweichungen von einer mittleren Sprechgeschwindigkeit
- Abweichungen von einer mittleren Sprechlautstärke
- längere Bereiche mit nicht-modaler Phonation

### **2.1.3.5      Erweiterte Etikettierung in den *Video Task* Daten**

Für die Analyse der tunfinalen prosodischen Grenzen sind zusätzliche manuelle Etikettierungen notwendig, die vor allem die zeitliche Beziehung der Sprachsignale beider Sprecher zueinander erfassen. Hierbei werden signalorientiert akustische Eigenschaften wie z.B. Überlagerung von Redebeiträgen beider Sprecher oder das Auftreten von Pausen bei Sprecherwechseln etikettiert und so die Voraussetzung für spätere automatische

Messungen dieser Parameter geschaffen. Informationen über überlagerndes Sprechen und die Dauer von Überlagerungen sowie über Pausen zwischen den Redebeiträgen verschiedener Sprecher sind wichtige Ergänzungen zu phonetischen Grenzmerkmalen wie Intonationsmustern oder segmenteller Dauerstruktur, denn sie spiegeln die Synchronisation von Redebeiträgen und somit das Verhalten der Sprecher zueinander und wider.

In der aktuellen Forschung werden vor allem zwei Transkriptionssysteme verwendet, die auf die Kodierung von Dialogstrukturen ausgerichtet sind: *DAMSL* (*Dialog Markup in Several Layers*) (Core & Allen 1997) und *GAT* (*Gesprächsanalytisches Transkriptionssystem*) (Selting et al. 1998). Der Schwerpunkt in der *DAMSL*-Dialogkodierung liegt auf der Abgrenzung und Symbolisierung von kommunikativen Aktivitäten im Rahmen der Sprechaktheorie. So werden unter anderem Diskurspartikel, Sprechakte und andere Äußerungsmerkmale symbolisch festgehalten. Das System ist also stark an die inhaltliche Interpretation des Dialogs gebunden. In *GAT* ist der interpretative Ansatz weniger stark ausgeprägt. Hier wird in einem Basistranskript der Wortlaut des Gesprächs festgehalten und durch phonetische Beschreibungen ergänzt. So werden prosodische Grenzen und Satzakkzentpositionen symbolisiert, F<sub>0</sub>-Muster rudimentär gekennzeichnet und verschiedene Arten von Turnübergängen markiert. In einem Feintranskript kann das Basistranskript durch beliebig viel zusätzliche Information ergänzt werden (z.B. Intensitätsverläufe oder feinere Beschreibungen melodischer Bewegungen).

Beide Systeme sind für die Aufbereitung der *Video Task*-Daten nicht geeignet. Kommunikative Akte, wie sie in *DAMSL* kodiert sind, sollen erst durch die phonetische Analyse und die nachfolgende Interpretation herausgearbeitet werden und nicht schon in die Datenaufbereitung eingehen. Die Anfertigung einer *GAT*-Transkription ist nicht sinnvoll, da die meisten in *GAT* vorgesehenen Arbeitsschritte (z.B. Transliteration und prosodische Etikettierung) bereits nach den Vorgaben der Kieler Konventionen zur Datenaufbereitung (Kohler et al. 1995) vorgenommen sind. Weiterhin ist es für die automatische Korpusabfrage wichtig, dass sich die zusätzlichen Etikettierungen in die Struktur der bereits vorhandenen Etikettierung einpassen. Deshalb wird ein neues Verfahren entwickelt und angewendet, das ohne eine inhaltliche Interpretation die akustische und phonetische Struktur von Dialogen - angepasst an die spezielle Fragestellung dieser Arbeit - erfasst. So werden signalorientiert akustische Merkmale wie z.B. Überlagerung von Redebeiträgen oder das Auftreten von Pausen beim Sprecherwechsel in die Etikettie-

zung der prosodischen Grenzen einbezogen. Anhand dieser zusätzlichen Etikettierung können in den nachfolgenden Korpusanalysen labelbasiert verschiedene Typen prosodischer Grenzen getrennt analysiert werden. Die ursprünglich in der Etikettierung der Dialoge verwendete Kategorie PGn (prosodische Grenze) wird nach folgenden Kriterien subklassifiziert:

- Position der prosodischen Grenze im Inneren oder am Ende eines Turns
- Auftreten von Pausen an prosodischen Grenzen innerhalb von Turns
- Auftreten von Pausen zwischen dem Redebeitrag eines Sprechers am Turnende und dem Beginn der Äußerung des Dialogpartners
- Überlagerung durch Redebeiträge des Dialogpartners
- Kategorisierung von Redebeiträgen als Rezeptionssignale oder Hintergrundkommentare

Die Nomenklatur der subklassifizierenden Etiketten orientiert sich am Englischen, trifft aber in den meisten Fällen auch auf die entsprechenden deutschen Wörter zu. So wird z.B. für turninterne Grenzen ein *i* in das Etikett an der prosodischen Grenze integriert (PG*i*). Das *i* steht für *internal* bzw. *intern*. Wenn keine Entsprechung zwischen den beiden Sprachen besteht, hat die englischsprachige Terminologie Vorrang.

Die nachfolgenden Abbildungen illustrieren verschiedene Typen prosodischer Grenzen und einige andere Elemente, die in der Etikettierung der *Video Task*-Dialoge differenziert werden. Die Illustrationen zeigen, stilisiert durch die schwarzen Balken, die Phrasen eines Dialogs zwischen Sprecher A und Sprecher B. Die Beispiele beziehen sich immer auf die Etikettierung der Rede von Sprecher A.

Die ursprünglich vorgenommene prosodische Etikettierung mit PROLAB unterscheidet nur zwei phonetisch verschiedene Typen symbolisch getrennter prosodischer Grenzen:

- Grenze mit nachfolgendem F0-Reset: PGn
- Grenze ohne nachfolgendes F0-Reset: =PGn

Eine weitere, syntaktisch motivierte Kategorie bezeichnet prosodische Grenzen an vsyntaktischen Abbruchstellen mit PG/. Diese Grenzmarkierungen werden in der ursprünglichen Etikettierung unabhängig von der Stärke des Einschnitts und der phonetischen Ausprägung vergeben, wenn ein deutlich wahrnehmbarer Einschnitt vorliegt. Auch die Position in der Äußerung (turnfinal vs. turnintern) spielt bei der Vergabe keine Rolle. Folgende Beispiele illustrieren die erweiterte Etikettierung.

**(1) Turninterne prosodische Grenze ohne überlagernde Rede**

Als erstes werden die Modifikationen beschrieben, die an turninternen prosodischen Grenzen zur Anwendung kommen. Das folgende Beispiel zeigt eine turninterne prosodische Grenze ohne Überlagerung durch lautliche Äußerungen des Dialogpartners und ohne eine Signalpause zwischen den Phrasen. Unter diesen Bedingungen wird das ursprüngliche Label PGn zu PGin geändert, wobei i für *intern* steht. Die Subklassifizierungen in PGin (mit nachfolgendem F0-Reset) und =PGin (ohne nachfolgenden F0-Reset) bleiben in beiden Fällen erhalten.

PGn → PGin

**Sprecher A**                    \_\_\_\_\_  
**Sprecher B**

Wenn sich eine Signalpause zwischen den prosodischen Phrasen befindet, wird das Label PGip vergeben. p steht für *Pause*.

PGn → PGip

**Sprecher A**                    \_\_\_\_\_  
**Sprecher B**

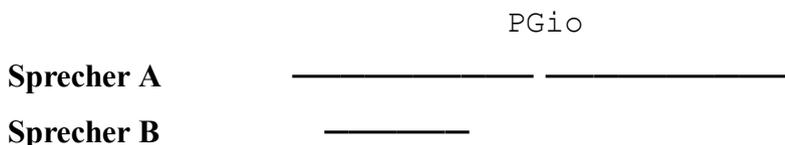
In der bisherigen Etikettierung sind prosodische Grenzen, die an syntaktischen Abbruchstellen auftreten mit PG/ gekennzeichnet. Die Entscheidung, ob ein syntaktischer Abbruch vorhanden ist, wird bereits während der Verschriftung der Dialoge getroffen und in der Transliteration durch spezielle Symbole markiert. In der segmentellen Etikettierung,

die auf der Transliteration basiert, wird die Position syntaktischer Abbrüche durch die Etiketten  $\$/+$  (Abbruch mit anschließender Wiederaufnahme) oder  $\$/-$  (Abbruch ohne Wiederaufnahme) gekennzeichnet. Die Integration der syntaktischen Information in das Notationssystem für prosodische Grenzen wird in der erweiterten Etikettierung aus drei Gründen aufgegeben.

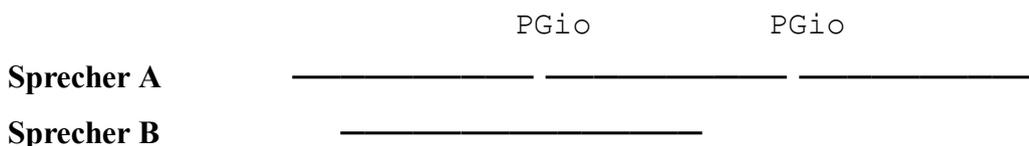
- (1) Die prosodische Etikettierung soll ausschließlich perzeptorisch relevante phonetische Merkmale gesprochener Sprache erfassen.
- (2) Das Symbolinventar soll so kompakt wie möglich gehalten werden.
- (3) Die Information über das Vorhandensein syntaktischer Abbrüche ist in der segmentellen Etikettierung vorhanden. Die zusätzliche Kennzeichnung auf der Ebene der prosodischen Etikettierung ist somit redundant.

**(2) Turninterne prosodische Grenze mit überlagernder Rede**

Dieses Beispiel zeigt die Etikettierung bei überlagerndem Sprechen der Dialogpartner. Die Äußerung von Sprecher B kann z.B. der Versuch einer Turnübernahme sein. Die Kategorie  $PG_{i0}$  erfasst nicht Überlagerungen durch Rezeptionssignale wie *mhm* (siehe 2.1.3.5. (11)).  $\circ$  in  $PG_{i0}$  steht für *overlap*.

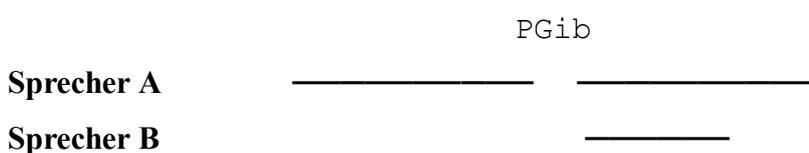


Bei längerem gleichzeitigem Sprechen können auch mehrere  $PG_{i0}$  in Folge auftreten, wie die folgende Abbildung zeigt. Auf die Etikettierung der prosodischen Grenze bei Sprecher B in dieser Konstellation wird später eingegangen.



### (3) Turninterne prosodische Grenze vor gleichzeitigem Redebeginn beider Sprecher

In den vorliegenden Sprachdaten ist folgendes Phänomen zu beobachten: Einer der Sprecher beendet seine Äußerung und es entsteht eine Gesprächspause, die wiederum mit einem gleichzeitigen Redebeginn beider Sprecher endet. Als *gleichzeitig* wird der Beginn des Sprechens unter der Bedingung markiert, dass die Äußerungen der Sprecher mit weniger als 100 ms zeitlichem Versatz anfangen. Es wird hierfür die Symbolisierung PGib verwendet. b in PGib steht für *beide* bzw. *both*.



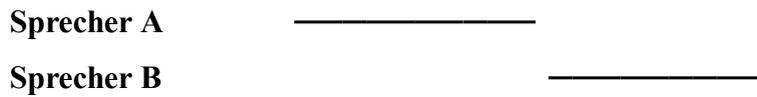
### (4) Turnfinale prosodische Grenzen

Als *turnfinal* wird eine Äußerungsgrenze klassifiziert, wenn im Anschluss an diese Grenze keine weiteren unmittelbaren Äußerungen desselben Sprechers folgen, sondern der Dialogpartner das Wort übernimmt. Die folgenden Abschnitte beschreiben die Etikettierung phonetisch unterschiedlicher turnfinaler prosodischer Grenzen.

### (5) Turnfinale prosodische Grenze ohne Überlagerung oder Pause

Der folgende Fall ist ein unmittelbarer Übergang ohne Überlagerung und ohne Pause zwischen den Äußerungen von Sprecher A und Sprecher B. *ohne Überlagerung* bedeutet, dass keine zeitliche Überlagerung der Äußerungen von Sprecher A und Sprecher B von mehr als 100 ms vorliegt. *Ohne Pause* bedeutet, dass zwischen dem Ende der Äußerung von A weniger als 100 ms vergehen bis B beginnt zu sprechen. Das t in PGtn steht für *Sprecherwechsel*. Für turnfinale prosodische Grenze wird generell nicht über Reset vs. fortlaufender Downstep unterschieden, da an dieser Stelle der Dialogpartner das Wort übernimmt. Dadurch entsteht normalerweise eine lange Sprechpause beim vormaligem Sprecher, der nun Hörer ist. Es ist nicht sinnvoll das Auftreten von Reset über eine so lange Unterbrechung hinweg zu kennzeichnen.

PG<sub>tn</sub>



**(6) Turnfinale prosodische Grenze mit Pause**

PG<sub>tp</sub> wird markiert, wenn zwischen dem Ende der Äußerung von A mehr als 100 ms vergehen, bis B beginnt zu sprechen. p in PG<sub>tp</sub> steht für *Pause*.

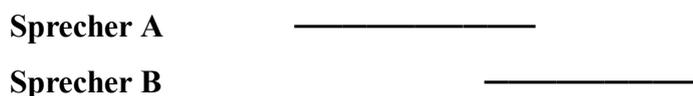
PG<sub>tp</sub>



**(7) Turnfinale prosodische Grenze mit Überlagerung**

In spontaner Konversation beginnen Hörer oft zu sprechen, während der Gesprächspartner noch redet. Die entstehende Überlagerung der Redebeiträge wird mit PG<sub>to</sub> markiert, sofern eine zeitliche Überlagerung der Äußerungen von Sprecher A und Sprecher B von mehr als 100 ms vorliegt. o in PG<sub>to</sub> steht für *overlap*.

PG<sub>to</sub>



**(8) Turnfinale prosodische Grenze nach gleichzeitigem Redebeginn beider**

**Sprecher**

Der hier illustrierte Fall ist ein Sprecherwechsel, der im Anschluss an einen Redebeginn passiert. In dieser Konstellation beendet Sprecher A seine Äußerung, während B weiter spricht. Auch hier wird an äußerungsfinalen Grenzen mit Überlagerung PG<sub>to</sub> etikettiert.





### (11) Prosodische Grenzen nach Rezeptionssignalen

Die Kennzeichnung einer Äußerung als Rezeptionssignal (*backchanneling*) (Schegloff 1982) ist im Gegensatz zu den übrigen vorgenommenen Klassifikationen nicht auf einer rein akustischen Basis möglich, da allein die Positionierung im Turn oder die Synchronisierung zur Äußerung des Dialogpartners nicht als Kriterium ausreicht, um eine Äußerung als Rezeptionssignal zu klassifizieren. Das liegt daran, dass Rezeptionssignale häufig überlagernd produziert werden und deshalb nicht ohne weiteres von überlagernden Äußerungen getrennt werden können, die nicht die Funktion der Rezeptionsbestätigung haben, sondern z.B. der Versuch einer Turnübernahme sein können. Um in späteren label-basierten Analysen die funktionale Klasse der Rezeptionssignale nicht mit den anderen zu untersuchenden Kategorien zu vermengen, sondern als separate Klasse behandeln zu können, werden folgende aus der Literatur bekannte definitorische Merkmale herangezogen, um eine Passage als Rezeptionssignal zu klassifizieren: Als Rezeptionssignale werden unaufgeforderte kurze Äußerungen wie *Aha*, *Mhm*, *Ja* oder *Find ich auch*. etc. verstanden. Sie haben zustimmenden Charakter und der Gesprächspartner reagiert nicht explizit darauf. Typische Rezeptionssignale stehen nicht unmittelbar vor einer Turnübernahme, sondern bestätigen den anderen Sprecher in seiner weiter fortlaufenden Rede. Phonetische Charakteristika sind eine herabgesetzte Lautstärke und tiefe Stimmlage (Schegloff 1982; French & Local 1983).

Auch Rezeptionssignale sind der Klasse der turnexternen Äußerungen zugeordnet. Die Grenzen nach Rezeptionssignalen werden mit  $P_{GeR}$  markiert wobei *e* für *extern* und *r* für *reception* bzw. *Rezeptionssignal* steht. In der Etikettierung der Äußerung von Sprecher B wird die Überlagerung durch Rezeptionssignale nicht gekennzeichnet (im Gegensatz zu Überlagerungen durch Äußerungen, die nicht als Rezeptionssignal klassifiziert sind).

PGer

**Sprecher A**

\_\_\_\_\_

**Sprecher B**

\_\_\_\_\_

### (12) Restkategorie

Die beschriebene subklassifizierende Etikettierung kann problemlos auf über 95% der bereits im Korpus etikettierten prosodischen Grenzen angewendet werden. In Zweifelsfällen wird das Label PG<sub>x</sub> vergeben. Zweifelsfälle sind z.B. überlagernde Äußerungen beider Sprecher ohne Klarheit darüber, wer der eigentliche Turninhaber ist, oder Unsicherheit darüber, ob ein bestimmter Abschnitt als Rezeptionssignal klassifiziert werden muss. Auch prosodische Grenzen nach verbalen Äußerungen, die in Lachen enden, werden mit PG<sub>x</sub> etikettiert.

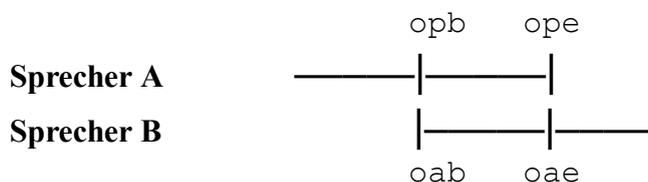
### (13) Zeitliche Grenzen von Überlagerungen und Pausen

Für eine phonetische Klassifikation verschiedener Typen prosodischer Grenzen im Kontext der Sprecherinteraktion ist es notwendig, Information über die Dauern von überlagerndem Sprechen und von Pausen zwischen Redebeiträgen verschiedener Sprecher einzubeziehen. Für eine zuverlässige automatische Bestimmung von Überlagerungs- und Pausendauern stellt die ursprüngliche Etikettierung nicht genügend Information zur Verfügung, deshalb werden manuell Marken eingefügt, die solche Messungen ermöglichen.

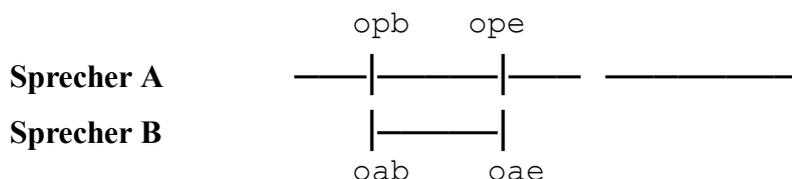
### (14) Zeitliche Abgrenzung von Überlagerungen

Der hier illustrierte Fall ist ein Sprecherwechsel mit Überlagerung (PG<sub>t</sub>o). Der Beginn der Überlagerung wird für den Turninhaber (Sprecher A) mit op<sub>b</sub> (*overlap passiv begin*) bezeichnet, das Ende der Überlagerung mit op<sub>e</sub> (*overlap passiv end*). Für den Kanal des Sprechers, der das Rederecht übernimmt (Sprecher B) werden die Etiketten oab (*overlap active begin*) und oae (*overlap active end*) vergeben. Als *passiv* wird also der Sprecher bezeichnet, der das Rederecht zum Zeitpunkt des Beginns der Überlagerung hat (der

Überlagerte); als *aktiv* derjenige, der zu reden beginnt, während der Dialogpartner selbst noch spricht (der Überlagernde).

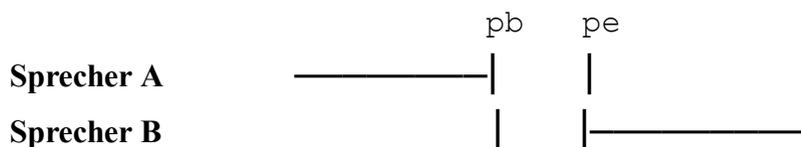


Die folgende Abbildung zeigt einen Fall turninterner Überlagerung der Rede von Sprecher A. Auch wenn die überlagernde Rede aus mehreren Phrasen besteht, werden Anfang und Ende der Überlagerung mit den Etiketten *opb*, *ope* bzw. *oab*, *oae* markiert. Die Positionierung der Marken für *opb* und *oab*, d.h. der Anfang der Überlagerung orientiert sich auf beiden Kanälen am aktiven Sprecher (hier Sprecher B). Die Marken *ope* und *oae* werden dort gesetzt wo die überlagernde Rede aufhört.



### (15) Zeitliche Abgrenzung von Pausen

Die Abbildung zeigt den Fall eines Sprecherwechsels mit einer Pause zwischen den Äußerungen von Sprecher A und Sprecher B (*PGtp*). Die Dauer von Pausen wird nur auf einem Kanal, nämlich auf dem des Sprechers, der das Rederecht abgibt (hier A), markiert. Die Labels *pb* bzw. *pe* bedeuten *pause begin* und *pause end*. Atmungsphasen werden genauso behandelt wie Pausen.



### 2.1.3.6 KielDat Datenbank

Die Erstellung einer Datenbank im *KielDat*-Format (Pätzold 1997) ist ein automatisierter Weiterverarbeitungsschritt der fertigen Etikettierdateien. Die Dateien werden über Funktionen der Programmiersprache *awk* (Close et al. 1993) in das Datenbankformat der *KielDat* Datenbank konvertiert. Hierbei handelt es sich um ASCII-Dateien, in denen die Informationen aus den Etikettierdateien wortbasiert in verschiedenen Feldern angeordnet werden. Auf die so neu strukturierten Daten kann auch ohne umfangreiche Computerkenntnisse durch etwa 60 verschiedene Datenbankfunktionen zugegriffen werden. Die Funktionen beinhalten Konvertierungen wie die Umrechnung von Hertz in mel, Extraktionsfunktionen für die Suche nach bestimmten Etiketten und mathematische Funktionen, z.B. für die Errechnung verschiedener statistischer Werte.

### 2.1.4 Zusammenfassung der Datenbasis

Folgende Daten gehen in die Korpusanalyse ein:

- das spontansprachliche Korpus der CD-ROM Vol. 2 und Vol. 3 (IPDS 1995, 1996) aus dem Verbmobil-Szenario der Terminabsprache: 77 Dialoge aus 11 Dialogsitzungen mit 22 Sprechern (13 männlich, 9 weiblich), 2,5 Stunden, 25.000 fortlaufende Wörter, 5100 etikettierte prosodische Grenzen,
- 6 überlappende Dialoge aus dem *Video Task* Szenario (LINDENSTRASSE-Korpus), (Peters 2001, 2005a): 12 Sprecher, 4 männlich, 8 weiblich, 1,5 Stunden, 13.500 fortlaufende Wörter, 3900 etikettierte prosodische Grenzen.

## 3 Turninterne prosodische Grenzen in flüssiger Sprache

### 3.1 Stand der Forschung

Der folgende Literaturüberblick konzentriert sich auf prosodische Grenzen, die das Ende prosodischer Phrasen in flüssiger Sprache markieren, sowie um die interne Struktur prosodischer Phrasen. Zuerst werden die grundlegenden Einheiten in der Beschreibung und Modellierung eingeführt, in weiteren Abschnitten Formen und Funktionen prosodischer Grenzen und prosodischer Phrasierung dargestellt.

#### 3.1.1 Grundlegende Beschreibungseinheiten

Keine phonetische Analyse fortlaufender Sprache kommt ohne eine Beschreibung dessen aus, was in dieser Arbeit *prosodische Phrase* genannt wird. Innerhalb längerer Äußerungen ist die prosodische Phrase die zentrale prosodische Strukturierungseinheit und wird in vielen Untersuchungen als die kognitive Verarbeitungsdomäne für Intonationskonturen angesehen. Prosodische Phrasen in flüssiger Sprache werden durch prosodische Grenzen delimitiert.

Der heute am häufigsten verwendete Begriff für die Einheit, ist *Intonationsphrase* (*intonation phrase*) und geht auf Selkirk (1978) zurück. Vergleichbare Begriffe sind *rhetorische Phrase* (v. Essen 1964), *tone group* (Halliday 1967; O'Connor & Arnold 1973), *prosodische Phrase* (Kohler 1995a), *Tongruppe* (Pheby 1984) oder *Äußerungseinheit* (Schwitalla 1997). Da an der phonetischen Konstitution dieser Einheit nicht nur die Intonation, sondern auch segmentelle Dauerrelationen, Intensitätsmuster und phonatorische Veränderungen beteiligt sein können, ist für phonetisch orientierte Beschreibungen der Terminus der *prosodischen Phrase* m.E. am geeignetsten.

Aufgrund der hohen Variabilität sprachlichen Verhaltens und der vielen unterschiedlichen Textsorten und Stilebenen gibt es keine einheitliche Definition der prosodischen Phrase. In fast allen Beschreibungen finden sich aber zumindest zwei phonetische Parameter, die als definitorische Kriterien herangezogen werden. Zum einen ein globaler Melodiebogen, der sich über der prosodischen Phrase ausprägt und ihr einen

melodischen Zusammenhalt gibt, und zum anderen eine Reihe von Merkmalen, die am Ende prosodischer Phrasen vorkommen und die zu einer Abgrenzung zur nachfolgenden Phrase führen. Diese beiden Parameter stehen in enger Verbindung mit den für diese Arbeit zentralen Konzepten der phonetischen Kohäsion im Inneren prosodischer Phrasen und der Separierung zwischen benachbarten Phrasen. Als erstes wird nun der interne Aufbau prosodischer Phrasen beschrieben, der ihnen eine Gestalt verleiht und sie als perzeptorische Einheit erscheinen lässt. Darauf folgt ein Abschnitt über die Delimitationssignale, die sich an Grenzen von Phrasen finden und zur Separierung von Folgephrasen führen. Es sei vorangestellt, dass die Größe einer prosodischen Phrase sehr variabel ist und von einer Silbe bis zu komplexen syntaktischen Konstruktionen reicht, woraus sich Probleme für eine einheitliche Definition dieser Einheit ergeben.

Die Basiseinheiten der internen Struktur prosodischer Phrasen, die auch heute noch verwendet werden, wurden vor allem durch die Britische Schule bestimmt (z.B. Halliday 1967; O'Connor & Arnold 1973). Die Untergliederung prosodischer Phrasen (*tone groups*) orientiert sich an deren Akzentstruktur. Jede komplette Phrase enthält als obligatorisches Element mindestens eine satzakzentuierte Silbe und wird in mindestens zwei Basiseinheiten unterteilt: den Teil, der der letzten satzakzentuierten Silbe vorangeht und den Rest der Phrase. Die betonte Silbe des letzten Satzakkzents wird *nucleus* (O'Connor & Arnold 1973) oder *tonic syllable* (Halliday 1967) genannt. Halliday bezeichnet den Teil der vor der *tonic syllable* als *pretonic segment* und den Rest der Phrase als *tonic segment*. O'Connor & Arnold unterteilen Hallidays *pretonic segment* in *prehead* und *head* und das *tonic segment* in *nucleus* und *tail*. Der *head* beginnt mit der ersten satzakzentuierten Silbe vor dem *nucleus* und endet mit der Silbe unmittelbar vor dem *nucleus*. Der *prehead* umfasst alle Silben vor dem *head*. Als *tail* werden alle Silben bezeichnet, die dem *nucleus* bis zum Ende der Phrase folgen. So kann die Terminologie Phrasen mit mehreren Satzakkzenten besser erfassen.

Der melodische Verlauf über Nukleus und möglichem Nachlauf wird auch *nuclear tone (nukleare Kontur)* genannt (Cruttenden 1986). Grabe (1998) stellt hinsichtlich dieser Terminologie allerdings, eine Inkonsistenz fest, denn die *nukleare Kontur* ist nicht, wie die Bezeichnung nahelegt, der tonale Verlauf auf dem Nukleus, sondern bezieht auch den (meist vorhandenen) Nachlauf mit ein. Um terminologische Klarheit zu gewährleisten wird in der vorliegenden Arbeit in Anlehnung an Kohler

(1995a) und Gilles (2005) folgende Bezeichnung gewählt: Der *Nukleus* ist (wie bei Halliday 1967) die gesamte Silbenfolge von der letzten akzentuierten Silbe bis zum Ende der prosodischen Phrase. Die erste Silbe eines Nukleus (die immer die akzentuierte Silbe ist) heißt *Nukleussilbe*. Zur Kennzeichnung der Silbenfolge nach der Nukleussilbe wird *Nachlauf* verwendet.

In den intonatorischen Beschreibungen, die im Gefolge des Tonsequenzmodells der autosegmentell-metrischen Phonologie und entstanden sind, wird die Konstituentenstruktur prosodischer Phrasen folgendermaßen beschrieben. Jede Phrase ist mit einem *tune* assoziiert: einer linearen Sequenz von Tönen auf zwei Ebenen. Jeder *Tune* enthält einen oder mehrere *pitch accents*, die als Einzeltöne oder bitonale Sequenzen realisiert sind und die Fokussierung bzw. Satzakkentuierung bestimmter Wörter bewirken. Meist werden zwei Ebenen der Phrasierung unterschieden: die (kleine) *intermediate phrase* (Intermediärphrase) und die (große) *intonation phrase* (Intonationsphrase). Eine Intonationsphrase besteht aus mindestens einer Intermediärphrase, kann aber auch aus einer Aneinanderreihung mehrerer Intermediärphrasen zusammengesetzt sein. Die *boundary tones* (Grenztöne an Intonationsphrasen) beschreiben den tonalen Verlauf vom letzten Satzakkent (*phrase accent*) bis zum Ende der jeweiligen Intonationsphrase, während der Verlauf nach einem Satzakkent vor einer intermediären Phrasengrenze oft als *phrasal tone* bezeichnet wird (vgl. Pierrehumbert 1980; Ladd 1996).

### 3.1.2 Phonologische Beschreibungen

Durch die Literatur zieht sich eine Trennung zwischen phonologisch orientierten Analysen, die linguistisch modellierend vorgehen, und phonetisch orientierten Untersuchungen, die eher messend und beschreibend sind. Die phonologische Herangehensweise bedingt eine Ausrichtung auf hierarchische linguistische Systemüberlegungen. Form und Funktion der prosodischen Phrasierung wird vor allem im Kontext der morphologischen und syntaktischen Ebene betrachtet, und die entwickelten Modelle bemühen sich um die Schaffung direkter Beziehungen zwischen diesen Ebenen. Ein typisches Vorgehen bei der Erarbeitung phonologischer Modelle ist, dass von bereits definierten Kategorien, wie Wortgrenzen oder syntaktischen Grenzen ausgegangen wird, diese in einer Hierarchie angeordnet werden und nach den phonetischen Korrelaten dieser Grenzen

in konstruiertem lesesprachlichem Material gesucht wird. Die gefundenen Korrelate werden dann in abstrakte, formalisierte phonologische Repräsentationen überführt, die in das linguistische Regelwerk integriert werden.

Ein solches Vorgehen findet sich z.B. bei Trubetzkoy (1939), der verschiedene Junkturphänomene und deren phonetische Delimitationssignale beschreibt, die zur Abgrenzung von Wörtern und Morphemen dienen; so z.B. der morpheminitiale Glottalverschluss vor Vokalen (harter Vokaleinsatz). Bei Trubetzkoy gibt es allerdings noch keine Beschreibungen tonaler Grenzsingale an prosodischen Grenzen, wie sie sich später z.B. bei Trager & Smith (1951) finden, die zwischen einer vierstufigen Junktur mit ansteigender Separierung unterscheiden:

- (1) internal open (plus) juncture (Junktur in *stress groups*)
- (2) single-bar juncture (terminale Junktur mit ebenem tonalen Verlauf)
- (3) double-bar juncture (terminale Junktur mit steigendem tonalen Verlauf)
- (4) double-cross juncture (terminale Junktur mit fallendem tonalen Verlauf)

Der erste Junkturtyp betrifft die Wortstrukturierung und ist phonetisch auf der segmentellen Ebene kodiert durch Lautdauer und allophonische Varianten und dient zur Trennung zwischen Wortpaaren wie *nitrate* und *night-rate*, wobei für letzteres eine *internal open juncture* zwischen *night* und *rate* angesetzt wird. Die drei terminalen Junkturen hingegen beziehen sich auf syntaktische Strukturen und sind auf prosodische Mittel gegründet, insbesondere auf die Intonation aber auch auf Pausen und segmentelle Längung. Diese Junkturen dienen entweder der Strukturierung von Äußerungen in einzelne Phrasen (vor allem die *single-bar juncture*) oder zeigen pragmatisch unterschiedliche Typen von Äußerungsenden an. Bei der Festlegung dieser Junkturen war das strukturalistische Prinzip der distinktiven phonologischen Opposition in Zitierformen mit konstruierten Sätzen maßgebend, und es wird eine weitgehende Übereinstimmung zwischen der Position der letzten drei Junkturtypen und hohen syntaktischen Grenzen (meist Sätzen) angenommen. In dieser Modellierung sind mehrere Ansätze enthalten, die auch für die vorliegende Studie und viele andere aktuelle Untersuchungen wichtig sind. Erstens wird die Einteilung – zumindest für die letzten drei Junkturtypen - an der Form des phrasenfinalen Melodieverlaufs festgemacht, und zweitens ist über diese Einteilung eine pragma-

tische Form-Funktions-Beziehung angelegt. Diese Form-Funktions-Beziehung ist phonetisch kodiert, und zeigt sowohl phonetisch als auch in der funktionalen Zuordnung einen zunehmenden Grad von Separierung. Weiterhin definieren Trager & Smith folgende Einheiten, die sich in vielen neueren Modellierungen wiederfinden:

- das *phonemic word*, eine morphologisch definierte phonologische Einheit ohne interne Junktur (z.B. das Wort *nitrate*),
- die *phonemic clause*, eine minimale komplette Äußerung ohne interne terminale Junktoren mit einem primären Satzakzent und einer finalen terminalen Junktur,
- die *phonemic phrase*, als phonologische Abstraktion der *phonemic clause*, definiert als eine Reihe von Segmenten mit oder ohne intervenierende interne Junktoren, mit einem primären Akzent und mehreren optionalen Nebenakzenten.

Eine vergleichbare Einteilung nimmt Selkirk (1978, 1984) vor. Ihre phonologische Einteilung sieht folgende Bestandteile vor.

*Utterance* (U)

*Intonation Phrase* (I)

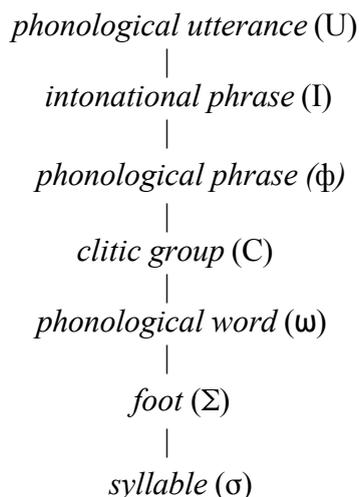
*Phonological Phrase* ( $\Phi$ )

*Prosodic Word* ( $\omega$ )

*Foot* ( $\Sigma$ )

*Syllable* ( $\sigma$ )

Die Einheiten dieses Modells stehen in einer hierarchischen Beziehung, die durch folgende Regeln ausgedrückt ist: Eine prosodische Konstituente der Stufe n ist unmittelbar dominiert durch eine Komponente der Stufe n+1. Jede Konstituente der Stufe n+1 lässt sich ohne Rest in Komponenten des Ranges n auflösen (*strict layer hypothesis*; Selkirk 1984). Selkirks Hierarchie überlappt zum großen Teil mit dem häufig zitierten System von Nespor & Vogel (1986):



Das Hauptinteresse solcher Systematisierungen gilt also der Beziehung zwischen morpho-syntaktischen und prosodischen Strukturen, die im Rahmen phonologischer Regeln aufeinander abgebildet werden sollen, was allerdings bisher nicht erreicht wurde:

*I think it is fair to say that up to now there has been no general theory of the mapping between syntactic structure and prosodic structure to accompany this theory of prosodic structure qua representation.* (Selkirk 1986: 384).

Auch de de Pijper & Sanderman (1994) und Price et al. (1991) versuchen in Perzeptions-experimenten eine systematische Beziehungen zwischen den phonologischen Ebenen von Nespor & Vogel und phonetischen Merkmalsbündeln herzustellen und kommen zu keiner phonetischen Systematik, die die Hierarchie von Nespor und Vogel stützen kann.

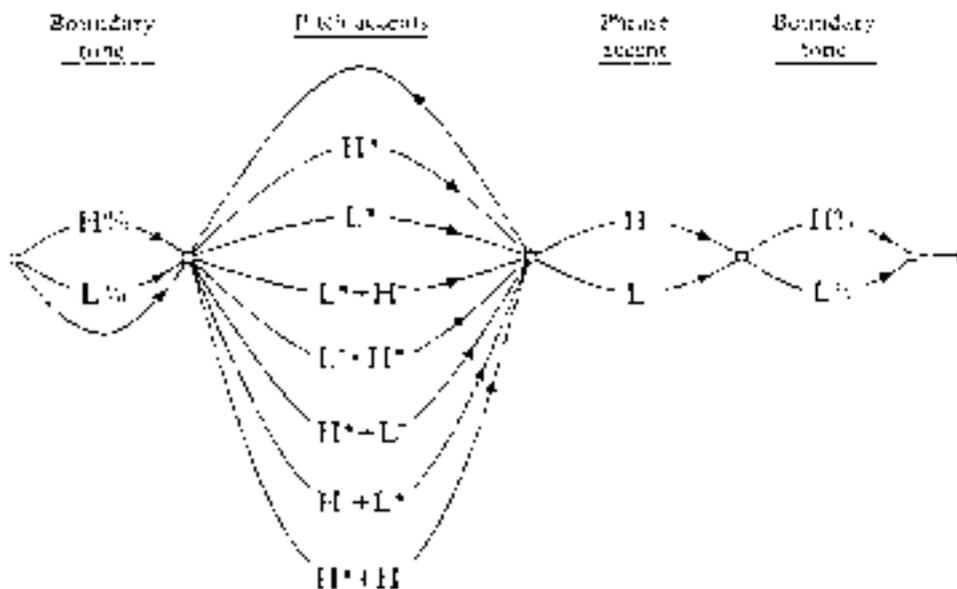
Die Vorgehensweisen von Trager & Smith, Selkirk und Nespor und Vogel zeigt die modellierende phonologische Ausrichtung, die komplexe Äußerungen in immer kleinere Konstituenten zerlegt. Es wird bei der Definition der Konstituenten immer ein Bezug zu morphosyntaktischen Strukturen definiert. Auf der anderen Seite werden phonetische Strukturen wie Pausen oder Grundfrequenzmuster, sowie pragmatische Funktionen berücksichtigt, ohne dass ein wirklich perzeptorischer Ansatz verfolgt wird oder systematische Bezüge zur Gesprächsstruktur entstehen. Diese Mischung definitiver Kriterien zeigt sich in zwei Zitaten aus Selkirk (1978:130):

*There is one syntactic generalization concerning the ( $\Phi$ ) which may be important in a model of speech perception. Every ( $\Phi$ ) ends in the **head** of a syntactic phrase, that is, the structuring of an utterance into phonological phrases provides fairly direct and easy access to a certain type of information – that about „headhood“, a type of information which seems particularly important in unravelling the syntax and semantics of a sentence.*

und

*One of the characteristics of the prosodic unit intonational phrase (I), is that it is the domain over which an intonational contour is "spread".*

Die eben skizzierten Ansätze sind die Basis für die heute am weitesten verbreitete Sichtweise auf prosodische Phänomene: das autosegmentell-metrische Modell. Das Modell geht zurück auf eine Studie zum amerikanischen Englisch von Pierrehumbert (1980) und wurde in vielen Arbeiten bis heute verfeinert, verändert und auf unterschiedlichste Sprachen übertragen (z.B. Beckman & Pierrehumbert 1986; Silverman et al. 1992; Ladd 1996). Dieses Modell ist nicht nur auf die Beschreibung hierarchischer Einheiten und deren Delimitation ausgerichtet, sondern modelliert melodische Muster über eine abstrakte phonologische Repräsentation durch Sequenzen aus zwei Tönen: einem Hochton (*H*) und einem Tiefton (*L*). Diese Töne werden im Rahmen eines sprachspezifischen Regelwerks verknüpft und repräsentieren den melodischen Verlauf über Intonationsphrasen. Abbildung 1 zeigt mögliche intonatorische Strukturen einer Intonationsphrase (IP) im amerikanischen Englisch:



**Abbildung 1:** Zustandsgrammatik (*finite-state grammar*), durch die sämtliche Intonationsphrasen des amerikanischen Englisch generiert werden können (vgl. Pierrehumbert 1980).

Durch diese Grammatik werden Sequenzen von hohen und tiefen Tönen generiert, die in Intonationsphrasen (*intonational phrase*) vorkommen können. Beckman & Pierrehumbert setzen eine Einheit an, die sie *intermediate phrase* nennen und die mit der *phonological phrase* von Nespor & Vogel (1986) vergleichbar ist. Eine volle Intonationsphrase kann aus mehreren intermediären Phrasen bestehen. Jede intermediäre Phrasen endet mit einem Grenzton (*phrase accent*) eine volle Intonationsphrase erhält zusätzlich noch einen finalen Grenzton den *boundary tone*. Die Intonationsphrase ist also vor allem durch die Sprechmelodie definiert und beschreibt Abschnitte von Äußerungen, die durch Grenztöne, aber auch durch mögliche Pausen delimitiert sind:

*Like other researchers, we will take the melody for an intonational phrase to be the "tune" whose internal makeup is to be described. As a rule of thumb, an intonational phrase boundary (transcribed here as %) can be taken to occur where there is a nonhesitation pause or where a pause could be felicitously inserted without perturbing the pitch contour. (Pierrehumbert 1980:19).*

Auf eine umfassende Darstellung des autosegmentell-metrischen (AM) Ansatzes soll hier verzichtet werden, da vorliegende Arbeit nur wenige Verbindungen zur AM-Phonologie und dessen Tonsequenzansatz hat und die intonatorischen Analysen im Rahmen des *Kieler Intonationsmodells* angesiedelt sind.

Auf der Basis der AM-Phonologie ist das Etikettierungssystem ToBI entwickelt worden, das zu einem weit verbreiteten System zur prosodischen Transkription geworden ist und für viele Sprachen adaptiert wurde (Silverman et al. 1992; Beckman & Ayers 1993; Batliner & Reyelt 1994; Grice & Benzmüller 1995; Grice & Baumann 2002). Auch die stark messphonetisch orientierte Transkription der tonalen Sequenzen, die die *pitch accents* konstituieren, soll hier nicht weiter thematisiert werden. Wohl aber wird kurz auf die getrennt von den tonalen Mustern etikettierte Ebene der *break indices* eingegangen, da diese unterschiedlich starke prosodische Einschnitte erfassen soll. Hieraus leitet sich auch der Name des Systems ab: tonale Ebene (*To*) und *break indices* (*BI*) (vgl. Price et al. 1991). Diese beiden Ebenen wurden von getrennten Forschungsgruppen entwickelt.

Folgende wortbasierte Systematik von fünf *break indices* wird von Ladd (1996) beschrieben und kann als Basis für etliche modifizierte Formen gelten, die sich in der Literatur finden: *Break index 0* beschreibt die größtmögliche phonetische Integration zwischen angrenzenden Wörtern und wird z.B. zwischen Wörtern etikettiert, die eine klitische Gruppe bilden. *Break index 1* wird zwischen Wörtern etikettiert, die zwar keine klitische Gruppe bilden, aber auch weiter keine Merkmale phonetischer Separierung voneinander zeigen (*default*-Wortgrenze). *Break index 2* bedeutet einen hörbaren Einschnitt, ohne tonale Korrelate. *Break index 3* wird an intermediären Grenzen etikettiert, *break index 4* am Ende von Intonationsphrasen. Die *break indices 3* und *4* implizieren das Auftreten von Grenztönen. Die Grenztöne dieser Phrasen beschreiben die Kontur vom letzten Ton des nuklearen Tonakzents bis zum Ende der jeweiligen Phrase.

Diese Einteilung birgt verschiedene Probleme. Erstens werden die systemorientierten phonologischen Kategorien der tonalen Ebene nicht sauber von den z.T. perzeptorisch definierten *break indices* getrennt. Dies führt zu undurchsichtigen Wechselbeziehungen zwischen den Beschreibungsebenen. So kann z.B. unter einem perzeptorischen Gesichtspunkt die Grenze nach einer intermediären Phrase die gleiche Stärke haben wie die Grenze nach einer Intonationsphrase, ohne dass dies durch das System erfassbar

ist. Zweitens ist das System, was die phonetische Adäquatheit betrifft, angreifbar, da hinsichtlich der phonetischen Struktur prosodischer Grenzen hauptsächlich tonale Exponenten der *break indices* definiert werden, dies gilt in besonderem Maße für die *break indices 3* und *4*, die im Kern das erfassen, was in der vorliegenden Arbeit als flüssige prosodische Grenzen verstanden wird. Verschiedenste Arbeiten zeigen aber, dass bei der Beschreibung und Modellierung auch die temporale Ebene, Pausen und Atmungsphasen, phonatorische Veränderungen und Intensitätsverläufe systematisch einbezogen werden müssen. Die Bindung des tonalen *tiers* an das *break index tier* geht so weit, dass in GToBI die *break indices 3* und *4* nicht per Hand etikettiert werden, sondern aus den bereits auf der tonalen Ebene etikettierten Grenztönen abgeleitet und automatisch eingefügt werden. So verliert die Ebene der *break indices* ihre Eigenständigkeit. Drittens wird in der Etikettierung großer spontansprachlicher Korpora deutlich, dass eine definitorische Abhängigkeit der prosodisch gesteuerten Äußerungsgliederung von Syntax und Morphologie (vor allem bei *break index 3* vs. *break index 4*) die Etikettierer vor unlösbare Probleme stellt. In natürlicher Sprache werden nämlich häufig agrammatische Strukturen produziert, die nicht korrigiert werden, aber durchaus mit kompletten melodischen Bögen intoniert sind und im Rahmen des Dialogkontextes vom Hörer verstanden werden. Hier zeigt sich eine partielle Unabhängigkeit syntaktischer und prosodischer Strukturen, die mit dem *break index*-System nicht erfasst werden kann. Auf der anderen Seite stehen komplette syntaktische Strukturen (z.B. vollständige Sätze), die ohne perzeptiv relevante Einschnitte von der nächsten syntaktischen Einheit gefolgt werden. Auch unter diesem Gesichtspunkt ist eine von linguistischen Systematiken unabhängige Etikettierung auf perzeptiver Basis vorzuziehen.

### 3.1.3 Phonetische Beschreibungen

Die phonetischen Untersuchungen, die in diesem Abschnitt beschrieben werden, sind im Gegensatz zu den phonologischen Ansätzen und Systematisierungen stärker an perzeptiv relevanten Signaleigenschaften orientiert und gelangen nicht zu Systemen mit dem Anspruch auf eine vollständig systematisierte Beschreibung des Phänomenbereichs. Es werden vor allem Untersuchungen referiert, die einen Zusammenhang zwischen ganzheitlich perzipierten Strukturen und kohäsiven prosodischen Mitteln herstellen, sowie solche, die separierende prosodische Parameter an den Grenzen der prosodischen Einheiten untersuchen.

Eine extreme phonetische, physiologisch ausgerichtete Position vertritt Lieberman (1967), der eine Opposition zwischen nur zwei Typen von *breath groups* aufbaut; der *marked breath group* und der *unmarked breath group*. Wobei die unmarkierte *breath group* aufgrund des abfallenden subglottalen Drucks eine fallende Intonation zeigt, die als Funktion des abfallenden subglottalen Drucks bei einer relativ konstanten Spannung der laryngalen Muskulatur angesehen wird. Die *marked breath group* weist ein steigendes Intonationsmuster auf, das durch einen Anstieg der Spannung in der laryngalen Muskulatur verursacht wird, das ein Abfallen von F0 verhindert. Dieser mechanistische Ansatz geht bereits auf Sweet (1877, 1890) zurück, der bemerkt, dass Sprecher nicht in der Lage sind, mehr als eine gewisse Anzahl von Lauten zu erzeugen, ohne die Luft in den Lungen zu erneuern. Sicher ist eine solche Vereinfachung nicht geeignet, die Strukturierung von Äußerungen in prosodische Phrasen oder die komplexen Merkmalsbündel zu erklären, die an prosodischen Grenze auftreten. Ohala & Hirano (1967) können in EMG-Untersuchungen experimentell belegen, dass die Larynxmuskeln ständig an der Steuerung der Grundfrequenz mitwirken und sich nicht wie von Lieberman postuliert in der unmarkierten Position passiv verhalten. Andererseits ist die Notwendigkeit der Atmung aber mit großer Wahrscheinlichkeit ein sprachentwicklungsgeschichtlicher Ausgangspunkt für die Untergliederung von Äußerungen in Phrasierungseinheiten, die sich sprachspezifisch immer weiter fortentwickelt hat.

Neuere Untersuchungen zeigen, dass die Grenzen von prosodischen Phrasen durch Bündel separierender phonetischer Merkmale gebildet werden, die von Hörern als Einschnitt perzipiert werden. Einige Studien befassen sich mit der Bündelung der

Merkmale. Die meisten Untersuchungen greifen allerdings einzelne phonetische Merkmale heraus und führen separate Analysen durch. Folgende Merkmale werden in der Literatur beschrieben:

- Pausen und Atmungsphasen
- Tonhöhenveränderungen
- Dehnung von Silben am Phrasenende
- globale Verlangsamung der Sprechgeschwindigkeit
- F0-Reset bzw. Neueinsatz der Deklinationslinie
- Phonatorische Veränderungen
- Intensitätsveränderungen
- Beschleunigung der Silben am Phrasenbeginn
- Ausbleiben von Koartikulation über prosodische Grenzen hinweg
- Registerwechsel

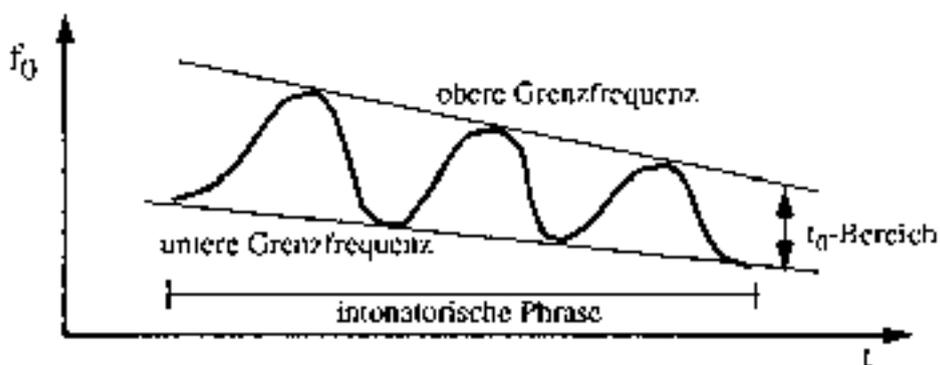
Alle diese phonetischen Parameter haben die Eigenschaft, eine wahrnehmbare Veränderung - einen nicht-kontinuierlichen phonetischen Ablauf - zu bewirken. Ihnen kann also ein Signalcharakter zugeschrieben werden, der auf prosodischer Basis eine Separierung bewirkt. Nähere perzeptorische Ergebnisse zur separierenden Wirkung der genannten phonetischen Faktoren aus unterschiedlichen Experimenten sind in Abschnitt 3.3.2 zu finden. Die meisten der genannten Parameter führen lokal an der Fuge zwischen zwei Phrasen zu einem perzipierbaren Einschnitt. Auf der anderen Seite stehen prosodische Strukturen, die sich global über die prosodische Phrase legen und diese zusammenhalten – ihr Kohäsion verleihen.

Als kohäsives Mittel wird in diversen Studien vor allem ein globales Melodiemuster über der prosodischen Phrase angenommen. Gilles (2005:6) beschreibt eine *globale Intonationskontur ohne prosodische Brüche, die vergleichbar einem musikalischen Motiv eine interne intonatorische Kohäsion stiftet*. Selting (1993) nennt diese Melodieverläufe *kohäsive Akzentsequenzen*. In Untersuchungen von Couper-Kuhlen (1983) und Schwitalla (1997) ist von *globalen Intonationsbögen* die Rede. Welche Formen melodischer Verläufe explizit kohäsionsstiftend sind, wird in der Literatur meist nicht definiert. Lediglich das Hutmuster (t'Hart et al. 1990; Peters et al. 2005a) wird

weithin als intonatorisches Verschmelzungsprodukt zweier Gipfelkonturen angesehen, das nur im Inneren prosodischer Phrasen vorkommt und Kohäsion stiftet. Peters et al. (2005a) gehen der Frage nach, ob die Verschmelzung zweier Satzakkente zu einer einzigen Gipfelkontur eine zeitlich gesteuerte Konsequenz aus der schnellen Aufeinanderfolge zweier Satzakkente ist oder ob der Hutmusterbildung semantische Funktionen zugrunde liegen. Sie untersuchen hierfür über 400 prosodische Phrasen aus Spontansprache, die zwei perzeptiv etwa gleich prominente Satzakkente (realisiert als Gipfelkonturen) enthalten, und kommen zu folgenden Ergebnissen. Das Verhältnis von Hutmustern zu Gipfelsequenzen (mit F0-Einbuchtung zwischen den Maxima) liegt bei etwa 2 zu 1. Die Dauern zwischen den satzakzentuierten Silbennuklei ist für die Gruppe der Hutmuster und der Gipfelsequenzen kaum unterschiedlich. Der Abstand in Hutmustern liegt bei durchschnittlich 421 ms, der in Gipfelsequenzen bei 482 ms. Somit wird eine zeitliche Steuerung der Hutmusterbildung ausgeschlossen. In inhaltlichen Interpretationen von Korpusbeispielen wird dann der Frage nachgegangen, ob die akustische Kohäsion von Hutmustern mit einer inhaltlichen Kohärenz der vereinigten Elemente einhergeht. Peters et al. zeigen, dass Hutmuster eher auftreten, wenn bereits bekannte Information wiedergegeben wird, während Gipfelsequenzen eher auftreten, wenn neue Information eingeführt wird. Hutmuster bilden hierbei einen breiten Fokus über die beiden satzakzentuierten Einheiten und stellen Alternativen neutral nebeneinander. Z.B. in der Frage *Bei Ihnen oder bei mir?*. Gipfelsequenzen hingegen bilden einen Doppelfokus und stellen Alternativen in ihrer Unterschiedlichkeit gegenüber. Hutmuster fassen also die akzentuierten Komponenten sowohl akustisch als auch inhaltlich zusammen, Gipfelsequenzen trennen die fokussierten Einzelkomponenten und unterstreichen ihre Eigenständigkeit. Ein weiteres Ergebnis der Untersuchung ist, dass innerhalb einer prosodischen Phrase mit mehreren Satzakkenten häufig Konturen in Reihung vorkommen, die sich in ihrer Gestalt ähneln. So finden sich z.B. oft Reihungen früher Gipfel und auch Reihungen von Talkonturen. Diese 'Konturharmonie' innerhalb prosodischer Phrasen ist ein weiteres, kaum erforschtes Mittel, die melodische und semantische Einheitlichkeit prosodischer Phrasen zu erhöhen. Peters et al. weisen darauf hin, dass verschiedene melodische Konturen an Satzakkentpositionen unterschiedliche semantisch-pragmatische Konnotationen haben. So signalisiert der frühe Gipfel Abgeschlossenheit, während ein späterer Gipfel neue Information oder auch Überraschung vermitteln kann (Kohler

1991b). Die Aneinanderreihung mehrerer Konturen gleicher Gestalt kann also dazu dienen, den funktionalen Aspekt einer Phrase auf der pragmatischen Ebene mehrfach durch die Verwendung funktional äquivalenter F0-Muster zu signalisieren.

Ein anderer F0-bezogener Parameter ist die *Deklination* (Cohen & 't Hart 1967; Cruttenden 1986; 'tHart et al. 1990; Pompino-Marschall 1995). Deklination wird als ein zeitabhängiger Abwärtstrend von F0 im Verlauf einer Phrase (auch im Verlauf von Äußerungen) beschrieben. Hierbei ist auch das Absinken der oberen und unteren Grenzfrequenz zu beobachten, das zu einer Verringerung des *pitch range* führt.



**Abbildung 2:** Deklination: Schematischer Grundfrequenzverlauf innerhalb einer intonatorischen Phrase (Pompino-Marschall 1995:235).

Einige Untersuchungen beschreiben das Phänomen Deklination über eine statistische Abstraktion, indem über einer großen Anzahl von Äußerungen der zeitliche Verlauf der Grundfrequenz gemittelt wird. In Messungen dieser Art zeigt sich dann ein Abwärtstrend von F0 über der Zeit (Willems 1983). Levelt (1989) wendet allerdings ein, dass sich Deklination eher in Lesesprache als in Spontansprache beobachten lässt. Es gibt bis heute keinen Konsens darüber, ob Deklination ein für spontansprachliche Äußerungen relevantes Phänomen ist. Kohler (1991a) argumentiert gegen die Einbeziehung der zeitlich gesteuerten Deklination in die Modellierung spontaner Sprache und spricht sich für das strukturell bedingte Prinzip des *downstep* aus. Downstep bedeutet ein sukzessives Absinken der Höhe von Grundfrequenzgipfeln an Satzakkzentpositionen innerhalb prosodischer Phrasen. Dieses Absinken ist unabhängig von der Zeit zwischen den Satzakkzenten. Die Erniedrigung von Gipfel zu Gipfel liegt im Bereich von 10%.

Neben diesen kohäsiven melodischen Ausprägungen zeichnen sich prosodische Phrasen durch Kontinuität in Register, Sprechgeschwindigkeit und Stimmqualität aus, also genau das gegenteilige Prinzip zum Wechsel prosodischer Eigenschaften von einer Phrase zur nächsten.

Zusammenfassend ist zu sagen, dass die interne Struktur prosodische Phrasen und deren Delimitation durch eine Vielzahl phonetischer Faktoren bedingt ist und sich in spontansprachlichen Korpora eine starke Variationsbreite in der Ausprägung prosodischer Phrasen beobachten lässt. Ein Teil dieser Variation kann durch unterschiedliche Funktionen von Phrasen erklärt werden, womit sich der nächste Absatz befasst.

### **3.1.4 Form-Funktions-Beziehungen**

Ein seit langem verbreiteter Ansatz in der Analyse der sprachlichen Funktion und der phonetischen Form prosodischer Grenzen besteht darin, die Beziehung zwischen syntaktischen und prosodischen Phrasen zu untersuchen. Dieses Konzept spiegelt sich z.B. in Experimenten von Price et al. (1991) wider, die feststellen, dass das syntaktische Parsing oberflächlich ambiger Sätze von Sprechern durch prosodische Phrasierung kenntlich gemacht wird und bei Hörern zur syntaktischen Disambiguierung führt (siehe 3.3.1). In zahlreichen Untersuchungen konnte eine teilweise Übereinstimmung zwischen prosodischen Phrasen und syntaktischen Teilsätzen (Neben- oder Hauptsätzen) nachgewiesen werden. Es zeigt sich aber besonders in Spontansprache, dass syntaktische Strukturen und prosodische Phrasen nicht zwangsläufig koextensiv sind (Croft 1995). So finden sich gleichermaßen hohe syntaktische Grenzen ohne Merkmale prosodischer Phrasierung und auch prosodische Grenzen innerhalb syntaktischer Konstituenten.

Die vorliegende Arbeit verfolgt einen Ansatz, der sich dem Problem der Form-Funktions-Zuordnung von der semantisch-pragmatischen Seite nähert. Auch dieser Ansatz ist nicht neu, gewinnt aber durch die zunehmende Beschäftigung mit Spontansprache immer mehr an Bedeutung. So schlagen verschiedene Forscher vor, von einer Übereinstimmung zwischen intonatorischen Einheiten und 'Sinneinheiten' auszugehen (z.B. Kingdon 1958). Klinghard & Klemm (1920:67) definieren den *Sprechtakt* durch den *intonatorischen Sinntakt*:

*Was ist es dann aber, das auch unter solchen umständen den sprechtakt deutlich zu gehör bringt? Die antwort kann m. e. nur lauten: die tonbewegung, die intonation des schlusses eines jeden sprechtakts. Darum genügt zu dessen genauer bezeichnung nicht nur das wort 'sinntakt', sondern wir müssen vom intonatorischen sinntakte als der grundlage aller takteinteilung unserer rede sprechen.*

und weiter:

*Nachdem ich nun so, wie ich hoffe, mit ausreichender beweiskraft nachgewiesen habe, dass der sprechtakt ein intonatorischer sinntakt ist, erübrigt nur noch, dass ich eine förmliche definition des begriffs zu geben versuche. Und zwar möchte ich ihn dahin bestimmen, dass 'sprechtakt' ein unterteil des begriffs 'satz' ist und aus einer gruppe von worten besteht, die 1. durch ihren gemeinschaftlichen sinn enger miteinander verbunden sind als mit einer der beiden sie umgebenden wortgruppen, und die 2. mit der tonbewegung ihrer letzten silben andeuten, ob die rede fortgesetzt wird oder nicht. Ein sprechtakt kann auch aus einem einzigen worte bestehn.*

In diesen Zitaten sind die entscheidenden Ansätze enthalten, die in der aktuellen, pragmatisch orientierten Intonationsforschung eine Rolle spielen: die enge Bindung intonatorischer Strukturen an die semantische Kohäsion, die Bindung an die Gesprächssteuerung und die Orientierung der Form-Funktions-Beziehung auf den melodischen Verlauf am Phrasenende. Auch andere frühe Veröffentlichungen erfassen die elementare Verbindung zwischen der pragmatischen Gliederung von Gesprächen und prosodischen Strukturen (z.B. v. Essen 1964). Trotzdem blieben phonetische Forschungen lange weitgehend in der Untersuchung der Beziehung zwischen grammatischen und intonatorischen Strukturen verhaftet. Eine entscheidende Wendung brachte erst das Aufkommen der Pragmatik und insbesondere der ethnomethodologischen Konversationsanalyse, die

*eine funktionale Perspektive formuliert, die die intonatorische Gestaltung als ein zentrales sprachliches Mittel zur Strukturierung von Konversationen auffasst. Daher ist nach House (1990:43) die pragmatische Hauptfunktion der Intonation "to anchor an utterance to its context, enabling the hearer to make suitable inferences to enrich his interpretation". (Gilles 2005:20).*

Die Konversationsanalyse postuliert vor allem zwei Hauptfunktionen der prosodischen Strukturierung: die Gliederung eines Redebeitrags gemäß seiner Informationsstruktur und die Signalisierung über Weiterweisung vs. Abschluss innerhalb und am Ende von Redebeiträgen in Gesprächen. Gilles (2005:17):

*Durch die Phrasierung in IPs werden insbesondere längere Äußerungen prosodisch gegliedert, wodurch die Hörerin /der Hörer bei der Sprachverarbeitung kognitiv geleitet und entlastet wird. Nach Bannert/Schwitalla (1999:314) ist es in Gesprächen wichtig, "dass Sprecher und Hörer wissen, wann eine Sprechereinheit zu Ende ist. Inhalte von Äußerungen zu verstehen fällt um so leichter, je deutlicher der Sprecher Einschnitte beim Sprechen macht".*

Da die Konversationsanalyse das sprachlichen Verhalten von Sprechern immer mit Blick auf die Interpretation von Äußerungen im interaktiven Kontext des Gesprächs vornimmt, wird vor allem die weiterweisende vs. abschließende Funktion thematisiert. *Weiterweisung* wird hierbei als die Projektion einer geplanten Äußerungsfortsetzung definiert. Also der Signalisierung seitens des Sprechers, dass der aktuelle Redebeitrag noch nicht abgeschlossen ist und der Hörer das Wort noch nicht übernehmen soll. *Abschluss* bedeutet ein Zuendekommen, das dem Hörer erlaubt in die Sprecherrolle überzuwechseln. Gilles (2005:139):

*Weiterweisung kann prinzipiell immer angenommen werden, wenn keine Merkmale auf einen Abschluss hindeuten. Folglich sind weiterweisende Kontexte immer innerhalb eines Turns anzutreffen.*

*Auch für die Bestimmung dieser Funktion hat es sich als notwendig erwiesen, zwischen potenziell und eindeutig weiterweisenden Kontexten zu differenzieren. Eine Phrase ist eindeutig weiterweisend, wenn ihre syntaktische Konstruktion eine Fortsetzungserwartung aufbaut, die in einer Folgephrase eingelöst werden muss. Bei potenziell weiterweisenden Phrasen wird durch den Aktivitätstyp (z.B. Erzählung, längere Sachverhaltsdarstellung) eine globale pragmatische Projektion aufgebaut.*

Das Konzept von Abschluss vs. Weiterweisung wird vor allem wichtig, wenn es um die Differenzierung zwischen turninternen und turnfinalen prosodischen Grenzen geht (siehe Kapitel 4). In der vorliegenden Arbeit soll aber auch im Inneren von Redebeiträgen zwischen einer abschließenden und einer weiterweisenden Funktion differenziert werden, denn längere Redebeiträge können aus verschiedenen inhaltlichen Blöcken zusammengesetzt sein, wobei prosodische Grenzen im Inneren dieser Blöcke durch ihre phonetische Struktur stärkere inhaltliche Bindung an die Folgephrase signalisieren als dies an prosodischen Grenzen am Ende der inhaltlichen Blöcke der Fall ist. v. Essen (1964: 37):

*Solange die Hauptsache noch nicht gesagt, das sinnwichtigste Wort noch nicht gesprochen, der Schwerpunkt noch nicht gebildet ist, sind alle Redeteile 'progreident' und weisen weiter auf einen 'terminalen' Teil.*

Kohler (1995a: 196):

*Die fallende Intonation vermittelt im Gegensatz zur steigenden die Bedeutung, dass der Sprecher zu einem Abschluß gekommen ist, dass er weder zum selben Thema fortfahren wird noch seinen Gesprächspartner einlädt, dies zu tun. Deshalb kommt fallende Intonation insbesondere in (kategorischen) Behauptungen und Feststellungen, steigende in nicht abgeschlossenen, weiterführenden Äußerungen und Fragen vor.*

Couper-Kuhlen (1983:80) benutzt zur Beschreibung der Weiterweisung über prosodische Grenzen hinweg das Konzept der Kohäsion: Steigende Tonhöhenbewegung hat

*[...]kohäsive Kraft, weil durch sie Erwartungen geweckt werden, die nach späterer Auflösung verlangen. [...] Sie] kommt auch zwischen Toneinheiten vor, wobei eine auslaufende, steigende Bewegung ein Moment der musikalischen Spannung mit sich bringt.*

Aus den vorangehenden Zusammenfassungen und Zitaten geht bereits hervor, dass im Rahmen der phonetischen und funktionalen Analyse prosodischer Grenzen vor allem der melodische Verlauf von der letzten Satzakkzentposition bis zur Grenze thematisiert wird. Ein typisches Vorgehen ist die Zuordnung einer melodischen Bewegung zu einer sprachlichen bzw. konversationellen Funktion. Dieser Blickwinkel bedarf aber der Ergänzung durch zwei Aspekte. Erstens die Einbeziehung von Information, die nicht auf der phonetischen Ebene angesiedelt ist: Gilles (2000:22):

*Um diese einzelnen Funktionsbestandteile herauszuarbeiten, ist die Bezugnahme auf den weiteren konversationellen Kontext, in den ein Konturbeleg eingebettet ist, notwendig. Die Form der Kontur allein bzw. die Analyse einer isolierten Phrase kann nur bedingt zu einer erschöpfenden Funktionsbeschreibung führen. Erst eine kombinierte Analyse von Intonation, syntaktischer, sequentieller und interaktionaler Struktur ermöglicht die Zuweisung von Funktionen. Die intonatorische Struktur ist somit immer in einen Komplex von Kontextualisierungshinweisen eingebettet; sie besitzt keine Funktion an sich.*

Zweitens gilt es neben der Einbeziehung nicht-phonetischer Faktoren auch, die phonetische Analyse zu verfeinern und die reine Ausrichtung auf (oft stark abstrahierte) melodische Bewegungen zu erweitern. Wie bereits beschrieben, sind an prosodischen Grenzen eine Reihe phonetischer Veränderungen zu beobachten, die sich zu einem perzeptorisch relevanten Bündel akkumulieren. Die vorliegende Arbeit hat u.a. das Ziel,

über die Art der Bündelungen die funktional relevanten Interaktionen der akkumulierten Merkmale zu untersuchen. Da aber in der Literatur vor allem die Intonation bzw. andere einzelne Merkmale an prosodischen Grenzen thematisiert werden, folgt ein kurzer Überblick über Form-Funktions-Zuordnungen, der sich weniger auf Merkmalsbündel als auf Einzelmerkmale bezieht. Hierbei werden vor allem die Funktionen von Pausen, melodischen Mustern und segmentellen Dauerstrukturen thematisiert. Eine kurze Zusammenfassung zur Rolle phonatorischer Strukturen findet sich unter 3.2.11.

### 3.1.4.1 Melodische Muster

Das Zentrum der meisten intonatorischen Analysen bildet der Nukleus der Intonationsphrase (Silbenfolge beginnend mit dem letzten Satzakzent bis zum Phrasenende). So auch in der wohl am häufigsten zitierten Veröffentlichung zur Form-Funktions-Problematik im Deutschen (v. Essen, 1964). Hierzu Gilles (2005:19):

*v. Essen (1964:58ff.) nimmt drei 'Intonationstypen' (=nukleare Intonationskonturen) an:*

- **terminale Intonation** (in Aussagen, Aufforderungen, Ausrufen, vorangestellten Anreden, Ergänzungsfragen, indirekten Reden, den zweiten Gliedern von Doppelfragen)
- **progrediente Intonation** (bei unvollendeten Redeteilen, denen die Schwerpunktbildung erst folgen soll)
- **interrogative Intonation** (angewandt in Entscheidungsfragen, Nachfragen)

*Der terminale Intonationstyp wird fallend, der interrogative wird steigend ausgeführt. "Progrediente (weiterweisende) Aussprüche laufen mit gehobener Stimme aus, im übrigen unterscheiden sie sich nicht von den terminalen" (von Essen 1964:59). Dieses 'minimale'*

*System wird bei Pheby (1984) auf fünf und bei Kohler (1995a) auf sechs Tonmuster erweitert. An dieser unterschiedlichen Anzahl und auch aus den angeführten Verwendungskontexten, die von Essen für seine drei Tonmuster vorsieht, wird deutlich, dass die Autoren tatsächlich kein konsistentes Modell entwerfen, in dem die Intonation eindeutig abhängig von der Syntax ist. Wenn bei von Essen sowohl ein Aussagesatz als auch eine Ergänzungsfrage mit dem gleichen Intonationsmuster realisiert werden, dann demonstriert dies doch gerade die Unabhängigkeit der Intonation. Darüber hinaus handelt es sich bei von Essens Funktionskontexten nicht ausschließlich um syntaktische Kontexte. Vielmehr liegt hier eine Überschneidung von syntaktischen und pragmatisch-textorganisatorischen Merkmalen vor. Damit wird die Annahme der primär syntaktischen Funktion der Intonation ad absurdum geführt.*

Die von Gilles kritisierte starre Bindung an syntaktische Strukturen war und ist im *Kieler Intonationsmodell* nicht vorgesehen. So findet sich bei Kohler (1995a:196) ein Zitat, das eine große Nähe zum Standpunkt von Gilles erkennen lässt:

*Es kann allerdings nicht ausdrücklich genug darauf hingewiesen werden, dass kontrastive Intonationsstrukturen in einer Sprache keine konsistente, von allen Äußerungen losgelöste Bedeutung haben, wie dies oft in solchen Termini wie „Frageintonation“ impliziert ist. Segmenteller Aufbau und Intonationsmuster, morphologische und syntaktische Strukturen, Wortbedeutung und Situationskontext leisten alle ihren Beitrag zur Gesamtbedeutung eines sprachlichen Kommunikationsvorgangs und werden in ihrer Bedeutungsfunktion durch die anderen Komponenten wesentlich mitbestimmt.*

Auf dem heutigen Stand werden im *Kieler Intonationsmodell* 12 unterschiedliche phrasenfinale Melodiemuster differenziert, die als perzeptorisch distinktiv angesehen werden (siehe 2.1.3.4). Allerdings fehlt hierbei zu weiten Teilen eine Validierung der perzeptorischen Distinktivität dieser Muster sowie die Einordnung in funktionale Kontexte.

Als Lösung des Problems der fehlenden eindeutigen Zuordnung von Satzart und Intonationsverlauf schlagen Altmann et al. 1989 die abstrakte Zwischenebene *Satzmodus* vor und führen fünf semantische Funktionstypen ein: *Assertion*, *Interrogation*, *Wunsch*, *Aufforderung* und *Exklamativ*. Diese Modi können jeweils mit syntaktisch verschiedenen Satztypen realisiert werden (je nach Intonation). Aber auch mit dieser Einteilung kann über Perzeptionsexperimente keine durchgehende Form-Funktions-Zuordnung erreicht werden. Eine semantisch-pragmatische Orientierung findet sich wiederum bei Kohler (1995a:197). Er weist auf eine weitere Ebene hin, auf der intonatorische Bedeutung angesiedelt ist und eine Wechselwirkung zum syntaktisch orientierten Sprecherverhalten zeigt: die der Einstellung von Sprechern. Es zeigt sich,

*[...]dass die Frage keineswegs immer mit der steigenden „Frageintonation“ verbunden ist. So sind Sätze in der syntaktischen Form der Fragewortsätze mit fallender Intonation neutrale Bitten um Information ohne besonderen Ausdruck des Interesses und der gefühlsmäßigen Einstellung des Sprechers. Mit steigender Intonation wird Interesse an einer Antwort speziell signalisiert, das um so größer ist, je höher die Melodiebewegung endet.*

Dieser Ansatz, dessen funktionaler Hintergrund oft in den Bereich der *Paralinguistik* eingeordnet und somit aus linguistischen Systemüberlegungen ausgeschlossen wird, findet in neueren Untersuchungen wieder einen Anschluss. So wird in verschiedenen Experimenten z.B. von Grabe et al. (1997) und Rietveld et al. (2002) Fragestellungen nachgegangen, die Zusammenhänge zwischen tonalen Eigenschaften von Sprache und Kategorien wie *Höflichkeit*, *Überraschung* oder *Unsicherheit* untersuchen.

Zusammenfassend soll noch einmal folgende Frage gestellt werden: *Welche tonalen Muster an prosodischen Grenzen kodieren welche Bedeutung?* Die Antwort allerdings muss aufgrund der komplexen Interaktionen zwischen den verschiedenen Ebenen der sprachlichen Kommunikation offen bleiben, bzw. die Frage muss neu formuliert werden: *Welchen funktionalen Beitrag zur kommunikativen Bedeutung einer Äußerung liefern die tonalen Muster am Phrasenende?* Wird die Frage so formuliert, finden sich in fast allen Veröffentlichungen Hinweise auf eine grobe funktionale Bindung: Tief fallende Konturen signalisieren einen Abschluss, und das Kontinuum von leicht fallend über eben bis hin zu steigenden Konturen signalisiert Weiterweisung bzw. Unabgeschlossenheit. Im Prinzip kann jedes Muster in einem entsprechenden Gesprächskontext mit einer Frage an den Gesprächspartner verbunden sein. Féry (1993) differenziert zwischen steigenden Konturen in Fragen als *other directed* und steigenden Konturen bei turninterner Weiterweisung als *self directed*. Sowohl in Fragen als auch bei Aussagen signalisiert die Intonation Bedeutungsnuancen in der Einstellung des Sprechers, wobei ein Kontinuum zwischen Sicherheit, Kategorialität und z.T. auch Aggressivität bei tief fallenden Konturen und Unsicherheit, Höflichkeit und Unterordnung in Richtung der steigenden Konturen zu beobachten ist.

Der Frage nach der sprachgeschichtlichen Entstehung dieses in vielen Sprachen vorhandenen Gegensatzes zwischen fallenden und nicht fallenden Mustern erklärt Ohala (1983, 1984) aus einer biologischen Perspektive: Bei Vokalisationen von Säugetieren lässt sich beobachten, dass dominante oder aggressive Individuen Vokalisationen mit einer tieferen Grundfrequenz produzieren als submissive Vertreter derselben Spezies. Dies gilt vor allem in Streit- oder Kampfsituationen. Der Hintergrund hierbei ist, dass die Höhe der Grundfrequenz mit der Größe des Larynx korreliert und die Kehlkopfgröße wiederum in Abhängigkeit von der Gesamtkörpergröße steht. Tiefe Vokalisationen signalisieren also körperliche Größe, Stärke, Macht und somit Führungsanspruch im sozialen Gefüge. Ohala untermauert das Prinzip des *frequency code* weiterhin dadurch, dass beim Menschen das Verhältnis der Kehlkopfgröße zwischen Mann und Frau bei 2:1 liegt und somit dem Mann im Verhältnis zum Körpermasseverhältnis eine überproportional tiefe Grundfrequenz ermöglicht. Weiterhin senkt sich beim Mann während der Pubertät der Kehlkopf, was die Resonanzfrequenzen des Ansatzrohrs senkt und so zur Tiefe von Vokalisationen beiträgt. Ohala vergleicht dieses biologische Prinzip akustischer Natur mit dem Aufrichten von

Fell oder Federkleid und der männlichen Gesichtsbehaarung. Diese Faktoren lassen den Körper über den visuellen Kanal größer und furchteinflößender erscheinen. Aus diesen und anderen biologischen Beobachtungen werden zwei zentrale Folgerungen für menschliches Sprachverhalten gezogen. Erstens kann so erklärt werden, dass bestimmte emotionale und soziale Einstellungen eines Sprechers, die submissive Züge tragen, zu einem globalen Anstieg der Sprechstimme führen. So z.B. bei Freundlichkeit und Höflichkeit, während Dominanz und Aggressivität mit tiefem F0 einhergehen. Zweites ergibt sich eine Erklärung für unterschiedliche Funktionen melodischer Muster. Bereits 1964 kann Uldall in Perzeptionsexperimenten unter Verwendung des semantischen Differentials eine Verbindung zwischen hoch endenden Äußerungen und den Einstellungen 'submissiveness' und 'pleasantness' belegen. Neuere Untersuchungen z.B. zum Niederländischen zeigen, dass vor allem der F0-Verlauf am Ende von Äußerungen bei der Signalisierung der Sprechereinstellung relevant ist (Rietveld et al. 2002). Hierauf aufbauend ergibt sich eine Erklärung der Form-Funktions-Beziehung fallender und steigender melodischer Muster. Fallende Muster haben in vielen Sprachen eine abschließende Funktion und signalisieren Sicherheit, evtl. sogar Kategorialität einer Aussage und haben somit einen dominanten Charakter. Steigende Konturen signalisieren oft Fragen und implizieren die Bitte um Information, setzen den Sprecher also eher in eine submissive Position. Der häufigste funktionale Zusammenhang, in dem final steigende oder hoch-eben endende Konturen zu beobachten sind, ist im Deutschen und wohl auch in vielen anderen Sprachen aber die Weiterweisung innerhalb eines Redebeitrags. Hier könnten diese Konturen im Rahmen des *frequency code* als eine Art Bitte um das weitere Rederecht interpretiert werden.

#### **3.1.4.2 Pausen**

Der funktionale Status von Pausen in phonetischen und phonologischen Analysen ist ein anderer als der von Intonationsmustern. Ihr Einfluss auf die semantische und pragmatische Struktur wird als wesentlich geringer eingestuft als der von Intonationsmustern und folglich spielen Pausen in der Beschreibungen von Form-Funktions-Zusammenhängen eine kleinere Rolle als die Intonation. Für die geringere funktionale Relevanz gibt es eine einfache Erklärung, die in der phonetischen Flexibilität des Parameters Pause liegt: Eine

Pause im Signal ist entweder vorhanden oder nicht, und wenn sie vorhanden ist, kann sie lediglich in ihrer Dauer variieren. Das bedeutet, dass die Möglichkeit der Bildung perzeptorisch relevanter Kontraste durch diesen Parameter viel geringer ist als bei einem so flexiblen Parameter wie der Tonhöhe. Trotzdem wurden eine Vielzahl von Untersuchungen zur Wahrnehmung und Funktion von Pausen durchgeführt. Der folgende Abschnitt konzentriert sich auf Beschreibungen der funktionalen Relevanz von Pausen. Literaturzusammenfassungen und experimentelle Untersuchungen zur Perzeption von Pausen sind in Abschnitt 3.3.1 dargestellt.

In den untersuchten Korpusdaten ist zu beobachten, dass in einem großen Anteil von Pausen eingeatmet wird. Da das Einatmen z.T. fast unhörbar geschieht, ist eine Trennung zwischen Sprechpausen und Atmungsphasen anhand akustischer Daten oft kaum möglich. Deshalb stellen die folgenden Betrachtungen das Phänomen der Pausen und das der Atmungsphasen in enge Verbindung zueinander und führen keine strikte Trennung zwischen Pausen und Atmungsphasen ein.

Sprechpausen und Atmungsphasen verursachen perzeptiv auffällige Zäsuren (siehe 3.3.2). Sie repräsentieren eine extreme Art phonetischer Separierung, da für eine bestimmte Zeitspanne kein Sprachsignal produziert wird. Die sprachentwicklungsgeschichtliche Herkunft von Pausen innerhalb längerer Äußerungen wird oft auf die Notwendigkeit des Nachfüllens der Lungenluft zurückgeführt, um weitere Sprachlaute produzieren zu können (Jonson 1640; Sweet 1877, 1890; Jones 1960). Das Vorkommen von Pausen ist universell, da ihnen diese physiologische Notwendigkeit zugrunde liegt. Sprecher müssen also in längeren Äußerungen Pausen machen, um zu atmen (allerdings wird nicht immer geatmet, wenn eine Pause gemacht wird).

In der synchronen Analyse von Sprechpausen müssen neben diesem physiologischen Zwang zwei Aspekte der Entstehung und Funktion von Pausen/Atmen berücksichtigt werden, um die Häufigkeit, die Dauer und die Positionen von Pausen in Äußerungen zu erklären:

- An welchen Positionen in der Äußerung werden Pausen gemacht?
- Welche kognitiven Prozesse auf der Sprecherseite stehen in Verbindung mit Pausen?

Viele Untersuchungen heben die Funktion von Pausen innerhalb des semantisch determinierten Gliederungsprozesses hervor. So z.B. Jones (1960:254):

*Pauses are continually being made in speaking. They are made (1) for the purpose of taking breath (2) for the purpose of making the meaning of the words clearer [...] Pauses for breath are normally made at points where pauses are necessary or allowable from the point of view of meaning. Sentences are usually divisible into smaller groups between which pauses may be made, though they are not essential. The shortest of such groups (i.e. groups which are not capable of being further subdivided by pauses) are called „sense groups“.*

Beobachtungen dieser Art sind in der Literatur häufig und reichen weit in die Geschichte sprachlicher Forschung zurück (Jonson 1640; Webster 1789; Sweet 1877). Generell wird von einer Übereinstimmung zwischen vorhandenen Grenzen in Äußerungen und der Position von Pausen ausgegangen. In einigen Beschreibungen wird hierbei eher die Verbindung mit syntaktischen Grenzen (auch mit der Interpunktion) gesucht, während andere Studien sich eher auf inhaltliche Gruppierungen beziehen (deren Grenzen allerdings meist auch mit syntaktischen Grenzen zusammenfallen). Lehiste (1979) sieht eine Funktion kürzerer Pausen in der Signalisierung von Satzgrenzen, während längere Pausen Texte in Absätze untergliedern. Sie weist in Experimenten mit delexikalisierter Lesesprache nach, dass Versuchspersonen bei Pausen zwischen 1000 und 2000 ms angeben, dass sich an der Pausenposition eine Satzgrenze befindet und bei Pausen über 2000 ms eine Absatzgrenze. Diese Ergebnisse zeigen, dass auch hinsichtlich der Signalfunktion eine Beziehung zwischen inhaltlicher und phonetischer Separierung vorliegt. Lehiste nimmt in ihrer Untersuchung allerdings keinen Bezug zu den melodischen Mustern an den Pausenpositionen und vernachlässigt so den funktionalen Beitrag der Intonation in der Interpretation des Stimulusmaterials durch die Hörer.

Die beschriebenen Beobachtungen sind auf der Analyse von gelesenen Texten gegründet. Erst in den letzten 30 bis 40 Jahren ist die spontan von Sprechern erzeugte Sprache in den Blickpunkt der Forschung gerückt. Dies hat zu entscheidenden

Veränderungen in der Interpretation von Sprechpausen geführt. Megyesi & Gustafson-Capkova (2002) beobachten in professioneller schwedischer Lesesprache eine hohe Koinzidenz von Pausen und hohen inhaltlichen und syntaktischen Grenzen, während Pausen in schwedischen *Map Task*-Dialogen eher im Inneren thematischer Einheiten und an sämtlichen syntaktischen Strukturstellen auftreten.

Eine Analyse der Terminabsprachen des *Kiel Corpus of Spontaneous Speech Vol. I und Vol. II* (IPDS 1995, 1996) belegt diese Divergenz zwischen syntaktischer Phrasierung und der Position von Sprechpausen auch für das Deutsche. Da für die untersuchten Daten keine syntaktische Analyse vorliegt, orientiert sich die Untersuchung an den Satzzeichen der transliterierten Dialoge. Dieses Vorgehen ist sinnvoll, da die Interpunktion im Deutschen viele der vorhandenen hohen syntaktischen Grenzen markiert. Es wird über Datenbankabfragen festgestellt, an wie vielen der Satzzeichen tatsächlich Sprechpausen vorkommen bzw. wie viele der messbaren Sprechpausen mit einem Satzzeichen einhergehen. In den zweieinhalb Stunden spontaner Sprache, die im Rahmen von Terminabsprachen aufgenommen wurden gibt es im Inneren der Redebeiträge 4051 Satzzeichen und 3774 Pausen. 35,2% der Positionen von Interpunktionszeichen fallen mit Pausen zusammen. 37,8% der Pausen befinden sich an der Position eines Interpunktionszeichens. Das bedeutet, dass etwa zwei Drittel der syntaktischen Grenzen, die ein Interpunktionszeichen tragen, keine Pause aufweisen und zwei Drittel der Pausen an Positionen ohne Interpunktion vorkommen. Sicher kommt der Bezug auf die Interpunktion nicht einer syntaktischen Analyse gleich, denn nicht jede syntaktische Konstituente wird durch Interpunktionszeichen abgegrenzt. Trotzdem zeigen diese Zahlen, dass ein hoher Anteil der Pausen nicht aus syntaktischen Strukturen abgeleitet werden kann. Diese große Zahl nicht syntaktisch motivierter Pausen wird erklärbar, wenn Pausen auch als externe Reflektionen von kognitiven Prozessen der Sprachproduktion angesehen werden und nicht nur als inhaltlich oder syntaktisch motivierte Gliederungssignale (Goldman-Eisler 1968, 1972). Pausen entstehen also häufig an Positionen der Äußerung, an denen der weitere Verlauf der Rede kognitiv und motorisch vorbereitet wird. Dies schließt die Planung lexikalischer und syntaktischer Strukturen und des inhaltlichen Gesamtverlaufs ebenso ein wie die Vorbereitung artikulatorischer Prozesse.

Diese grundsätzliche Unterscheidung findet in ältere Literatur keinen Eingang, da in der dort analysierten gelesenen Sprache, abgesehen von der motorischen Planung, keine

Sprachplanung stattfindet, da lediglich ein bereits bestehender Text abgelesen wird und die Sprecher hierbei tatsächlich nur die inhaltliche und syntaktische Struktur in Einklang mit eventuellen Sprechpausen bringen müssen. Folglich gibt es in Lesesprache wesentlich weniger Pausen kognitiven Ursprungs als in spontaner Sprache (Goldman-Eisler 1968, 1972). Grosjean & Deschamps (1975) untermauern diese Sichtweise durch Befunde, die zeigen dass Sprecher bei komplexeren sprachlichen Aufgaben mehr und längere Pausen machen als bei einfacheren Aufgaben. Sie vergleichen hierfür das sprachliche Verhalten beim Beschreiben von Cartoons im Gegensatz zur Beantwortung einfacher Interviewfragen. Grosjean & Dommergues (1983) weisen darauf hin, dass die psycholinguistischen Einheiten, in denen Sprache geplant wird, nicht homomorph mit syntaktischen Strukturen sind, die Syntax also für Pausen kognitiven Ursprungs kein geeigneter Prädiktor ist. Weiterhin ist anzunehmen, dass Dauern und Positionen von Pausen sehr stark sprecherspezifischen Schwankungen unterliegen, da es sowohl in der Atmungstechnik und respiratorischen Kapazität als auch im Ablauf und der Geschwindigkeit des Sprachplanungsprozesses individuelle Unterschiede gibt.

Pausen als Spiegel kognitiver Aktivität sowie inhaltlich/syntaktisch gliedernde Pausen können als ein großer Beitrag zum Verstehen von Sprache angesehen werden, sofern Hörer die Herkunft bzw. Funktion einer Pause richtig einordnen. Da aber eine Signaltypische Pause allein, wie bereits beschrieben, ein relativ unflexibler Parameter ist, bedarf sie in ihrer funktionalen Einordnung durch einen Hörer der Ergänzung durch andere prosodische Merkmale, wie z.B. der Intonation oder der Variationen segmenteller Dauerstrukturen. So beschreiben Shriberg et al. (2000), dass im Englischen dem Beginn einer neuen thematischen Einheit (*new topic*) oft eine Pause, begleitet von einem tiefen Grenztönen vorangeht und zum Beginn der neuen thematischen Einheit ein F0-Reset folgt.

Eigene Untersuchungen zeigen, dass Pausen, die als 'Planungspausen' eingestuft werden, oft eine sehr starke Längung, begleitet von einem ebenen melodischen Verlauf vorangeht. Bei starken Planungsschwierigkeiten kann der gesamte Sprechfluss ins Stocken kommen und sich die Vorkommenshäufigkeit von Pausen stark erhöhen. Bei vielen Sprechern ist dann das Phänomen der *gefüllten Pausen* zu beobachten: die Pausen werden mit Vokalisationen wie *ähm* gefüllt. Dieses Verhalten zeigt dem Hörer klar an, dass die entstandene Unterbrechung mit Planungsschwierigkeiten verbunden ist und dass der Sprecher sich nach Beendigung der Planung weiter äußern wird. Dieses und

verwandte Phänomene zu prosodischen Grenzen in unflüssiger Sprache werden im Kapitel 5 weiter ausgeführt.

### 3.1.4.3 Segmentelle Dauerstrukturen

Segmentelle Dauerrelationen sind ein wichtiger Träger sprachlicher Information. Klatt (1976) führt folgende Bereiche an, in denen im Englischen die Dauerstruktur zur Differenzierbarkeit sprachlicher Strukturen beiträgt:

- Im segmentellen Bereich werden Lang- vs. Kurzvokale (auch als *gespannte* vs. *ungespannte* Vokale bezeichnet, z.B. Kohler 1995a) und stimmhafte vs. stimmlose Frikative u.a. durch Dauerunterschiede kodiert.
- Auf der Äußerungsebene zeigen satzakzentuierte Silben eine größere Dauer als unakzentuierte Silben.
- Emotionale Einflüsse (z.B. Emphase) können phonetisch in Dauerstrukturen kodiert sein (Kohler 2005a).
- Und natürlich wird auch die Zunahme von Dauern am Ende prosodischer Phrasen beschrieben. Dieses Phänomen wird oft als *phrasenfinale Längung* bezeichnet.

Dauerstrukturen sind somit Informationsträger auf ganz unterschiedlichen Ebenen, angefangen bei der Kodierung von Segmenten, über die Prominenz von Silben im prosodischen Syntagma, bis hin zu Beziehungen zur emotionalen Verfassung von Sprechern. Hier soll vor allem die Ausprägung und die Funktion der Dauer von Segmenten an prosodischen Grenzen kurz zusammengefasst werden.

Eine Reihe von Untersuchungen zeigt, dass die Dauer von Segmenten an prosodischen Grenzen zunimmt (Klatt 1975; Price et al. 1991; Wightman et al. 1992). Als Domäne dieser finalen Längung wird von den genannten Autoren der letzte Reim vor der Grenze angesetzt. Peters, Kohler & Wesener (2005b) belegen für über 2000 prosodische Grenzen in spontaner Sprache einen Anteil von ca. 70% deutlich gelängter Segmentdauern in phrasenfinalen Reimen. Einige Autoren beziehen sich bei ihren Analysen nicht direkt auf das Konzept der prosodischen Grenzen bzw. der prosodischen Phrasen, sondern messen die segmentelle Dauerstruktur vor Pausen und stellen segmentelle Längung vor

Pausen fest (*prepausal lengthening*) (Oller 1973; Klatt 1975). Die gemessenen Silben vor Pausen sind zwischen 60 ms und 200 ms länger als vergleichbare Silben in anderen Positionen. Ein Zitat aus Klatt (1976:1211) zeigt deutlich die Fixierung auf das Konzept des *prepausal lengthening* und die angenommene Bindung der Längung an nachfolgende Pausen:

*Prepausal lengthening also seems to occur at phrase and clause boundaries when there is no physical pause present in the acoustic signal.*

Andere Untersuchungen beschreiben eine globale Verlangsamung der Sprechgeschwindigkeit an prosodischen Grenzen. Lehiste (1979) geht von einem solchen Faktor am Ende von Absätzen aus, kann allerdings die perzeptorische Relevanz dieser Längung nicht belegen. Local & Kelly (1986) und Local, Kelly & Wells (1986) beobachten in ihren gesprächsanalytischen Studien vor allem am Ende von Redebeiträgen eine globale Verlangsamung.

Abgesehen von den konversationsanalytischen Untersuchungen sind, ähnlich wie bei der Analyse von Pausen und ihrer Funktion, die meisten Arbeiten mit der Beziehung zwischen Längung und syntaktischen Strukturen befasst. Typisch hierbei ist der Nachweis von Längung am Ende hoher syntaktischer Grenzen (Martin 1970; Klatt & Cooper 1975; Klatt 1975). Klatt (1975) zeigt, dass Segmente, die unmittelbar am Ende syntaktischer Phrasen stehen, etwa 30 ms länger sind als an anderen Positionen. Diese Ergebnisse werden auf die perzeptorische Relevanz des Parameters projiziert, indem angenommen wird, dass Hörer größere Dauern am Ende von Sätzen erwarten und diese Dauern den Hörern als Hilfe bei der Auflösung der syntaktischen Konstituentenstruktur dienen (Lehiste 1979; Klatt & Cooper 1975). Ein überzeugender Nachweis hierfür bleibt allerdings aus. Lehistes Experimente von 1979 mit delexikalisierte Lesesprache können nicht belegen, dass die Hörer bei der Erkennung von Satzgrenzen von der Dauerstruktur geleitet werden. Sie scheinen sich für ihre Urteile vielmehr an der Position von Pausen zu orientieren. Wie schon bei der Analyse der Pausen gelingt es also nicht, eine klare Beziehung zwischen Syntax und segmenteller Dauerstruktur zu etablieren. Und wie bei den Pausen übergehen die Autoren vieler Veröffentlichungen auch hier die Bedeutung der

Dauerstruktur als Reflexion kognitiver Prozesse bei der Sprachplanung. Die Gründe hierfür sind wiederum das Denken in linguistischen Kategorien und die fehlende Ausrichtung der Analysen auf die konversationelle Bedeutung phonetischer Strukturen in spontaner Sprache.

Als Beleg für die Notwendigkeit eines neuen Forschungsparadigmas, das die kognitive und interaktionale Bedeutung stärker hervorhebt, kann eine einfache Statistik über das in dieser Arbeit untersuchte Korpus angeführt werden. Im Sprachmaterial des *Kiel Corpus of Spontaneous Speech Vol. I und II* sind Längungserscheinungen, die als "Zögerungslängung" wahrgenommen wurden durch die Etikettierer gekennzeichnet worden. In den zweieinhalb Stunden Sprachmaterial gibt es 2010 Fälle, in denen perzeptiv auffällige Längung vorkommt, die als Zögerungslängung interpretiert wurden. Das bedeutet, dass durchschnittlich alle viereinhalb Sekunden eine etikettierte Längung dieser Art auftritt. Kontextbezogene Analysen dieser Fälle zeigen klar, dass diese Form z.T. extremer Längung an Positionen in der Äußerung vorkommt, an denen Sprecher überlegen, wie sie ihre Rede weiter fortsetzen wollen, auch innerhalb syntaktischer und inhaltlicher Einheiten. Es handelt sich hierbei also um ein vergleichbares Phänomen wie das der Pausen kognitiven Ursprungs, die dem Zeitgewinn im Sprachplanungsprozess dienen und gleichzeitig dem Hörer signalisieren, dass die Äußerung noch nicht abgeschlossen ist. Es gilt also auch hier, Längungserscheinungen, die die syntaktische und inhaltliche Struktur einer wohlgeformten Äußerung signalisieren, von denen zu trennen, die von Planungsprozessen herrühren. Eine experimentelle Untersuchung hierzu findet sich unter 5.2.

Im Bereich der intendierten prosodischen Gliederung in Übereinstimmung mit Textstruktur und syntaktischer Phrasierung scheint der Parameter Längung für Hörer keine starke perzeptive Relevanz bei der Identifizierung von prosodischer Grenzen zu haben (siehe 3.3.2), ist aber in den untersuchten Sprachdaten sehr häufig (siehe 3.2.9). Es kann also die Frage nach der Ursache für das gehäufte Vorkommen der Längung am Ende prosodischer Phrasen gestellt werden. Ist die finale Längung ein erlerntes syntax-bezogenes sprachspezifisches Verhalten, das dem Hörer bei der Dekodierung der sprachlichen Botschaft hilft, oder handelt es sich um eine produktionsbezogene generelle Tendenz, am Ende motorischer Einheiten eine Verlangsamung der motorischen Aktivität zu haben (*speech construction constraints*)?

Lindblom (1978) widmet sich dieser Frage und führt aus, dass finale Längung (*increased relative duration*) nicht nur in spontaner Sprache, sondern als Katalexis auch in Lyrik und Musik (*fermate, ritardando*) zu beobachten ist. Er kommt allerdings nicht zu einer klaren Zuordnung, da sich für beide Erklärungsansätze Belege finden lassen. So variiert der Grad der finalen Längung in verschiedenen Sprachen, was für den sprachspezifischen Hörerbezogenen Ansatz spricht. Auch lässt sich in vorsprachlichen Äußerungen von Kindern und in der Sprache von Gehörlosen ein schwächerer Grad finaler Längung, bis hin zu deren Fehlen beobachten, was ebenfalls gegen eine produktionsbezogene Erklärung spricht. Auf der anderen Seite kommt finale Längung laut Lindblom nicht nur vor Pausen, sondern auch an syntaktischen Grenzen vor (Lindblom bezieht sich hier offensichtlich auf prosodische Grenzen ohne Pausen). Er interpretiert die finale Längung in solchen Fällen als Hörerorientierte Einschränkung von Anpassungs- und Verschmelzungsprozessen über die prosodische Grenze hinweg, mit dem Ziel, die Konstituentenstruktur anzuzeigen. Diese Sichtweise stützt den grammatischen Erklärungsansatz. Lindblom kommt zu dem Schluss, dass sich hier evtl. mehrere Erklärungen des Phänomens überlagern.

Die folgenden Abschnitte dieses Kapitels beschreiben Korpusanalysen und Perzeptionsexperimente, die im Rahmen dieser Arbeit zu flüssigen turninternen prosodischen Grenzen durchgeführt wurden.

## 3.2 Symbolbasierte Analyse der Produktionsdaten

### 3.2.1 Methoden der automatischen Merkmalsextraktion

Dieser Abschnitt befasst sich mit der Verteilung der phonetischen Signale prosodischer Grenzen im Korpus und gibt anhand von Häufigkeitsverteilungen einen Überblick über Vorkommen und Ausprägung der Merkmale Pause/Atmen, melodischer Verlauf, segmentelle Längung, F0-Reset und Phonation an prosodischen Grenzen. Zuerst wird ein Überblick über Häufigkeit und Ausprägung der einzelnen Merkmale gegeben, anschließend über ihre Bündelungen. Es werden sowohl die Einzelmerkmale als auch deren Bündelungen dargestellt, da in der Literatur häufig Bezug auf einzelne phonetische Parameter genommen wird und diese Arbeit auch einen Bezug zu diesen Publikationen herstellen soll. Darüber hinaus ist eine vorangehende separate Darstellung der Einzelmerkmale und bestimmter funktionaler Aspekte isolierter phonetischer Strukturen übersichtlicher und kann z.T. den Zusammenhang zwischen phonetischer Form und sprachlicher Funktion klarer herstellen als die Analyse der Bündelungen, da sich zeigt, dass in einigen funktionalen Bereichen einzelne Merkmale entscheidend für die funktionale Struktur verantwortlich sind und die begleitenden anderen Merkmale keinen wesentlichen Einfluss auf die Funktion haben.

Die Etikettierung des untersuchten Datenkorpus enthält Informationen über artikulatorische und temporale Strukturen der segmentellen Ebene, über suprasegmentelle melodische Muster und über die Position und die Dauer von Pausen und Atmungsphasen. Im Rahmen einer zusätzlichen Etikettierung wurde außerdem das Vorhandensein von laryngalisierten Signalabschnitten im Bereich prosodischer Grenzen mit terminaler Intonation markiert und zeitlich delimitiert (nur in den *Video Task* Daten). Alle diese phonetischen Parameter spielen bei der Signalisierung prosodischer Grenzen eine Rolle und werden deshalb für jede etikettierte prosodische Grenze automatisch abgefragt. So entsteht für jede Grenze ein symbolisch repräsentiertes Merkmalsbündel. Diese symbolischen Repräsentationen phonetischer Strukturen bilden die Basis für statistische Aussagen über die Form prosodischer Grenzen im Korpus einerseits und sind andererseits der Ausgangspunkt für die Überführung der im Korpus beobachteten phonetischen Variation in Stimulusmaterial für Perzeptionsexperimente.

### 3.2.1.1 Symbolbasierte Extraktion suprasegmenteller Merkmale

Die folgenden Abschnitte erläutern die verschiedenen Arbeitsschritte, in denen die Merkmale extrahiert und in ein komplexes gebündeltes Etikett pro prosodische Grenze integriert werden. Die beschriebene Verarbeitung der Etikettierung bezieht sich nur auf etikettierte prosodische Grenzen im Inneren von Redebeiträgen, die weder Häsitationslängung noch Häsitationspartikeln oder syntaktische Abbrüche aufweisen. Außerdem werden prosodische Grenzen ausgeschlossen, die in Signalabschnitten mit überlagerndem Sprechen beider Sprecher vorkommen. Diese Kriterien führen zu einer Grundgesamtheit von 3399 flüssigen turninternen prosodischen Grenzen für diesen Untersuchungsbereich (LINDENSTRASSE und Terminabsprachen). Für einige Fragestellungen werden allerdings nur Teile des Gesamtkorpus in die Analyse einbezogen (siehe 3.2.2)

Die prosodische Etikettierung der Datenbasis bietet direkten symbolbasierten Zugriff auf die phonetischen Merkmale F0-Reset, phrasenfinale Intonationskontur und Satzakkentuierung. Für jede einzelne prosodische Grenze wird aus der manuellen Etikettierung automatisch ein Label abgeleitet, das die phrasenfinale Intonationskontur, das Vorhandensein bzw. Fehlen eines F0-Resets sowie die Stärke der Satzakkentuierung des phrasenfinalen Reims erfasst. Die Information wird extrahiert, in *7-Bit ASCII* kodiert und in einer abgeleiteten Etikettierdatei an das Label angehängt, dass die prosodische Grenze markiert.

Auf diese Weise entstehen Etiketten wie z.B. `<=Pgin_#2._A2>`. Das Gleichheitszeichen kodiert hierbei das Auftreten eines F0-Resets nach der prosodischen Grenze und `<#2.>` einen tief fallenden melodischen Verlauf am Ende der prosodischen Phrase. Die Sequenz `<A2>` symbolisiert, dass der phrasenfinale Reim Träger eines Satzakkents der Akzentstufe 2 ist. Alle weiteren Merkmale werden jeweils mit dem Trennzeichen `<_>` an dieses Etikett angehängt.

### **3.2.1.2 Symbolbasierte und numerische Extraktion von Pausen, Atmen und phonatorischen Veränderungen**

Pausen und Atmungsphasen sind in der segmentellen (nicht in der prosodischen) Etikettierung symbolisch markiert und zeitlich delimitiert. Das gleiche gilt für Laryngalisierung im Zusammenhang mit terminalen Intonationskonturen (siehe 3.2.11). Die Integration dieser Merkmale in das Merkmalsetikett verläuft deshalb in zwei Schritten. Im ersten Schritt wird die symbolisch repräsentierte Information in das Merkmalsetikett übernommen und in einem zweiten Schritt wird die Dauer der Pausen und Atmungsphasen bzw. der laryngalisierten Signalabschnitte gemessen und als numerischer Wert in das Merkmalsetikett eingefügt. So entstehen Etiketten wie das folgende: `<=Pgin_#2._A2_L150_P310>`. Eine solche Symbolisierung repräsentiert einen laryngalisierten phrasenfinalen Signalabschnitt von 150 ms, dem eine Pause von 310 ms folgt.

### **3.2.1.3 Numerische Extraktion der Längungsinformation**

Für die Erfassung der finalen Längung müssen phrasenfinale mit phraseninternen Segmentdauern verglichen werden. Die Segmentdauern lassen sich aus den Etikettierdateien extrahieren. Um den Vergleich durchführen zu können, muss zunächst der Bereich am Ende der Phrase ausgewählt werden, in dem die Längung bestimmt werden soll. Messungen von Wightman et al. (1992) weisen darauf hin, dass der Reim der phrasenfinalen Silbe und nicht der Takt oder die letzte betonte Silbe die Domäne ist, in der sich die phrasenfinale Längung hauptsächlich manifestiert. Deshalb wird der letzte Reim vor der prosodischen Grenze für die Untersuchung ausgewählt. Für den Vergleich mit Segmentdauern im Phraseninneren werden die phraseninternen wortfinalen Reime herangezogen.

Ziel der numerischen Verarbeitung ist ein von Sprecher und Segmenttyp unabhängiger Vergleich der Lautdauer in den beiden Phrasenpositionen. Hierzu muss eine Normalisierung durchgeführt werden. In Anlehnung an Price et al. (1991) wird folgendes Normalisierungsverfahren gewählt:

- Für jeden Sprecher und jeden Lauttyp werden in *phraseninternen* wortfinalen Reimen Mittelwert und Standardabweichung der Lautdauer berechnet. In die Berechnung gehen nur reguläre Wörter ein, in denen keine Zögerungslängung auftritt.
- Jede Lautdauer im *phrasenfinalen* Reim wird mit Bezug auf die so gewonnenen laut- und sprecherspezifischen Mittelwerte und Standardabweichungen nach folgender Formel normalisiert:

$$d^{\sim} = (d - \mu_{\alpha}) / \sigma_{\alpha}$$

mit

$d$  = Dauer des jeweiligen Segments im phrasenfinalen Reim,

$\mu_{\alpha}$  = sprecherspezifischer Mittelwert des jeweiligen Lauttyps in phraseninternen Reimen,

$\sigma_{\alpha}$  = zugehörige Standardabweichung.

- Der resultierende Wert  $d^{\sim}$  erfasst die relative Dauer eines Segments als Anzahl der Standardabweichungen bezogen auf die mittlere Segmentdauer. Ist die Segmentdauer  $d$  genauso groß wie der Mittelwert  $\mu_{\alpha}$ , so ist die normalisierte Dauer  $d^{\sim} = 0$ . Ist die Segmentdauer größer (kleiner) als der Mittelwert, so ist die normalisierte Dauer positiv (negativ). Für Segmentdauern innerhalb der Standardabweichung ergeben sich Werte von -1 bis 1.
- Akzentuierung ist ebenfalls mit Längung verbunden. Um diesen Faktor zu kontrollieren, werden Mittelwertbildung und Normalisierung für drei in der prosodischen Etikettierung symbolisierte Akzentstufen im wortfinalen Reim getrennt durchgeführt: unbetont, schwach akzentuiert (Akzentstufe 1), akzentuiert (Akzentstufe 2). Die ebenfalls markierte emphatische Akzentuierung (Akzentstufe 3) wird aufgrund des

seltenen Vorkommens von der Analyse ausgeschlossen. Die Kategorie 'unbetont' umfasst neben Reimen in nicht satzakzentuierten Wörtern auch lexikalisch unbetonte Reime in satzakzentuierten Wörtern.

Um für jeden phrasenfinalen Reim einen Kennwert zu erhalten, der dessen normalisierte Dauer insgesamt beschreibt, wird das arithmetische Mittel der normalisierten Dauern der Reimsegmente gebildet. Diese Kennwerte werden in drei Klassen eingeteilt:

$$\begin{aligned} N0 < -1 & \rightarrow \text{Kürzung} \\ -1 \leq N1 \leq 1 & \rightarrow \text{weder Längung noch Kürzung} \\ N2 > 1 & \rightarrow \end{aligned}$$

Längung

Da für die Einteilung in Dauerklassen keine zuverlässigen, in Perzeptionsexperimenten ermittelten Schwellenwerte zur Verfügung stehen, beruhen die Grenzwerte der Klassen auf einem Vergleich der Verteilungen der Segmentdauern in phraseninterner und phrasenfinaler Position. Ist die Berechnung der Längungsstärke erfolgt, so wird an das Merkmalslabel ein entsprechender Zusatz angehängt. Hierbei kodiert `_L0` finale Kürzung, `_L1` weder Längung noch Kürzung und `_L2` Längung.

### 3.2.2 Teilkorpora

Die Datengrundlagen für die verschiedenen Analysen sind aus folgenden Gründen z.T. unterschiedlich:

- Nicht für jede prosodische Grenze kann automatisch ein Merkmalsbündel abgeleitet werden, in dem Information über den melodischen Verlauf, die finale Längung, F0-Reset und Pause/Atmen vorhanden ist. Dies hat verschiedene Gründe.

- (1) Melodische Verläufe am Phrasenende werden nur dann in der prosodischen Etikettierung erfasst, wenn die Phrase mindestens einen Satzaccent enthält. Im

Korpus gibt es aber häufig abgebrochene Phrasen, die keinen Satzakzent enthalten. In diesen Fällen kann automatisch keine Information über den melodischen Verlauf an der prosodischen Grenze erfasst werden.

- (2) Die Berechnung der finalen Längung wird über einen Vergleich von Segmentdauern im Phraseninneren und an prosodischen Grenzen durchgeführt. Um einen aussagekräftigen Messwert abzuleiten, wird die Berechnung für jeden Segmenttyp und für jeden Sprecher getrennt durchgeführt und später durch ein Normalisierungsverfahren in einen segment- und sprecherunabhängigen Wert überführt. Wenn für einen Sprecher bestimmte Segmente in einer der beiden Positionen zu selten vorkommen, kann kein Wert für diese Segmente errechnet und somit keine Angabe zur Längung im finalen Reim gemacht werden.
- (3) Bei bestimmten melodischen Konstellationen am Phrasenende werden im Rahmen der prosodischen Etikettierung keine Beurteilung eines nachfolgenden F0-Resets vorgenommen. Für diese prosodischen Grenzen fehlt folglich die Angabe über das Vorhandensein eines F0-Resets. Für die Gesamtdarstellung der Merkmalsbündel wird nur Bezug auf die prosodischen Grenzen genommen, für die alle Merkmale automatisch abgeleitet werden können.

Durch diese Einschränkung kommt es zu einer Anzahl von insgesamt 3399 flüssigen turninternen prosodischen Grenzen aus beiden Korpora, für die automatisch ein komplettes Merkmalsbündel abgeleitet werden kann.

- Für einige Fragestellungen müssen zusätzliche Etikettierungen oder interpretative Beurteilungen vorgenommen werden. Wegen des großen damit verbundenen Arbeitsaufwands können nur Datenausschnitte diesem Verfahren unterzogen werden.

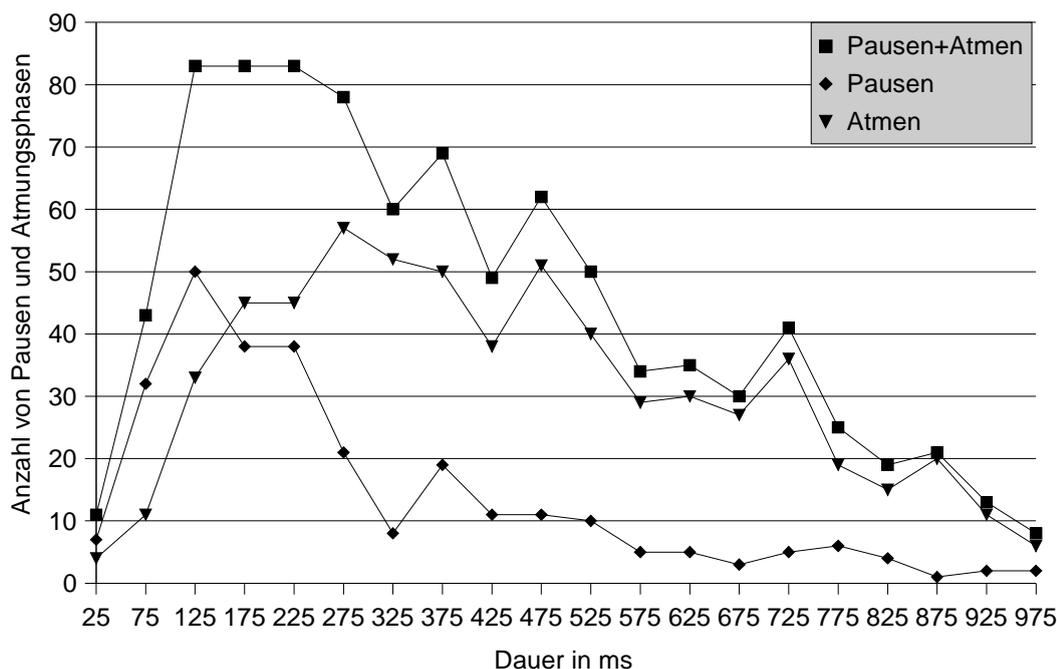
Entsprechend wird für jeden Bereich der Korpusanalyse angegeben, welcher Teil der Datengrundlage den Ausgangspunkt der jeweiligen Teiluntersuchung bildet.

### 3.2.3 Pausen und Atmungsphasen an turninternen prosodischen Grenzen

Pausen und Atmungsphasen im Inneren von Turns werden vor allem anhand der Terminabsprachedaten untersucht, da dieses Korpus aufgrund der nicht-überlagernden Aufnahme für diese Fragestellung zuverlässiger ausgewertet werden kann als die Video Task Daten. Es stellt sich die Frage, ob anhand der Analyse von Pausen- und Atmungsdauern eine Trennung in mehrere Klassen (z.B. lange vs. kurze Pausen) erreicht werden kann. Ein Hinweis auf eine solche Klassenbildung wäre eine bimodale Verteilungen von Pausen- und Atmungsdauern. Butcher (1981) weist eine solche Bimodalität in lesesprachlichen Daten nach.

Über die automatische Analyse der segmentellen Etikettierung ist keine eindeutige Trennung zwischen Pausen und Atmungsphasen zu erreichen. Dies hat zwei Gründe: Erstens kann ein als Atmen etikettierter Signalabschnitt neben einem hörbaren Atmungsgeräusch auch noch eine Pause enthalten und zweitens kann nicht-hörbares Atmen nicht anhand des akustischen Signals detektiert werden. So können als Pausen etikettierte Abschnitte auch mit nicht hörbarer Atmungsaktivität verbunden sein. Trotz dieser Einschränkungen wird im Rahmen der Untersuchungen jeder als Atmen etikettierte Abschnitt als 'Atmungsphase' angesehen und dessen Dauer bestimmt. Genauso werden etikettierte Pausen rein anhand der Etikettierung gemessen.

An 945 der 2470 Phrasengrenzen, die in den Terminabsprachedialogen untersucht werden, treten Pausen oder Atmungsphasen auf (38,3%). Abbildung 3 zeigt die Dauern der Pausen und Atmungsphasen im Terminabsprachekorpus. Es gehen 283 Pausen und 662 Atmungsphasen in die Darstellung ein. Die Pausen- und Atmungsdauern sind für die Abbildung in Klassen mit einer Breite von 50 ms eingeteilt. Die Zahlen auf der x-Achse zeigen den Mittelwert der jeweiligen Klasse. Für die Abbildung sind die oberen 5% der Werte ausgespart, da in diesem Bereich eine breite Streuung der wenigen Fälle bis zu Werten von 2000 ms vorliegt und diese die grafische Darstellung sehr unübersichtlich machen würden.



**Abbildung 3:** Verteilung der Pausen- und Atmungsdauern im Terminabsprachekorpus.

Folgende Kennwerte der Verteilungen der Pausen- und Atmungsdauern ergeben sich (hier inklusive der oberen 5% der Werte, Angaben in ms):

**Tabelle 2:** Kennwerte der Verteilung von Pausen- und Atmungsdauern (Terminabsprachen, Angaben in ms).

	Mittelwert	Standardabweichung	Median	Modalwert (Klasse von)
<b>Pausen</b>	298	236	220	100-150
<b>Atmungsphasen</b>	509	334	268	250-300

Die Verteilungen von Pausen und Atmen zeigen weder einzeln noch zusammen eine deutliche Bimodalität, sondern zusammen eher ein breites Maximum zwischen 100 und 300 ms. Pausen sind durchschnittlich kürzer als Atmungsphasen. Diese Verteilungseigenschaften bieten keinen Ansatzpunkt für eine produktionsdatengestützte Einteilung in mehrere Klassen unterschiedlich langer Pausen- und Atmungsdauern.



Audiobeispiele für Pausen in turninterner Position: 1.2 und 1.6

### **3.2.4 Pausen und Atmungsphasen im Inneren prosodischer Phrasen**

Wenn Aussagen über die separierende Wirkung phonetischer Parameter gemacht werden sollen, stellt sich die Frage, ob und wie häufig phonetische Merkmale, die sich an etikettierten prosodischen Grenzen finden, auch im Inneren prosodischer Phrasen vorkommen. Je häufiger dies für ein Merkmal der Fall ist, desto geringer ist dessen perzeptorisch separierende Wirkung einzuschätzen. In der Etikettierung des Korpus wurden prosodische Grenzen dort gesetzt, wo ein deutlich hörbarer phonetischer Einschnitt wahrgenommen wurde.

Im gesamten Korpus (Terminabsprachen und LINDENSTRASSE) finden sich 1957 etikettierte Pausen und 3821 Atmungsphasen. Davon gibt es nur 31 Pausen (1,6% aller Pausen) und 13 Atmungsphasen (0,3% aller Atmungsphasen), an denen keine prosodische Grenze etikettiert ist. Die auditive Prüfung dieser Fälle zeigt, dass es sich in den meisten Fällen um Etikettierungsfehler handelt, da Pause/Atmen auch in den nicht mit einer prosodischen Grenze etikettierten Fällen einen deutlichen perzeptorischen Einschnitt bewirkt. Pausen und Atmen sind also perzeptiv stark hervortretende akustische Korrelate prosodischer Grenzen.

### **3.2.5 Melodische Muster an turninternen prosodischen Grenzen**

In der PROLAB-Etikettierung werden 12 phrasenfinale melodische Muster unterschieden. Diese Konturen verlaufen von der letzten satzakzentuierten Silbe bis zum Phrasenende. In der Literatur werden einige dieser Konturen als besonders prominent hinsichtlich ihrer separierenden perzeptorischen Wirkung beschrieben. Hierbei handelt es sich in allen Fällen um Konturen mit einer starken melodischen Bewegung. Die Ursache für die separierende Wirkung dieser Konturen wird in zweierlei Eigenschaften gesehen. Zum einen kann eine tonal sehr umfangreiche melodische Bewegung selbst als separierendes Element angesehen werden (Peters et al. 2005b), zum anderen enden die als separierend eingestuften Konturen im oberen oder im unteren Bereich der Sprechstimme. Der auf solche tonalen Bewegungen folgende Einsatz von F<sub>0</sub> liegt aber bei den tief fallenden Konturen meist höher als der vorher erreichte Endpunkt der melodischen Bewegung und bei den hoch steigenden Konturen niedriger. Dadurch entstehen tonale Brüche, die zu

einer perceptiven Separierung führen (Lehiste 1975; Couper-Kuhlen 1983; Batliner et al. 1995).

Insgesamt handelt es sich um 7 der 12 Intonationsverläufe, die in der weiteren Untersuchung als separierende tonale Merkmale eingestuft werden. Die übrigen 5 sind ebene oder leicht steigende oder leicht fallende Muster, die ohne das Hinzukommen weiterer Merkmale nicht zu einem perceptorischen Einschnitt führen. Aufgrund des seltenen Vorkommens sind die vier fallend-steigenden Muster, die in der Etikettierung differenziert werden, in der Untersuchung zusammengefasst.

Tabelle 3 zeigt die Verteilung der melodischen Muster, getrennt in separierende und nicht-separierende Konturen. Da sich die Distributionen in den beiden untersuchten Korpora stark ähneln, werden die Ergebnisse in einer gemeinsamen Tabelle zusammengefasst.

**Tabelle 3:** Melodische Muster an prosodischen Grenzen (Turninterne prosodische Grenzen, Terminabsprachen und LINDENSTRASSE, n=3399).

	<b>Anzahl</b>	<b>%</b>
<b>Starke melodische Bewegung</b>		
Tief fallend	915	26,9
Tief fallend + sehr leicht steigend	202	5,9
Fallend-steigend	121	3,6
Hoch steigend	65	1,9
<b>Gesamt starke melodische Bewegung</b>	<b>1303</b>	<b>38,3</b>
<b>Schwache melodische Bewegung</b>		
Eben	562	16,5
Leicht fallend	962	28,3
Leicht bis mittelhoch steigend	438	12,9
Eben + sehr leicht steigend	23	0,7
Leicht fallend + sehr leicht steigend	111	3,3
<b>Gesamt schwache melodische Bewegung</b>	<b>2096</b>	<b>61,7</b>
<b>Gesamt</b>	<b>3399</b>	<b>100</b>



Audiobeispiele für melodische Muster in turninterner Position: 1.1 – 1.7

### 3.2.6 Melodische Muster im Inneren prosodischer Phrasen

Bei der Anwendung des Etikettierungssystems PROLAB auf Lese- und Spontansprache zeigte sich, dass ein Teil der beobachtbaren Konturen gleichermaßen im Inneren von Phrasen (zwischen zwei Satzakkzenten) und auch am Ende von Phrasen vorkommen. Dies betrifft z.B. die leicht fallende Kontur, die sehr häufig zwischen zwei Gipfelkonturen auftritt (Peters et al. 2005a). Andere melodische Muster, z.B. fallend-steigende Konturen wurden hingegen ausschließlich am Ende prosodischer Phrasen beobachtet. Aus diesem Grund ist schon in den Richtlinien für die prosodische Etikettierung (Peters & Kohler 2004) festgelegt, dass die Etikettierung dieser Muster nur in phrasenfinaler Position vorgesehen ist. Ähnliches gilt auch für die hoch steigenden und die tief fallenden + sehr leicht steigenden Konturen. In neueren, bislang informellen Experimenten mit synthetischem Sprachmaterial konnte allerdings gezeigt werden, dass kombinierte Konturen auch im Phraseninneren vorkommen können. Da aufgrund der Etikettierungspraxis von den separierenden Konturen also nur die tief fallende Kontur im Phraseninneren vorkommen kann, ist die Analyse des phraseninternen Vorkommens separierender Konturen auf diesen Konturtyp beschränkt. Dies ist auch unter dem Gesichtspunkt sinnvoll, dass die tief fallenden Konturen über zwei Drittel aller separierenden Konturen an flüssigen turninternen Phrasengrenzen ausmachen.

Die Analyse der melodischen Muster zwischen den Satzakkzenten innerhalb prosodischer Phrasen in den Terminabsprachedaten ergibt, dass von den 5103 Phrasen des Terminabsprachekorpus 459 Phrasen (9%) im Phraseninneren eine melodische Bewegung bis an die untere Grenze der Sprechstimme zeigen. 155 der 459 Fälle bilden hierbei allerdings den tiefsten Punkt einer steigenden Kontur (Talkontur). Dieser Anteil ist sehr hoch, wenn in Betracht gezogen wird, dass nur 839 der untersuchten 5103 Phrasen überhaupt eine Talkontur haben. Das tief fallende melodische Muster kann in diesem Zusammenhang als Erreichen eines initialen tiefen Zielpunktes von steigenden Konturen angesehen werden, der entscheidend für die auditive Prominenz der Talkonturen ist. Der Rest der tief fallenden Verläufe liegt zwischen zwei Gipfelkonturen im Phraseninneren (Peters et al. 2005a).

Diese Daten zeigen, dass das Konturmerkmal in seiner separierenden Wirkung hinter dem Merkmal Pause zurücksteht, da tief fallende Konturen im Gegensatz zu Pausen auch im

Phraseninneren vorkommen können, ohne eine deutliche perzeptorische Separierung zu verursachen.

### 3.2.7 F0-Reset an turninternen prosodischen Grenzen

Tabelle 4 zeigt die Verteilung von prosodischen Grenzen mit und ohne Reset in den beiden Korpora. Da auch hier sehr ähnliche Zahlen für die Daten der beiden Korpora vorliegen, werden wiederum die Ergebnisse zusammengefasst.

**Tabelle 4:** F0-Reset in der nachfolgenden Phrase (Turninterne prosodische Grenzen, Terminabsprachen und LINDENSTRASSE, n=3399).

<b>Reset</b>	<b>Kein Reset</b>	<b>Keine Entscheidung</b>
2165 (63,7%)	610 (18%)	624 (18,4%)

Etwa zwei Drittel aller analysierten prosodischen Grenzen werden von einem Reset gefolgt, und nur bei 18% gibt es einen fortlaufenden Downstep über die prosodische Grenze hinweg. Die dritte Spalte fasst die Fälle zusammen, in denen im Rahmen der prosodischen Etikettierung keine Entscheidung über Reset an prosodischen Grenzen gefällt wurde. Dies betrifft vor allem prosodische Phrasen, die steigende melodische Muster (Talkonturen oder kombinierte Konturen) enthalten. Der Hintergrund dieser fehlenden Resetentscheidung ist, dass als Bezugspunkt für die Entscheidung, ob ein Reset zwischen zwei Phrasen vorkommt, das F0-Maximum der letzten Gipfelkontur der Phrase festgesetzt ist. Dieser Bezugspunkt wird dann mit der Höhe des ersten F0-Gipfels der Folgephrase verglichen. Wenn der erste Gipfel der Folgephrase tiefer liegt, wird diese als 'kein Reset' etikettiert. Die Erfahrungen im Rahmen der prosodischen Etikettierung haben gezeigt, dass steigende Konturen den kontinuierlichen Downstep zwischen den F0-Maxima innerhalb von Phrasen oft unterbrechen und eine Art phraseninternen Reset verursachen. Deshalb ist am Ende dieser Phrasen auf eine Resetentscheidung verzichtet worden.

### 3.2.8 F0-Reset im Inneren prosodischer Phrasen

Tabelle 5 zeigt die Auftretenshäufigkeiten von prosodischen Phrasen mit durchgehendem Downstep gegenüber Phrasen, in denen der Downstep durch einen oder mehrere Upsteps unterbrochen wird. Diese phraseninternen Upsteps können als Neueinsatz des Downsteps angesehen werden und folglich als phraseninterner Reset.

Für die Analyse wird folgende Stichprobe gewählt: Alle prosodischen Phrasen der Terminabsprachedialoge, die ausschließlich Gipfel der Akzentstufe 2 ('default-Akzentstärke') enthalten (n=1019). Eine solche Auswahl wird getroffen, um Interferenzen mit den Akzenten der Stufe 3 ('emphatische Akzentuierung'), die in der Regel ein sehr hohes F0-Maximum zeigen, auszuschließen. Auch die schwachen Akzente der Stufe 1 ('partielle Deakzentuierung') sollen ausgeschlossen werden, da diese meist nur leichte melodische Bewegungen implizieren. Die gewählte Stichprobe gewährleistet also die Vergleichbarkeit zwischen den untersuchten Gipfelkonturen.

**Tabelle 5:** Upstep bzw. Downstep in Phrasen mit nur 2er-Akzenten (Turninterne prosodische Grenzen, Terminabsprachen, n=1019).

	<b>Downstep (n=790)</b>	<b>Upstep (n=229)</b>
<b>2 Akzente (n1=743)</b>	622 (83,7%)	121 (16,3%)
<b>3 Akzente (n2=214)</b>	138 (64,5%)	76 (35,5%)
<b>4 Akzente (n3=46)</b>	22 (47,8%)	24 (52,2%)
<b>5 Akzente (n4=16)</b>	8 (50,0%)	8 (50,0%)
<b>Gesamt (n=1019)</b>	<b>790 (77,5%)</b>	<b>229 (22,5%)</b>

Die Zahlen in Tabelle 5 zeigen deutlich, dass ein phraseninterner Upstep besonders in Phrasen mit mehr als zwei Gipfelkonturen oft vorkommt. Dieser Upstep ist im Prinzip dasselbe, was an prosodischen Grenzen mit 'Reset' bezeichnet wird: ein Abweichen von der Absenkung der Gipfelpositionen von Gipfel zu Gipfel. Reset kann also in spontaner Sprache seine primäre Funktion nicht in der Abgrenzung prosodischer Phrasen haben. Aus den Erfahrungen in der prosodischen Etikettierung von Spontansprache ist bekannt, dass ein Upstep innerhalb einer prosodischen Phrase meistens mit einer verstärkten Prominenz des erhöhten F0-Gipfels einhergeht. Insofern scheint das phonetische Phänomen Upstep vor allem im Bereich der Fokussierung lexikalischer Einheiten funktionalisiert zu sein und

als Reset im Bereich des Neueinsatzes prosodischer Phrasen eine perzeptorisch weniger relevante Begleiterscheinung zu sein.

Trotzdem soll hier festgehalten werden, dass durchgehender Downstep zwischen Gipfelkonturen in den meisten untersuchten prosodischen Phrasen auftritt (77,5%) und damit auch in spontaner Sprache ein wichtiges phonetisches Phänomen im Rahmen der kohäsiven Gestaltung prosodischer Phrasen ist.

### **3.2.9 Längung an turninternen prosodischen Grenzen**

Für die Erfassung der segmentellen Dauerstruktur in phrasenfinalen Reimen wird ein Vergleich der Segmentdauern in *phraseninternen* Reimen mit den Dauern in *phrasenfinalen* Reimen durchgeführt. Am Ende wird ein Durchschnittswert für jeden phrasenfinalen Reim errechnet, der die Stärke der Längung relativ zu Segmenten in wortfinalen phraseninternen Reimen beschreibt. Die Grundlage für den Vergleich bilden ca. 50.000 Segmente aus phraseninternen Reimen. Ähnliche Verfahren kommen in den Untersuchungen von Ostendorf et al. (1990), Price et al. (1991) und Wightman et al. (1992) zum Einsatz.

Die temporale Struktur der phrasenfinalen Reime wird in drei Klassen eingeteilt: Kürzung, Längung sowie weder Kürzung noch Längung. Als gekürzt werden die Reime angesehen, deren durchschnittliche Segmentdauern kleiner sind als der phraseninterne Mittelwert minus eine Standardabweichung. Als gelängt werden Reime klassifiziert, deren Segmente durchschnittlich länger sind als der phraseninterne Mittelwert plus eine Standardabweichung. Weder gelängt noch gekürzt sind alle Reime dazwischen (siehe Tabelle 6).

Es ergibt sich ein globaler Mittelwert von 64,6 ms für alle phraseninternen Segmente mit einer Standardabweichung von 20 ms. Die Klasse 'Längung' (Dauern oberhalb Mittelwert plus eine Standardabweichung) würde dementsprechend bei einem Faktor von 1,31 beginnen. D.h., dass Reime in die Klasse 'Längung' fallen wenn ihre Dauer sich gegenüber der phraseninternen Dauer um mehr als das 1,3-fache vergrößert.

**Tabelle 6:** Längung am Phrasenende (Terminabsprachen und LINDENSTRASSE, n=3399).

<b>Kürzung</b>	<b>Weder Kürzung noch Längung</b>	<b>Längung</b>
75 (2,2%)	1087 (32%)	2237 (65,8%)

Die folgenden Tabellen zeigen die Häufigkeiten des gemeinsamen Auftretens von separierenden Konturen und den verschiedenen Längungsklassen (Tabelle 7) bzw. von Pausen und den Längungsklassen (Tabelle 8). Ziel der Analyse ist es festzustellen, ob eine Kopplung von segmenteller Längung an einen der beiden Parameter vorliegt.

**Tabelle 7:** Verteilung der Längungsklassen bei separierenden und nicht-separierenden Konturen (Terminabsprachen, n=2470).

	<b>Kürzung</b>	<b>Weder Kürzung noch Längung</b>	<b>Längung</b>
<b>Starke tonale Bewegung (n=1008 (40,8%))</b>	18 (1,8%)	317 (31,5%)	673 (66,7%)
<b>Keine starke tonale Bewegung (n=1462 (59,2%))</b>	41 (2,8%)	460 (31,5%)	961 (65,7%)

**Tabelle 8:** Verteilung der Längungsklassen an prosodischen Grenzen mit und ohne Pausen (Terminabsprachen, n=2470).

	<b>Kürzung</b>	<b>Weder Kürzung noch Längung</b>	<b>Längung</b>
<b>Mit Pause/Atmungsphase (n=945 (38,3%))</b>	8 (0,8%)	143 (15,1%)	794 (84,0%)
<b>Ohne Pause/Atmungsphase (n=1525 (61,7%))</b>	51 (3,3%)	634 (41,6%)	840 (55,1%)

Die Zahlen in Tabelle 7 zeigen keine Verbindung zwischen der Dauer des phrasenfinalen Reims und der Art der melodischen Kontur. Für Phrasengrenzen mit Pausen hingegen gibt es ein deutlich erhöhtes Vorkommen von finaler Längung (84%), gegenüber dem durchschnittlichen Wert von 65,8% (siehe Tabelle 6) und besonders gegenüber den 55,1% gelängerter Reime an den prosodischen Grenzen ohne Pause. Pausen kommen also kaum ohne vorangehende Längung vor. Nur 15,9% aller prosodischen Grenzen mit Pause haben keine vorangehende Längung. Wenn also ein Sprecher innerhalb seiner Äußerung auf eine Pause "zusteuert", ist dies mit einer Verlangsamung der artikulatorischen Abläufe

verbunden. Oller (1973) und Klatt (1975) beschreiben dieses Phänomen für das Englische als 'prepausal lengthening'. Sie beobachten, dass die Silbe vor einer Pause um 60 bis 200 ms länger wird. Diese Längung betrifft vor allem den Vokal und die postvokalisches Konsonanten. Allerdings kann der Umkehrschluss, dass finale Längung zumeist an Pausen gebunden ist, nicht gezogen werden, da an den prosodischen Grenzen ohne Pause eine Längung des finalen Reims mit 55,1% ebenfalls recht häufig ist.

### **3.2.10 Längung im Inneren prosodischer Phrasen**

Eine entsprechende Analyse des Faktors Längung für die phraseninterne Position ist schwierig, da es im Phraseninneren zu einer starken Fluktuation der Dauerverhältnisse in Abhängigkeit von den Akzentuierungsverhältnissen kommt. Häufig sind außerdem Dauerveränderungen, die in Verbindung mit Störungen der Sprechflüssigkeit stehen. Somit sind segmentelle Längungserscheinungen im Inneren von prosodischen Phrasen üblich und stehen in dieser Position nicht in Verbindung mit prosodischer Phrasierung. Wenn Längung an prosodischen Grenzen in flüssiger Sprache vorkommt, dann fast immer in Kombination mit anderen phonetischen Merkmalen. Die automatische Korpusabfrage findet nur 256 prosodische Grenzen, an denen die Längung des finalen Reims das einzige extrahierbare Merkmal ist (7,5% der 3399 untersuchten Grenzen mit kompletten Merkmalsbündeln in beiden Korpora).

### **3.2.11 Phonation an turninternen prosodischen Grenzen**

Neben den oben untersuchten Merkmalen finden sich in der Literatur Hinweise auf phonatorische Veränderungen am Ende prosodischer Phrasen. So beschreiben Local, Wells & Sebba (1985) das Auftreten von Verhauchung und Knarrstimme am Ende von Redebeiträgen. Nakatani & Hirschberg (1994) beobachten Knarrstimme in unflüssiger Sprache am Ende syntaktischer Fragmente. Sie führen weiterhin eine terminologische Trennung zwischen *laryngealization* (an flüssigen prosodischen Grenzen) und *glottalization* (an Abbrüchen) ein, um diese beiden Typen von Knarrstimme zu differenzieren.

Um die Häufigkeit von Knarrstimme an flüssigen prosodischen Grenzen zu ermitteln, wurde jeder Fall von tief fallender Intonation im LINDENSTRASSE- Korpus hinsichtlich des Vorkommens unregelmäßiger Schwingungszyklen und Glottalverschlüssen einer auditiv-visuellen Analyse unterzogen. Im Rahmen dieser Analyse wurde deutlich, dass zwischen zwei Arten von 'Knarrstimme' zu unterscheiden ist, die sich auditiv relativ ähnlich sind, in der visuellen Darstellung in Oszillogramm und Sonagramm aber unterschieden werden können. Eine Art von Knarrstimme (in Tabelle 9 als 'Knarrstimme (unregelmäßig)' bezeichnet) zeigt Unregelmäßigkeiten in der Amplitude und der Frequenz der Schwingungszyklen. Die andere Art von Knarrstimme ('Knarrstimme (regelmäßig)') zeigt einen in Frequenz und Amplitude recht regelmäßigen Verlauf. Allerdings sinkt die Grundfrequenz hierbei in einen sehr tiefen Bereich ab, so dass ein knarrender Gehörseindruck entsteht. Dieses Phänomen ist vor allem bei tiefen Männerstimmen zu beobachten.

Die Restkategorie enthält Fälle, in denen plosivbezogene Knarrstimme oder harte Vokaleinsätze interferieren oder Unsicherheit in der Beurteilung des Phonationstyps vorliegt. Im Deutschen (wie auch im Englischen) können orale stimmlose Plosive unter bestimmten morphologischen und phonotaktischen Rahmenbedingungen durch Glottalverschlüsse oder durch Glottalisierungen ersetzt werden (Kohler 2001). Interferenz von Glottalisierung im Zusammenhang mit harten Vokaleinsätzen tritt auf, wenn morphem-initial Vokale mit Glottalverschluss oder Glottalisierungen produziert werden. Diese Fälle machen eine eindeutige Zuordnung schwierig und werden gekennzeichnet, damit sie später aus der Analyse ausgeschlossen werden können.

Um die zeitliche Extension der nicht-modalen Phonation zu bestimmen, wurden Zeitmarken eingefügt, die den Bereich der Knarrstimme delimitieren. Auch Anfang und Ende vorkommender Glottalverschlüsse wurden mit Zeitmarken abgegrenzt, so dass die Dauer von Glottalverschlüssen und geknarrten Signalabschnitten in deren Umgebung separat voneinander messbar sind.

**Tabelle 9:** Phonationstyp bei terminaler Intonation (LINDENSTRASSE, n=238).

<b>Modalstimme</b>	159 (66,8%)
<b>Knarrstimme (unregelmäßig)</b>	26 (10,9%)
<b>Knarrstimme (regelmäßig)</b>	20 (8,4%)
<b>Glottalverschluss</b>	- (0%)
<b>Rest (Interferenz und unsichere Beurteilung)</b>	33 (13,8%)

Die Zahlen in Tabelle 9 zeigen, dass beide Arten von Knarrstimme etwa gleich häufig vorkommen und zusammen etwa 20% der tief fallenden Intonationsmuster begleiten. Die durchschnittliche Dauer von Knarrstimme (unregelmäßig) beträgt 121 ms. In den Daten ist kein Fall zu beobachten, in dem die Knarrstimme in einem Glottalverschluss endet, wie dies an syntaktischen Abbrüchen vorkommen kann.

Es ist eine klare Sprecherabhängigkeit im Vorkommen der nicht-modalen Phonation beobachtbar. Dies betrifft sowohl das Auftreten von Knarrstimme generell - nur manche Sprecher zeigen dieses phonetische Phänomen - aber auch die Art der Knarrstimme. So häufen sich vor allem die Fälle von regelmäßiger Knarrstimme bei einigen Sprechern im Korpus. Dieser Phonationstyp ist also als sprechertypisch einzustufen.

### **3.2.12 Diskussion zu den Einzelmerkmalen an turninternen prosodischen Grenzen**

Pausen und Atmen sind sehr eng mit prosodischen Grenzen verbunden. Sie sind somit als stark separierende phonetische Merkmale einzuordnen, die fast immer zur Wahrnehmung einer prosodischen Grenze führen. Dies ist bei den tonalen Merkmalen anders. Melodische Bewegungen, die den unteren Bereich der Sprechstimme erreichen sind zwar selten (9% aller prosodischen Phrasen haben eine solche Kontur), aber keine wirkliche Ausnahmeerscheinung. Reset bzw. das Abweichen vom Downstep ist in spontaner Sprache offensichtlich auch in phraseninterner Position häufig; fast ein Viertel der hierauf untersuchten Phrasen zeigen Abweichungen vom Downstep.

Zum Merkmal Längung kann auch ohne eine umfassende Analyse der phraseninternen Dauerstruktur gesagt werden, dass segmentelle Längung keineswegs primär mit dem Ende prosodischer Phrasen assoziiert ist, sondern auch in Verbindung mit prominenten Silben und Sprechflussstörungen vorkommt. Möglicherweise liegt die

Funktion des Dauerparameters hauptsächlich in der Erzeugung auditiver Prominenz auf satzakzentuierten Silben und in der phonetischen Signalisierung von Planungsproblemen.

### 3.2.13 Merkmalsbündel an turninternen prosodischen Grenzen

Nach der separaten Beschreibung einzelner phonetischer Merkmale an turninternen prosodischen Grenzen soll nun deren Bündelung analysiert werden. Der Vergleich zwischen den Merkmalsbündeln im Terminabsprachekorpus und in den LINDENSTRASSE-Daten zeigt sehr ähnliche Verteilungen. Deshalb sind die Ergebnisse für beide Korpora in Tabelle 10 zusammengefasst. Für die Darstellung sind die Fälle ohne Resetentscheidung der Kategorie '-R' zugeordnet.

**Tabelle 10:** Merkmalsbündel an turninternen flüssigen prosodischen Grenzen ohne nachfolgende Häsitiation und ohne Überlagerungen (Terminabsprachen und LINDENSTRASSE, n=3399).

	-	L	P	K	LP	PK	LK	LPK	Ges.
<b>+R</b>	339	418	112	237	377	90	249	343	2165
<b>%</b>	10	12,3	3,3	7	11,1	2,6	7,3	10,1	63,7
<b>-R</b>	227	301	54	81	268	22	153	128	1234
<b>%</b>	6,7	8,9	1,6	2,4	7,9	0,7	4,5	3,8	36,5

Auf der Basis der Verteilung in den Terminabsprachedaten in Kombination mit der Interpretation von Korpusbeispielen nehmen Peters et al. (2005b) die Einteilung in die vier Kategorien PG1-PG4 vor:

- PG1: nur temporale und/oder tonale Merkmale (segmentelle Längung und/oder Reset und/oder starke melodische Bewegung)
- PG2: kurze/mittlere Pause/Atmen (unter 700 ms)
- PG3: lange Pause (über 700 ms) ohne tonale Merkmale
- PG4: lange Pause (über 700 ms) mit tonalem Merkmal

Es ergibt sich weiterhin eine Klasse von prosodischen Grenzen (n=171), für die bei der automatischen Merkmalsextraktion keine phonetischen Korrelate festgestellt werden

können. Im Rahmen einer Untersuchung dieser Fälle zeigt sich, dass diese prosodischen Grenzen häufig in Verbindung mit Register- oder Sprechgeschwindigkeitswechsel stehen und/oder mit Talkonturen enden, die nicht bis in den oberen Bereich der Sprechstimme ansteigen. Da diese Fälle immerhin 6,9% aller turninternen flüssigen Phrasengrenzen ausmachen, werden sie bei der Validierung des von Peters et al. (2005b) erarbeiteten Klassifikationssystems mit einbezogen und in die sich ergebende Neustrukturierung des Systems integriert. In den nachfolgenden Texten und Tabellen werden diese Fälle einer Klasse zugeordnet, die PG0 genannt wird.

Unter Hinzuziehung der in 3.1.4.2 gemessenen Atmungs- und Pausendauern im Terminabsprachekorpus kann die Matrix der Merkmalsbündel um den jeweiligen Atmungs- oder Pausenwert ergänzt werden. Dies ermöglicht die automatische Zuordnung jeder flüssigen turninternen Phrasengrenze zu einer der Kategorien PG1-PG4 (bzw. der Restkategorie PG0). Folgende Zuordnung ergibt sich:

**Tabelle 11:** Verteilung über die PG-Kategorien PG0-PG4 (Terminabsprachen, n=2470).

<b>PG0</b>	171 (6,9%)
<b>PG1</b>	1354 (54,8%)
<b>PG2</b>	770 (31,2%)
<b>PG3</b>	106 (4,3%)
<b>PG4</b>	69 (2,8%)

Da die Analyse der gemessenen Verteilung der Merkmalsbündel über die Klassen keine Möglichkeit einer funktionalen Einordnung gibt, wird der Form-Funktions-Beziehung in den Abschnitten 3.3 und 3.4 mit Hilfe von Perzeptionsexperimenten und Textinterpretationen nachgegangen.



Audiobeispiele für Merkmalsbündel an turninternen prosodischen Grenzen: 1.3, 1.4.

### 3.3 Experimentelle Untersuchungen

Es ist eines der Ziele dieser Arbeit, die unterschiedlich ausgeprägten prosodische Grenzen auf einer perzeptorischen Skala zwischen Kohäsion und Separierung aneinander-grenzender Phrasen anzuordnen und die unterschiedlichen Grade perzeptorischer Kohäsion bzw. Separierung mit sprachlichen Funktionen in Verbindung zu bringen. Die perzeptorische Stärke prosodischer Grenzen kann aber durch die Analyse der Produktionsdaten allein nicht objektiviert werden. Eine perzeptiv orientierte Klassifikation unterschiedlich stark separierender Merkmale und Merkmalsbündelung und deren sprachlicher Funktion kann nur in Verbindung mit systematischen Perzeptions-experimenten herausgearbeitet werden. Die Urteile von Probanden liefern eine wichtige Ergänzung zu den akustischen Analysen und den interpretativen Korpusanalysen dieser Arbeit.

Die Experimente basieren auf der Überführung der im Korpus beobachteten Variation in systematische Stimulusserien. So kann die große Variationsbreite der phonetischen Ausprägungen prosodischer Grenzen über resynthetisiertes Stimulusmaterial in systematischer Weise experimentell überprüft werden. Es werden hierbei u.a. verschiedene Grade von segmenteller Längung im finalen Reim, Pausen verschiedener Dauer und unterschiedliche melodische Muster synthetisch erzeugt und die perzeptorischen Auswirkungen dieser Veränderung im Experiment überprüft.

Die Ergebnisse aus den Perzeptionsexperimenten werden nachfolgend in die funktionale Interpretation von Korpusbeispielen einbezogen. So kann über die Kombination von Produktions- und Perzeptionsdaten ein komplexes Bild erarbeitet werden, das sowohl die Hörer- als auch die Sprecherseite berücksichtigt.

### 3.3.1 Stand der Forschung

Für die perzeptorische Analyse der prosodischen Grenzen spielen folgende Fragen eine wichtige Rolle:

- (1) Differenzieren Hörer konsistent prosodische Grenzen verschiedener Stärke?
- (2) Welche perzeptorische Wirkung haben verschiedene phonetische Merkmalskombinationen?
- (3) Welche Wechselwirkungen zwischen der morphologischen und syntaktischen Struktur und der Wahrnehmung prosodischer Grenzen gibt es?
- (4) Welche sprachliche Funktion haben prosodische Grenzen unterschiedlicher akustischer Ausprägung?
- (5) Welche Schwellenwerte akustischer Parameter gelten bei der Wahrnehmung prosodischer Grenzen?

Mit den oben genannten Fragestellungen befassen sich verschiedene Untersuchungen, von denen diejenigen kurz vorgestellt werden sollen, aus denen Anregungen und Fragen für die Experimente der vorliegenden Arbeit entstammen. Die Frage (1), nach der konsistenten Unterscheidung unterschiedlich starker Grenzen, untersuchten de Pijper & Sandermann (1994) an gelesenen niederländischen Sätzen. Die Teilnehmer eines Hörtests beurteilten hierbei die Stärke eines prosodischen Einschnitts innerhalb kurzer Äußerungen auf einer 10-stufigen Skala zwischen 1 und 10. 1 bedeutet, dass kein Einschnitt wahrgenommen wird, 10 ist ein sehr starker Einschnitt. In den einleitenden Bemerkungen zum Hörtest wurde den Teilnehmern kurz das Konzept der prosodischen Grenzen erklärt. Dann bekamen die Probanden 20 Sätze von 5 bis 15 Wörtern Länge präsentiert, die von je fünf Sprechern gelesen worden waren. Sie notierten die Stärke einer prosodischen Grenze innerhalb jeder Äußerungen auf der 10-Punkt Skala. Die Verwendung von natürlich produziertem Material (gegenüber Resynthesen) hat hierbei Vor- und Nachteile. Vorteilhaft ist, dass durch die natürlichen Produktionsdaten eine große Vielfalt verschiedener natürlicher Realisierungen einbezogen wird. Ein Nachteil ist, dass keine kontrollierte, systematische Variation der Grenzmerkmale möglich ist und so kaum Aussagen über die Interaktion zwischen unterschiedlichen Merkmalsbündelungen

und der wahrgenommenen Separierung gemacht werden können. Ebenso kann nicht systematisch beobachtet werden, zu welchen perzeptorischen Effekten Änderungen in der Ausprägung von Einzelmerkmalen (z.B. Pausendauer) führt. Wie in den Daten des LINDENSTRASSE- und Terminabsprachekorpus zeigte sich in den Testsätzen eine große Streuung in der phonetischen Ausprägung der prosodischen Grenzen. Die fünf Sprecher des verwendeten Korpus machten sehr unterschiedlichen Gebrauch von den Merkmalen 'finale Längung', 'Reset', 'Pause' und 'melodischer Verlauf', sowohl in der Häufigkeit als auch in der Stärke dieser Parameter. Infolgedessen ist die Interpretation der Hörerurteile problematisch. De Pijper und Sandermann kamen im Kern nur zu dem Ergebnis, dass die Wahrnehmung von Separierung zunimmt, je mehr phonetische Merkmale an einer Grenze vorkommen. Das Verfahren lieferte relativ konsistente Urteile naiver Probanden und ist ein guter Ausgangspunkt für die Skalierung prosodischer Kohäsion bzw. Separierung, da die Urteile bereits auf einer Skala abgegeben werden.

In einem parallel durchgeführten Experiment befassten sich die Autoren mit der unter (3) gestellten Frage nach dem Einfluss syntaktischer Strukturen auf die Perzeption prosodischer Grenzen. Hierfür wurden einige der im ersten Experiment präsentierten Sätze in einer delexikalisierten Version präsentiert, so dass die Hörer zwar noch die für die Phrasierung wichtigen phonetischen Eigenschaften der Äußerungen erkennen konnten, aber keine Information über Wortbedeutung und somit auch Syntax bekamen. Die delexikalisierten Stimuli wurden wieder auf der 10-stufigen Skala bewertet. Diese Tests zeigten, dass die Delexikalisierung kaum Einfluss auf die Wahrnehmung der Stärke der Grenze hat. Die Urteile waren vergleichbar mit denen für die nicht delexikalisierten Versionen.

Der Frage (2) nach der perzeptorischen Wirkung verschiedener Merkmalsbündel geht keine mir bekannte experimentelle Untersuchung systematisch nach.

Den Zusammenhang zwischen der morphosyntaktischen Struktur, der phonetischen Ausprägung prosodischer Grenzen und ihrer Wahrnehmung (Frage 3) untersuchten Price et al. (1991) in einer Kombination aus einem Produktions- und einem Perceptionsexperiment. Hier ging es vor allem um die Verteilung von tonalen Mustern und segmenteller Längung vor prosodischen Grenzen an unterschiedlich hohen syntaktischen Knoten. In den Experimenten wurde nachgewiesen, dass sich prosodische Grenzen in Abhängigkeit von syntaktischen Strukturen sowohl in der Art der tonalen

Muster als auch in Bezug auf die Segmentdauern vor der prosodischen Grenze unterschiedlich ausprägen.

In einem Produktionsexperiment lasen professionelle Radiosprecher 35 Satzpaare vor, die ambige syntaktische Konstruktionen enthalten. Die Paare waren in der Lautfolge identisch und unterschieden sich nur in der Position syntaktischer Grenzen. Jeder einzelne Zielsatz aus den 35 Paarungen war in einen Kontext eingebettet, der zu einer inhaltlichen Disambiguierung des Zielsatzes führte. Ein Beispiel für die Ambiguität zwischen Apposition vs. angehängter Präpositionalphrase mit den disambiguierenden Kontexten:

**Kontext I:**

*I can't remember whether you are going to Romania or Bulgaria.*

**Zielsatz:**

Wherever you are, in Romania or Bulgaria, remember me.

**Kontext II:**

I hope you enjoy your tours of Bucharest, Sofia and all the towns you are visiting in Romania and Bulgaria.

**Zielsatz:**

Wherever you are in Romania or Bulgaria remember me.

Im nachfolgenden Perzeptionsexperiment wurden die syntaktisch ambigen Sätze, die von den Radiosprechern innerhalb des disambiguierenden Kontextes vorgelesen worden waren ohne diesen Kontext präsentiert. Die Versuchspersonen bekamen parallel in schriftlicher Form die beiden Kontexte dargeboten und hatten die Aufgabe den jeweiligen Stimulus einem der beiden Kontexte zuzuordnen. Hierbei zeigte sich, dass in 84% der Fälle die Hörer in der Lage waren, die Stimuli aufgrund der phonetischen Realisierung dem Kontext zuzuordnen, in dem der Satz ursprünglich gelesen worden war. D.h. es konnte nachgewiesen werden, dass Sprecher über prosodische Mittel eine syntaktische Disambiguierung vornehmen. Price et. al. 1991 bezogen sich in ihrer Interpretation der Ergebnisse auf Ausprägung und Positionierung der prosodischen Grenzen als phonetische Signale, die den Hörern die Disambiguierung ermöglichen. Allerdings kam es durch den Aufbau des Produktionsexperiments zu unterschiedlichen Prominenzverhältnissen

zwischen den Wörtern in den Zielsatzpaaren, die vermutlich für die Disambiguierung mitverantwortlich waren. Diese Variable wurde weitgehend aus der Interpretation und Diskussion der Daten ausgeschlossen. Auch durch das stark stilisierte, nach syntaktischen Gesichtspunkten konstruierte Sprachmaterial und die Produktion desselben durch professionelle Radiosprecher ist ein Erkenntnisgewinn in Bezug auf spontane Sprache im Dialog kaum zu erzielen.

Neben der Frage nach dem Zusammenhang zwischen prosodischer Phrasierung und der morphosyntaktischen Struktur berührt die Untersuchung von Price et al. auch den thematischen Komplex der sprachlichen Funktionalisierung der prosodischen Phrasierung (Frage 4), nämlich in der Disambiguierung ambiger syntaktischer Strukturen. In Perzeptionsexperimenten ist die Verwendung syntaktisch ambiger Stimuli praktisch, da die Versuchspersonen gefragt werden können, ob sie die eine oder die andere Variante des ambigen Paares wahrgenommen haben. Dies Vorgehen fordert von den Probanden keine metasprachlichen Beurteilungen und erlaubt eine binäre Unterscheidung zwischen Vorhandensein und Abwesenheit prosodischer Phrasierung. Befragungen dieser Art geben aber letztlich nur wenig Aufschluss über allgemeinere Funktionen der prosodischen Phrasierung.

Zur Frage (5) - perzeptorische Schwellen in der Wahrnehmung prosodischer Grenzen - wurde eine große Zahl von Experimenten durchgeführt. Hierbei ist es ein gängiges Verfahren, einen oder mehrere akustische Parameter der Äußerung systematisch zu variieren und die erzeugten Stimulusserien durch Versuchspersonen beurteilen zu lassen. Die Untersuchungen konzentrierten sich meist auf die Wahrnehmung eines einzelnen phonetischen Merkmals, z.B. Pausen bei Megyesi & Gustavson-Capkova (2002) oder melodische Muster bei Isacenko & Schädlich (1970) oder bei House (1990).

In den Experimenten von House (1990) wurden tonale Muster über einer kurzen Äußerung variiert und der Einfluss der Musterbildung auf die perzeptorische Gruppierung der Äußerung untersucht. Der Ausgangsstimulus für die Resynthesen war eine Sequenz von fünf Äußerungen des Wortes [fɛm] (schwedisch für *fünf*), die mit gleicher Prominenz ohne interne prosodische Grenze produziert wurden. Über diese Äußerung wurden dann in F0-Resynthesen verschiedene F0-Muster gelegt, z.B. ein Hutmuster über die ersten drei Wortwiederholungen und ein weiteres Hutmuster über die nächsten zwei Wiederholungen. Weiterhin wurden verschiedene Kombinationen von steigenden und fallenden

Konturen auf gleichen oder unterschiedlichen tonalen Niveaus resynthetisiert. Auch der Neueinsatz der Deklinationslinie wurde variiert. Die Aufgabe der Versuchspersonen war zu entscheiden, ob die Gruppierung der fünf Wiederholungen von [fɛm] in der präsentierten Äußerung 2+3 oder 3+2 ist. House kommt zu dem Ergebnis, dass für die prosodische Gruppierung vor allem eine starke Differenz des tonalen Niveaus zwischen benachbarten tonalen Akzenten ausschlaggebend ist. Wenn z.B. die ersten drei Akzente mit einer gleich hohen Gipfelkontur realisiert sind und die letzten zwei mit Gipfelkonturen auf einem höheren oder niedrigeren tonalen Niveau liegen, wird die Gruppierung so vorgenommen, dass jeweils die gleich hohen Akzente eine Gruppe bilden. Auch hoch steigende und tief fallende Konturen erweisen sich als starke delimitative Merkmale, d.h. an Positionen wo es zu einem hohen Anstieg oder einem tiefen Fall von F0 kommt, wird die Grenze einer tonalen Gruppe wahrgenommen. Der dritte Faktor, der zu einer deutlichen Gruppierung der Sequenz führt ist die Realisierung mehrerer Akzente als Hutmuster: die als Hutmuster realisierten Akzente bilden eine wahrgenommene Gruppe.

Das Design dieser Experimente resultiert aus dem Forschungsziel, die tonalen Merkmale der prosodischen Phrasierung (oder auch Gruppierung) in Experimenten zu evaluieren. Die verwendeten Stimuli machen aber eine Übertragung der Ergebnisse auf anderes sprachliches Material problematisch. Dies hat vor allem zwei Ursachen:

- Die Struktur der Ausgangsäußerung ist sehr ungewöhnlich. Sie besteht aus fünf akzentuierten Silben in Folge, immer im gleichen Wort. Dies ist zum einen auf der lexikalischen und syntaktischen Ebene eine Ausnahmeerscheinung, zum anderen ist eine solche Häufung prominenter Silben sowohl in Lesesprache als auch in Spontansprache höchstens in Aufzählungen zu beobachten.
- Wie bereits in den Untersuchungen der Produktionsdaten gezeigt werden konnte, gehen tonal kodierte prosodische Grenzen im Deutschen fast immer mit anderen Merkmalen einher (siehe 3.1.13). Unter diesem Gesichtspunkt ist die von House untersuchte rein tonale Kodierung eine erhebliche Vereinfachung der tatsächlichen Verhältnisse, es sei denn die phonetische Kodierung von prosodischer Grenzen im Schwedischen verläuft grundsätzlich anders als im Deutschen.

Die Frage nach den Schwellenwerten phonetischer Parameter, die zu wahrgenommenen prosodischen Grenzen führen, hat in der Forschung aber vor allem eine intensive Beschäftigung mit Sprechpausen ausgelöst, da Pausen ein häufiges Korrelat prosodischer Grenzen sind und einen hohen Grad der Trennung signalisieren.

Megyesi & Gustavson-Capkova (2002) befragten Versuchspersonen über das Vorhandensein von Signalpausen in gelesenen und spontanem Schwedisch (*Map Task*). Die gelesenen Daten teilen sich in zwei Klassen: eine Klasse mit professionellen und eine mit nicht-professionellen Lesern. Die Autoren korrelierten die Antworten der Versuchspersonen mit den Ergebnissen einer automatischen Pausendetektion, bei der Signalpausen über 100 ms registriert werden. Die Ergebnisse dieser Untersuchung zeigen, dass die Perzeption von Pausen für die beiden Sprechstile sehr unterschiedlich ist. In gelesener Sprache werden ca. 90% der tatsächlichen Signalpausen erkannt, während es bei den spontansprachlichen Daten nur 35% sind. Ein weiterer interessanter Befund ist, dass die wahrgenommenen Pausen in den spontansprachlichen Daten zu etwa 70% an Stellen gehört werden, an denen auch das Signal eine akustische Pause zeigt. Bei den lese-sprachlichen Daten gibt es einen deutlichen Unterschied zwischen den professionell und den nicht-professionell gelesenen Daten. Unter beiden Bedingungen werden zwar etwa 90% der tatsächlichen Pausen erkannt, aber in den professionell gelesenen Daten haben nur 35% der gehörten Pausen auch eine Signalpause als akustisches Korrelat. Bei den nicht-professionellen Lesedaten sind es aber fast 90%. Diese Untersuchung macht sehr deutlich, dass die Wahrnehmung prosodischer Grenzen nicht allein durch ein Merkmal (z.B. Pausen) gesteuert wird und Versuchspersonen offensichtlich die Beurteilung von Signalpausen mit anderen phonetischen Merkmalen der Separierung vermischen. Hierzu Zwirner & Zwirner (1937:114) in "Phonometrischer Beitrag zur Frage der Leseпаusen":

*Andererseits gibt es .... Fälle, in denen die Wahrnehmung von Sprechpausen (vielleicht auch ihre Intention) einerseits, ihre Registrierung und Messung andererseits nicht übereinstimmen, in denen einwandfrei messbare Pausen nicht wahrgenommen werden (vielleicht auch nicht intendiert) werden und andere Fälle, in denen man derartige Pausen wahrzunehmen glaubt (vielleicht aufgrund*

*syntaktischer oder sprachmelodischer Faktoren), ohne dass sich solche Pausen messend feststellen lassen.*

Die Nicht-Übereinstimmung zwischen Signalpausen und Pausenwahrnehmung ist auch bei Megyesi und Gustafson-Capkova 2002 zu beobachten. Die Ursache für dieses Phänomen ist, dass die Kategorie "Pause" bei phonetisch naiven Hörern nicht mit dem phonetischen Konzept der Signalpause übereinstimmt. In der signalphonetischen Betrachtungsweise des Phonetikers ist eine Pause ein Zeitintervall, in dem die Amplitude des Sprachsignals gegen Null geht (abgesehen von Hintergrundgeräuschen). Signalpausen, die durch Plosivverschlüsse entstehen, sind bei dieser Definition ausgenommen. Die Beurteilung eines Signalabschnitts als Pause durch naive Hörer hingegen ist nicht allein durch tatsächliche Stille im Signal bedingt. Das Verhalten von Versuchspersonen in verschiedenen Perzeptionsexperimenten lässt vielmehr darauf schließen, dass naive Hörer verschiedene Phänomene vermengen, wenn sie über das Auftreten von "Pausen" im Stimulusmaterial befragt werden. So können auch syntaktische und suprasegmentelle Merkmale des Gesprochenen zur Wahrnehmung von "Pausen" führen. Versuchspersonen scheinen also eher über Zäsuren unterschiedlichen phonetischen oder linguistischen Ursprungs als über tatsächliche Signalpausen zu urteilen, wenn sie direkt nach Pausen befragt werden. Um diese Vermengung zu verdeutlichen werden im Folgenden exemplarisch zwei weitere Experimente zur Pausenwahrnehmung beschrieben, die sowohl in der Methode, als auch in den theoretischen Implikationen sehr unterschiedlich vorgehen.

Experimente von Boomer & Dittmann (1962) und Boomer (1965) liefern Ergebnisse, die zu einer funktionalen Differenzierung verschiedener Pausen führen. Sie weisen mit manipuliertem Sprachmaterial nach, dass Versuchspersonen Stimuli mit unterschiedlich langen Signalpausen schon bei einer niedrigeren Differenz der Pausendauern diskriminieren, wenn sich die Pausen innerhalb einer syntaktischen Einheit befinden, als wenn die Pausen an hohen syntaktischen Grenzen auftreten. Boomer & Dittmann gingen folgendermaßen vor: Aus einer aufgenommenen Radiodiskussion wurden Sätze herausgeschnitten, die a.) eine Pause zwischen zwei trennbaren syntaktischen Einheiten enthielten oder b.) eine Pause innerhalb einer syntaktischen Einheit aufwiesen. Ihre Hypothese war, dass Pausen innerhalb der Einheiten eine

niedrigere Perzeptionsschwelle haben als die sog. Junkturpausen zwischen den Einheiten. In der Vorbereitung des Experiments, das die Ausgangshypothese stützen sollte, wurden die Pausen in den ausgewählten Äußerungen auf 100, 200 und 500 ms geschnitten. Im Perzeptionsexperiment wurden den Versuchspersonen nun Paare derselben Äußerung präsentiert, wobei in einer Äußerung die Pause völlig herausgeschnitten war und die Pausendauer im zweiten Stimulus variierte. Die Versuchspersonen wurden gebeten, die Stimuli auf Gleichheit oder Unterschiedlichkeit zu beurteilen, wobei sie darauf hingewiesen wurden, dass mögliche Unterschiede nur in Pausen zwischen den Wörtern auftreten. Die Resultate zeigten, dass Pausen von 100 ms innerhalb der syntaktischen Einheit in etwa 40% der Fälle erkannt wurden, bei 200 ms Pausendauer zu 75% und bei 500 ms zu fast 100%. Die Resultate für die Pausen zwischen den syntaktischen Einheiten wichen erheblich von diesen Ergebnissen ab. Die 100 und 200 ms langen Pausen wurden nur in ca. 20% der Fälle bemerkt, die 500 ms lange Pause nur in knapp 50%. In den durchgeführten Experimenten beurteilten die Probanden Stimuli mit Signalpausen von 100 vs. 200 ms überwiegend als gleich, wenn die Pause am Ende einer syntaktischen Einheit stand. Stimuli mit denselben Pausendauern im Inneren syntaktischer Einheiten hingegen wurden von den Versuchspersonen als ungleich bewertet. Für die Paarungen mit 200 vs. 500 ms Pausen wurden die Stimuli unter beiden syntaktischen Bedingungen als "verschieden" bewertet.

Zur Erklärung der Ergebnisse adaptieren Boomer und Dittman das Konzept der *phonemic clause* (Trager & Smith 1951) (siehe 3.1.2). Eine klare Definition der Einheit *phonemic clause* wird nicht gegeben. Die *phonemic clause* wird als eine syntaktische Einheit beschrieben, die nur einen primären Akzent enthält und mit terminaler Junktur endet. Mit diesem Konzept im Hintergrund ergibt sich für Boomer und Dittman eine syntaktisch orientierte Erklärung; nämlich dass die Pausen innerhalb syntaktischer Einheiten aus syntaktischer Sicht falsch platziert sind und so für den Hörer auffälliger sind und bei geringeren Dauerunterschieden diskriminiert werden, als wenn die Pausen an den Grenzen syntaktischer Einheiten stehen. Aus diesem Ansatz heraus führen Boomer & Dittman die Termini *hesitation pause* vs. *junction pause* für die beiden Pausentypen ein und postulieren unterschiedliche syntaktische Funktionen der zwei Pausentypen.

Die syntaktisch gesteuerten Unterschiede im Hörerverhaltens sind aber nur eine Komponente einer wohl komplexeren Erklärung der experimentellen Befunde. Butcher

(1981) greift die Experimente wieder auf und bietet einen anderen Erklärungsansatz: Pausen haben unterschiedliche perzeptorische Eigenschaften, je nachdem ob sie einen einheitlichen globalen Melodiebogen unterbrechen oder ob sie am Ende eines solchen Bogens stehen. Dieses Konzept erweitert die syntaktisch definierte *phonemic clause* um die Komponente einer globalen Intonationskontur, die innerhalb einer prosodischen Phrase für ein hohes Maß an Kohäsion sorgt. Diese Intonationskontur kann mit einer syntaktischen Einheit deckungsgleich sein, bildet aber auch eine perzeptorische Einheit, wenn keine Deckungsgleichheit besteht. Folglich untersucht Butcher die Wahrnehmungsschwellen für Pausen in Abhängigkeit ihrer Position innerhalb eines einheitlichen Melodiebogens über einer Äußerung und kann eine phonetische Erklärung des von Boomer und Dittmann beobachteten Phänomens belegen: Innerhalb der melodischen Einheit liegt die Wahrnehmungsschwelle für Pausen wesentlich niedriger als zwischen zwei melodischen Einheiten.

Zwei Experimente belegen die phonetische Erklärung: In Stimuli mit monotonisiertem F<sub>0</sub>-Verlauf wurde an jeder Position der Äußerung ein einheitlicher Schwellenwert der Pausenwahrnehmung von 80-90 ms festgestellt. Die Erklärung hierfür ist, dass die intonatorische Strukturierung, d.h. die Einteilung in zusammenhängende Melodiebögen durch die Monotonisierung verloren geht und somit die perzeptorische Trennung von Junkturaussetzungen und Häsitationspausen entfällt. In einem zweiten Experiment wurde mit Nonsenssätzen gearbeitet, die aber mit einer natürlichen Intonation versehen sind. Hier wiederum zeigte sich, dass es zu einer signifikanten Diskrepanz der Schwellenwerte der Pausenperzeption zwischen Pausen innerhalb des Melodiebogens und Pausen zwischen zwei trennbaren globalen melodischen Konturen kommt. Durch diese Experimente wird deutlich, dass für das Phänomen der Gliederung von Äußerungen in syntagmatische Einheiten eher das phonetische Konzept der *prosodischen Phrase*, die mit syntaktischer Phrasierung übereinstimmen kann, es aber nicht muss, als das syntaktisch determinierte Prinzip der *phonemic clause* angemessen ist.

Die Entscheidung von Versuchspersonen über das Vorhandensein von Zäsuren und deren Beurteilung als "Pausen" wird also durch ein Netzwerk syntaktischer und phonetischer Faktoren gesteuert. Hierbei treffen naive Hörer offensichtlich Entscheidungen, die nicht ausschließlich durch Signalaussetzungen gesteuert sind, sondern eher vom Vorhandensein und der Interaktion von Zäsuren auf verschiedenen Ebenen (phonetisch,

syntaktisch, semantisch) abhängen. Diese Situation bei der Kategorisierung von Pausendauern in Perzeptionsexperimenten macht deutlich, dass Versuchspersonen nicht direkt über die Wahrnehmung von Signalpausen befragt werden sollten und dass bei der Erzeugung der Stimuli das gesamte Netzwerk phonetischer Einflussfaktoren berücksichtigt werden muss.

Des Weiteren repräsentieren die drei beschriebenen Untersuchungen zwei unterschiedliche Befragungsverfahren: eine binäre Entscheidung über das Vorkommen einer Pause bei Megyesi und Gustafson-Capkova und eine ebenfalls binäre Entscheidung über Gleichheit und Verschiedenheit in den Diskriminationstests von Boomer und Dittmann bzw. Butcher.

Ein Ziel der vorliegenden Arbeit ist die Untergliederung des Kontinuums von Kohäsion zur Separierung. Hierbei spielen Pausen als separierende phonetische Elemente eine große Rolle. Es soll aber nicht nur die Wirkung der Pause allein untersucht werden, sondern die Interaktion zwischen Pausen verschiedener Dauer und anderen phonetischen Merkmalen der syntagmatischen Separierung. Da Butchers Experimente die perzeptorische Wechselwirkung von Pausen und intonatorischen Strukturen belegen, wird als weitere Variable auch der melodische Verlauf variiert. Außerdem zeigen die Korpusanalysen unter 3.1.4.2 und 3.1.4.3, dass Pausen im untersuchten Korpus fast immer von phrasenfinaler Längung begleitet werden, also soll auch dieses Phänomen als weitere Variable in die Stimulusgenerierung einbezogen werden. Da von einem Kontinuum der Dauer von Pausen ausgegangen wird, das Auswirkungen auf eine ebenfalls kontinuierliche Skala von Kohäsion zu Separierung hat, sollen im Experiment keine binären Entscheidungen der Versuchspersonen abgefragt werden, sondern eine Beurteilung der Stimuli auf einem untergliederten Kontinuum abgegeben werden. Ausgangspunkt hierfür sind die oben beschriebenen Versuche von de Pijper & Sandermann, die auf einer 10-stufigen bipolaren Skala die Stärke prosodischer Grenzen beurteilen ließen. Eine solche Skala ist aus drei Gründen adäquat. Erstens weil sie zu einer direkten Unterteilung der Skala der Kohäsion innerhalb des Stimuluskontinuums führen kann, ohne einen Umweg über binäre Unterscheidungen zu gehen, und zweitens, weil de Pijper und Sanderman in ihren Experimenten ein konsistentes Urteilsverhalten der Versuchspersonen festgestellt haben. Zum dritten kann auf einer solchen Skala direkt ein metasprachliches Urteil über

den Grad der Trennung bzw. Kohäsion zwischen prosodischen Phrasen abgegeben werden.

### **3.3.2 Perzeptionsexperimente zur phonetischen Kohäsion an turninternen prosodischen Grenzen (Experiment 1a, 1b und 1c)**

Die untersuchten Produktionsdaten zeigen ein hohes Maß an Variabilität in der phonetischen Signalisierung prosodischer Grenzen. Sowohl die Zusammensetzung der Merkmalsbündel als auch die Ausprägung der Einzelmerkmale variiert stark. Es ist eines der Ziele dieser Arbeit, unterschiedlich ausgeprägte flüssige prosodische Grenzen auf einer perzeptorischen Skala zwischen Kohäsion und Separierung angrenzender prosodischer Phrasen anzuordnen. Es werden hierfür drei Experimente mit dem Ziel durchgeführt, eine systematische Beziehung zwischen verschiedenen Zusammensetzungen von Merkmalsbündeln und dem Grad der perzeptorischen Kohäsion/Separierung herzustellen. Hierbei werden experimentell die perzeptorischen Effekte von Pausen, finaler Längung und verschiedenen melodischen Mustern hinsichtlich ihrer separierenden Wirkung innerhalb unterschiedlicher Merkmalskombinationen untersucht.

Da Pilotexperimente gezeigt haben, dass F0-Reset keinen Einfluss auf die wahrgenommene Kohäsion bzw. Separierung zwischen prosodischen Phrasen hat, wird dieser Parameter nicht in die systematische Variation des Stimulusmaterials einbezogen. Diese Entscheidung wird auch durch die Korpusanalyse gestützt, die zeigt dass Reset zwischen benachbarten Gipfelkonturen auch im Inneren prosodischer Phrasen häufig vorkommt und so nicht als typisches Merkmal prosodischer Grenzen angesehen werden kann (siehe 3.2.7 und 3.2.8).

Das Stimulusmaterial der Perzeptionsexperimente geht von einer natürlich produzierten Äußerung aus, die ohne eine prosodische Grenze produziert wurde: *Ich möchte eine kleine Portion mit viel Salat.* In den Experimenten werden in systematischer Variation an der Grenze zwischen *Portion* und *mit* drei Parameter verändert:

- Vorhandensein und Dauer von Pausen zwischen *Portion* und *mit*,
- die Längung des Reims von *Portion*,
- melodische Muster auf dem Wort *Portion*.

So entstehen Stimuli mit unterschiedlichen phonetischen Merkmalsbündeln im Übergang zwischen *Portion* und *mit*. Diese Merkmalsbündel repräsentieren typische, in den Produktionsdaten beobachtete phonetische Strukturen. Im Hörtest sollen Probanden zu jedem Stimulus auf einer siebenstufigen bipolaren Skala entscheiden, wie stark getrennt ihnen die zwei Äußerungsteile (*Ich möchte eine kleine Portion* und *mit viel Salat*) erscheinen. Die Skala rangiert hierbei von "nicht getrennt" bis "extrem getrennt".

Das erste Telexperiment (Experiment 1a) soll vor allem die Rolle von Pausen in der wahrgenommenen Separierung prosodischer Grenzen beleuchten und Erkenntnisse über die Interaktion von Pausen mit anderen Merkmalen der Phrasierung bringen. Das zweite Telexperiment (Experiment 1b) konzentriert sich mehr auf die Wahrnehmung von finaler Längung und melodischen Mustern. Das dritte Telexperiment (Experiment 1c) untersucht Interaktionen in der Wahrnehmung prosodischer und syntaktischer Grenzen.

### **3.3.2.1 Experiment 1a**

#### **3.3.2.1.1 Hypothesen**

Folgende Hypothesen werden in Experiment 1a überprüft:

##### **Hypothese 1**

Jeder der drei variierten Parameter (Pausen, Längung, Kontur) beeinflusst die Stärke der wahrgenommenen Trennung zwischen den Äußerungsteilen.

##### **Hypothese 2a**

Pausen sind am stärksten für den Eindruck von Trennung verantwortlich. Die verschiedenen Pausendauern führen zu unterschiedlich wahrgenommenen Trennungsgraden auf der Skala.

### **Hypothese 2b**

Durch die Urteile können Pausenklassen herausgearbeitet werden, in denen trotz unterschiedlicher Dauer ein vergleichbarer Grad an perzeptorischer Separierung erreicht wird.

### **Hypothese 3**

Starke tonale Bewegungen, vor allem tief fallende Melodiemuster an der prosodischen Grenze, haben einen separierenden Effekt. Der tief fallende melodische Verlauf wird deshalb zu größerer perzeptorischer Separierung führen als der ebene und der leicht fallende Verlauf.

In einem Pilotexperiment zu dieser Studie zeigte sich die Tendenz, dass Kombinationen von finaler Längung und starken melodischen Bewegungen in Stimuli ohne Pausen oder mit kurzen Pausen zu einer Zunahme der wahrgenommenen Separierung führen. Bei Stimuli mit langen Pausen hingegen sind die Urteile der Versuchspersonen praktisch unabhängig von Längung oder starken melodischen Bewegungen. Hieraus ergibt sich

### **Hypothese 4:**

Pausen ab einer bestimmten Dauer maskieren die anderen phonetischen Parameter in ihrer trennenden Funktion. Wenn sich dieser Befund absichern lässt, bietet sich die Möglichkeit eine Kategoriengrenze zwischen kurzer und langer Pause zu etablieren, nämlich an dem Punkt, wo die Pausen beginnen, sich maskierend auf andere phonetische Merkmale auszuwirken.

#### **3.3.2.1.2 Stimulusmaterial**

Der Testsatz *Ich möchte eine kleine Portion mit viel Salat.* wird vom Autor gesprochen und in einer schallgedämpften Aufnahmekabine mit 44,1 kHz digital aufgezeichnet. Der Satz zeigt in der Ausgangsäußerung keine interne prosodische Grenze. Er hat Satzakzente auf den Wörtern *Portion* und *Salat* und ist mit einem mittleren Sprechtempo gesprochen. Die Änderungen der phonetischen Parameter betreffen im Fall der Längung und des melodischen Verlaufs den Reim von *Portion*. Die unterschiedlich langen Pausen werden an der Wortgrenze zwischen *Portion* und *mit* hereingeschnitten. Durch diese Manipulatio-

nen entstehen an dieser Stelle verschieden ausgeprägte prosodische Einschnitte, die die Äußerung in zwei Teile untergliedern.

### **(1) Längung des finalen Reims**

Die Dauer des finalen Reims von *Portion* wird in drei Stufen vergrößert. Ausgehend von einer kurzen Reimdauer der Originaläußerung, die nicht zur Wahrnehmung einer prosodischen Grenze führt, wird der Reim der Äußerung mit dem PSOLA-Algorithmus (Moulines & Carpentier 1990) in PRAAT (Boersma & Weenink) sukzessive vergrößert. Drei Längungsfaktoren werden angewendet: 1,2 und 1,4 und 1,6. Da aus Untersuchungen zur finalen Längung bekannt ist, dass die Domäne der phrasenfinalen Längung bei flüssiger Phrasierung in der Regel der phrasenfinale Reim ist (Klatt 1975; Whightmann et al. 1992) wird die Daueränderung auf die Segmente des Reims *Portion* angewendet. Durch dieses Verfahren ergeben sich folgende Dauern im Reim von *Portion*:

- Ausgangsstimulus ohne wahrnehmbare Längung (Faktor 1), Reimdauer: 230 ms,
- Erste Längungsstufe (Faktor 1,2), Reimdauer: 275 ms,
- Zweite Längungsstufe (Faktor 1,4), Reimdauer: 325 ms,
- Dritte Längungsstufe (Faktor 1,6), Reimdauer: 370 ms.

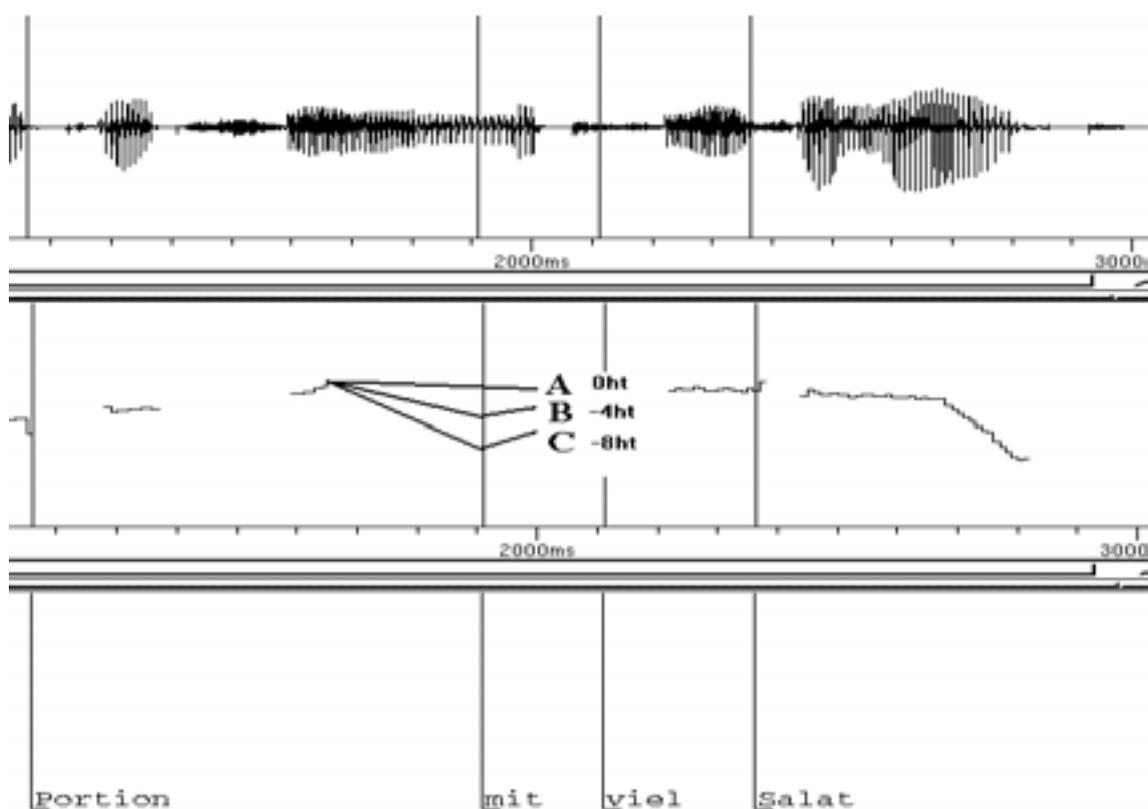
Keiner der so manipulierten Stimuli mit den Längungsfaktoren 1 bzw. 1,2 und 1,4 vermittelt den Eindruck einer zögerlichen oder unflüssigen Sprechweise, wie dies bei einer noch stärkeren Dehnung des finalen Reims auftreten kann. Der Stimulus mit der stärksten finalen Längung (Faktor 1,6) kann hingegen als leicht zögerlich interpretiert werden. Der mittlere Längungsfaktor (1,4) entspricht der durchschnittlichen, in den Produktionsdaten gemessenen Längung in finalen Reimen vor flüssigen prosodischen Grenzen.

### **(2) Verlauf des phrasenfinalen melodischen Musters**

Die Ausgangsäußerung hat zwei deutliche Satzakkente, die durch zwei F0-Gipfel mit einer leichten Einbuchtung dazwischen realisiert sind. Ein Gipfelpunkt liegt bei 130 Hz, in der Mitte des Silbennukleus von *Portion*. Der zweite, breitere Gipfel beginnt auf *viel* und

erstreckt sich bis in den Silbenukklus der satzakzentuierten Silbe von *Salat* bei etwa 120 Hz. Damit liegt ein Downstep von etwa 8% zwischen den beiden Gipfeln vor. Danach beginnt ein starkes Absinken der Grundfrequenz bis zum Äußerungsende.

Abbildung 4 zeigt die drei F0-Muster, die über die PSOLA-Komponente von PRAAT generiert wurden.



**Abbildung 4:** Melodische Muster im Stimulusmaterial.

Die drei resynthetisierten melodischen Muster repräsentieren drei Kategorien phrasenfinaler Intonationsmuster, die in der prosodischen Etikettierung mit *KIM* bzw. *PROLAB* differenziert werden.

Der Verlauf A ist eine direkte Verbindung zwischen dem Gipfel auf *Portion* und dem Gipfel auf *Salat*. Diese Art der Gipfelkonkatenation wird als Hutmuster bezeichnet (‘t Hart et al. 1990; Peters et al. 2005a).

Der Verlauf B weist ein Absinken von F0 nach dem Gipfel auf *Portion* auf, das seinen Endpunkt am Ende des alveolaren Nasals bei 105 Hz erreicht. Das entspricht einer tonalen Differenz von vier Halbtönen zum vorangehenden Gipfelpunkt. Zum Beginn des

nachfolgenden bilabialen Nasals von *mit* beginnt ein linearer Wiederanstieg, der sich bis zum Vokal von *viel* fortsetzt und bei 120 Hz endet.

Für Verlauf C wird der Endpunkt des Absinkens nach dem Gipfel auf *Portion*, ausgehend vom Endpunkt der Kontur B, nochmals um vier Halbtöne herabgesetzt. So entsteht ein Endpunkt der Bewegung bei 82 Hz. Der Wiederanstieg verläuft wiederum mit dem Vokal in *viel* als Zielpunkt, nur entsprechend steiler.

In den untersuchten Produktionsdaten finden sich diese drei Muster mit großer Häufigkeit an flüssigen turninternen Grenzen. Die ebene Kontur tritt mit einer Häufigkeit von 16,5% auf, die leicht fallende mit 28,3% und die tief fallende mit 26,9% (siehe Tabelle 3). Damit haben diese drei Muster die höchsten Vorkommenshäufigkeiten vor allen anderen melodischen Verläufen und werden deshalb im Stimulusmaterial abgebildet.

### **(3) Pausen**

Um das Kontinuum von einer nicht hörbaren Pause zu einer sehr langen Pause perzeptorisch zu untersuchen, werden Pausen verschiedener Dauer in den Ausgangsstimulus hereingeschnitten. Es werden sieben Pausendauern gewählt. Ausgehend von der kleinsten Pause mit 50 ms wird eine Multiplikation mit dem Faktor 1,6 vorgenommen. Folgende Pausendauern ergeben sich (in ms): 50, 80, 130, 210, 330, 560, 890. Diese Spannbreite zwischen der kürzesten und der längsten Pause (50 ms - 890 ms) entspricht in etwa der im Korpus beobachteten Variation von Pausen- und Atmungsdauern. Pausen im Inneren von Redebeiträgen oberhalb einer Dauer von 900 ms kommen im Korpus kaum vor; nur etwa 0,5% der Pausen- und Atmungsdauern im Terminabsprachekorpus liegen in diesem Bereich.

Das Hereinschneiden von Pausen in Äußerungen, in denen ursprünglich keine Pause vorhanden ist, kann sehr leicht zum Eindruck eines unatürlichen, ruckartigen Einschnitts im Signal führen, da Pausen in natürlicher Sprache in der Regel von einer Abnahme der Intensität und von vorangehender segmenteller Längung begleitet sind. Weiterhin kommt es auch über Wortgrenzen hinweg zu Koartikulation zwischen benachbarten Lauten. Wenn diese koartikulierten Laute künstlich voneinander getrennt werden, entsteht ein unnatürlicher Bruch an dieser Stelle. Wegen der genannten Gründe verursacht andererseits aber auch das Herausschneiden von ursprünglich vorhandenen Pausen oft den Eindruck von Unnatürlichkeit. Bei Pilotversuchen zu diesem Experiment

wurde speziell darauf geachtet, dass nicht der Eindruck eines technischen Artefakts entsteht. Hierbei zeigte sich, dass im Übergang zwischen zwei Wörtern manche Lautkombinationen recht gut auseinandergeschnitten und Pausen eingefügt werden können, ohne dass ein zu starker, technisch bedingter Einschnitt entsteht. So kann ohne weiteres eine Pause in den Übergang zwischen dem wortfinalen Nasal in *Portion* und dem wortinitialen Nasal von *mit* hereingeschnitten werden. Der Eindruck von Unnatürlichkeit ist hierbei nur sehr gering. Bei sehr genauem Abhören der Schnittstelle kann lediglich im finalen alveolaren Nasal von *Portion* eine leichte Bewegung in den bilabialen Verschluss festgestellt werden. Andere Lautkombinationen erwiesen sich als wesentlich schlechter trennbar.

Die Parametermanipulationen in den Bereichen Länge des finalen Reims (4 Längungsstufen), phrasenfinales Melodiemuster (3 Konturverläufe) und Pausendauer (keine Pause und sieben Pausendauern) und die systematische Kombination aller Merkmale ergibt 96 Stimuli mit verschiedenen Merkmalskombinationen.



Audiobeispiele für die Stimuli aus Experiment 1a: 4.1

### 3.3.2.1.3 Durchführung

Aus den erzeugten Stimuli wird eine Audiodatei erzeugt, die im Hörexperiment präsentiert wird. Die Testserie umfasst drei Wiederholungen jedes Stimulus. Die Abfolge aller Stimuli ist randomisiert. Nach je zehn Stimuli ist eine längere Pause und ein Piepton. Zwischen zwei Stimuli ist innerhalb der Blöcke eine Pause von vier Sekunden, in der die Probanden das jeweilige Urteil abgeben können. Die gesamte Stimulusserie ist 40 Minuten lang.

Vor Beginn des Experiments wird den Probanden eine Aufnahme mit Instruktionen für den Versuch präsentiert. Dieser Präsentationsmodus hat zwei Gründe: Erstens ist es wichtig, dass die Instruktionen in den verschiedenen Datenerhebungssitzungen identisch sind, um auszuschließen, dass Unterschiede in der Instruierung die Ergebnisse zwischen verschiedenen Sitzungen beeinflussen. Zweitens sind in die Instruktionen verschiedene Audiobeispiele eingebunden, die bei einer mündlichen Präsentation nicht exakt reproduziert werden könnten. Die Instruktionen umfassen eine kurze Einführung in das Konzept der Äußerungsgliederung durch

prosodische Grenzen und einige Audiobeispiele, die unterschiedlich ausgeprägte Grenzen veranschaulichen. Weiterhin wird die konkrete Fragestellung und der Ablauf des Experiments erklärt. Eine schriftliche Fassung der Instruktionen findet sich im Anhang.

Nach dem Abspielen der Instruktionen werden zehn Stimuli präsentiert, die das Spektrum der gesamten phonetischen Variation repräsentieren. Die Versuchspersonen geben für jeden Stimulus ein Urteil auf einer Skala ab. Danach gibt es eine Unterbrechung mit der Möglichkeit, Fragen zum Ablauf zu stellen. In keiner der Datenerhebungssitzungen gab es grundlegende Verständnisfragen.

Die Versuchspersonen beantworten im Experiment für jeden Stimulus folgende Frage *Wie stark sind der erste und der zweite Teil der Äußerung voneinander getrennt?* Für die Befragung wird eine 7-stufige bipolare Antwortskala gewählt, die folgendermaßen aufgebaut ist:

nicht getrennt			mittel			extrem getrennt
0	1	2	3	4	5	6

Skalen dieser Art werden *Equal appearing interval scales (EAI)* genannt und gehen auf Messungen in der experimentellen Psychologie zurück (Thurston & Chave 1929; Likert 1932). Sie sind geeignet, um die Urteile von Versuchspersonen als skalierte Messwerte aufzunehmen und werden in neuerer Zeit auch zunehmend häufiger in phonetischen Experimenten, vor allem im Bereich der Prosodieforschung, eingesetzt (de Pijper & Sanderman 1994; Grabe et al. 1997; Caspers 2000; Rietveld et al. 2002; Chen 2003).

An dem Experiment nahmen insgesamt 14 Versuchspersonen in zwei getrennten Datenerhebungssitzungen teil. Wie auch in allen weiteren Versuchen sind die Versuchspersonen zum größten Teil Studierende der Universität Kiel. Alle Versuchspersonen sind deutsche Muttersprachler.

### 3.3.2.1.4 Ergebnisse

In diesem Abschnitt werden die oben formulierten Hypothesen anhand der Ergebnisse des Perzeptionstests überprüft. Die Varianzanalyse liefert folgende Ergebnisse.

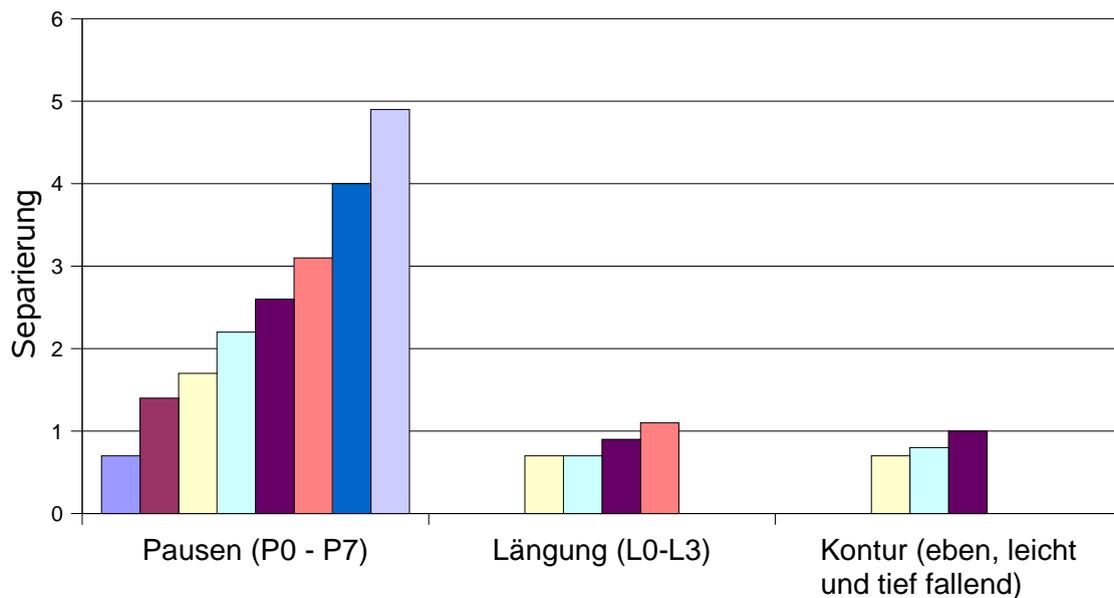
**Tabelle 12:** Ergebnisse der Varianzanalyse für Experiment 1a.

<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Längung des finalen Reims	2,13	8,57	0,001
Finale Kontur	1,29	1,91	0,185
Pause	2,11	250,86	0,000
Längung * Kontur	4,16	1,84	0,132
Längung * Pause	6,97	1,26	0,279
Kontur * Pause	6,06	2,3	0,042
Längung*Kontur*Pause	8,53	2,5	0,014

**Hypothese 1:**

Jeder der drei variierten Parameter beeinflusst die Stärke der wahrgenommenen Trennung zwischen den Äußerungsteilen.

Abbildung 5 zeigt die Auswirkungen der manipulierten Parameter Pause, Längung und Konturverlauf auf die wahrgenommene Separierung zwischen den Äußerungsteilen. Um eine anschauliche Darstellung zu erhalten, die den Effekt jedes einzelnen Parameters, getrennt von der Variation der anderen Parameter herausarbeitet, wird für alle drei Parameter von dem Stimulus mit den geringsten Anzeichen phonetischer Separierung ausgegangen. Dies ist der Stimulus ohne Pause, ohne finale Längung und mit einem ebenen F0-Verlauf. Für die Darstellung der Pausen werden die Parameter Längung und Kontur konstant gehalten, d.h. in der Pausenserie verändert sich nur die Dauer der Pausen. Ebenso wird für Längung und Kontur vorgegangen. Um die perzeptorische Wirkung der vier Längungsstufen zu zeigen, sind hier die Urteile der Stimuli ohne Pause und mit ebenem Konturverlauf abgebildet. Die Variable 'Kontur' wird im konstanten Kontext ohne Pause und ohne Längung dargestellt. Diese Darstellungsweise ist auch unter Einbeziehung der prüfstatistischen Ergebnisse gerechtfertigt, die zeigen dass die Interaktion zwischen den getesteten Parametern nur sehr gering ist.



**Abbildung 5:** Wahrgenommene Separierung für die Pausenstufen, die Längungsstufen und die Konturen. In der Abbildung entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteile, ein Wert von 6 einer maximalen Trennung.

Abbildung 5 zeigt deutlich, dass jede Pausenstufe sukzessive zu einer Zunahme der wahrgenommenen Separierung führt. Die Effekte der Längung des finalen Reims und des phrasenfinalen Konturverlaufs hingegen sind sehr schwach ausgeprägt. Die prüfstatistische Bearbeitung zeigt, dass Pausen hochsignifikant die Urteile beeinflussen ( $F=250,86$ ,  $p=0,000$ ; siehe Tabelle 12). Die Berechnung von Einzelkontrasten zeigt, dass der Unterschied von jeder Stufe zur nächsten hochsignifikant ist. Längung hat ebenfalls einen hochsignifikanten Einfluss ( $F=8,57$ ,  $p=0,001$ ; siehe Tabelle 12), der Einfluss der Kontur hingegen ist nicht signifikant ( $F=1,91$ ,  $p=0,185$ ; siehe Tabelle 12). Hierbei zeigt die deskriptive Darstellung (Abbildung 5) den großen Unterschied zwischen den Urteilen jeder Pausenstufe und die minimalen Unterschiede der Serien, in denen Längung und Konturmerkmal verändert wurden, sehr deutlich. Hypothese 1 kann also nur zum Teil bestätigt werden.

### **Hypothese 2a:**

Pausen sind am stärksten für den Eindruck von Trennung verantwortlich. Die verschiedenen Pausendauern führen zu unterschiedlich wahrgenommenen Trennungsgraden auf der Skala.

Hypothese 2a wird durch die Ergebnisse bestätigt.

### **Hypothese 2b:**

Durch die Urteile können Pausenklassen herausarbeitet werden, in denen trotz unterschiedlicher Dauer ein vergleichbarer Grad an perzeptorischer Separierung erreicht wird.

Die Hypothese 2b kann nicht bestätigt werden. Es kommt von jeder Pausenstufe zur nächsten zu einem hochsignifikanten Anstieg der wahrgenommenen Separierung. Allerdings zeigt sich, dass die Effektstärke variiert: Der Unterschied zwischen Stimuli ohne Pause (P0) und erster Pausendauer (P1) und die Differenzen von P5 nach P6 und P6 nach P7 sind etwas größer als die Unterschiede im mittleren Bereich.

### **Hypothese 3:**

Starke tonale Bewegungen, vor allem tief fallende Melodiemuster an der prosodischen Grenze haben einen separierenden Effekt. Der tief fallende melodische Verlauf wird deshalb zu größerer perzeptorischer Separierung führen als der ebene und der leicht fallende Verlauf.

Diese Hypothese kann durch die statistische Analyse der Ergebnisse nicht belegt werden, obwohl sich in den Daten ein Trend in Richtung stärkerer Separierung des tief fallenden Musters abzeichnet.

### **Hypothese 4:**

Pausen ab einer bestimmten Dauer maskieren die anderen phonetischen Parameter in ihrer trennenden Funktion. Wenn sich dieser Befund absichern lässt, bietet sich die Möglichkeit eine Kategoriengrenze zwischen kurzer und langer Pause zu etablieren, nämlich an dem Punkt, wo die Pausen beginnen, sich maskierend auf andere phonetische Merkmale auszuwirken.

Da der separierende Effekt der Konturverläufe und der finalen Längung selbst ohne das Vorhandensein von Pausen sehr schwach ausfällt, kann dieser Hypothese anhand der Ergebnisse nicht weiter nachgegangen werden. Auch die Varianzanalyse zeigt keine systematischen Interaktionen zwischen den Parametern.

### 3.3.2.1.5 Diskussion

Die Ergebnisse des Versuchs sind in zweierlei Hinsicht überraschend. Zum einen wegen des geringen Einflusses von Längung und Konturverlauf auf die Urteile der Versuchspersonen und zum anderen wegen der schwachen Interaktion zwischen den getesteten Parametern.

Die Ursache für dieses Ergebnis liegt möglicherweise im Aufbau des Experiments 1a, denn nur 12 der 96 verwendeten Stimuli haben keine Pause. Aus diesem Grund konzentrieren sich die Versuchspersonen möglicherweise vor allem auf das Vorkommen von Pausen und deren unterschiedliche Dauern und geben nur dort, wo Pausen zwischen den Äußerungsteilen vorkommen, ein Urteil für deutliche Separierung ab. Ein solches Verhalten kann die Effekte von Konturverlauf und Längung sehr gering ausfallen lassen. Um diese Annahme zu überprüfen, wird in Experiment 1b eine Auswahl aus den 96 Stimuli des ersten Experiments getroffen, die ein Gleichgewicht zwischen Stimuli mit und Stimuli ohne Pause herstellt (siehe 3.3.2.2.1).

Die Ergebnisse zeigen weiterhin, dass die Versuchspersonen bereits die erste Pausenstufe (50 ms) klar von den Stimuli ohne Pause differenzieren. Diese Ergebnisse stehen im Widerspruch zu verschiedenen Experimenten in denen untere Schwellenwerte für die Wahrnehmung von Sprechpausen im Bereich zwischen 80 ms und 250 ms angesetzt werden (Butcher 1981, Goldmann-Eisler 1968, Grosjean & Deschamps 1975). Die Ursache für diesen Befund liegt m.E. vor allem in der Art der Befragung in den verschiedenen Experimenten. In den Experimenten, die einen Schwellenwert belegen, werden die Versuchspersonen meist direkt nach dem Vorhandensein einer Pause befragt. Da die Versuchspersonen aber meist weder eine klare Vorstellung von Signalpausen noch ein ausgebildetes analytisches Hörvermögen haben, führt diese Art der Befragung zu zwei Artefakten: Erstens werden z.T. Urteile für das Vorhandensein von Pausen abgegeben, obwohl gar keine Pause vorhanden ist, und zweitens werden Urteile gegen das Vorhandensein von Pausen abgegeben, obwohl Pausen vorhanden sind. Die Nicht-Übereinstimmung zwischen Signalpausen und Pausenwahrnehmung ist auch in anderen Experimenten zu beobachten (z.B. Zwirner & Zwirner 1937; Megyesi & Gustafson-Capkova 2002; siehe 3.3.1 und 3.3.2.5).

### **3.3.2.2 Experiment 1b**

#### **3.3.2.2.1 Stimulusmaterial und Durchführung**

Während Experiment 1a vor allem auf die Wahrnehmung von Pausen ausgerichtet ist, steht die perzeptorische Wirkung der melodischen Muster und der phrasenfinalen Längung im Zentrum von Experiment 1b. Wie in Experiment 1a gezeigt wurde, steht die separierende Wirkung tonaler Brüche und phrasenfinaler Längung deutlich hinter dem Grad der Separierung zurück, die selbst kurze Pausen verursachen. Die Ursache hierfür liegt aber, wie bereits beschrieben, möglicherweise im Übergewicht der Stimuli mit Pausen im ersten Experiment. Experiment 1b arbeitet deshalb mit einer Auswahl von Stimuli aus Experiment 1a und stellt ein Gleichgewicht zwischen Stimuli mit und ohne Pausen her. Folgende Stimuli aus Experiment 1a werden ausgewählt und in derselben Weise wie in Experiment 1a präsentiert und auf der 7-stufigen Skala beurteilt:

- 3 Konturtypen: eben, leicht fallend, tief fallend (wie in Experiment 1a),
- 3 Längungsfaktoren: keine Längung des finalen Reims (Längungsfaktor 1), Längungsfaktor 1,2 und Längungsfaktor 1,4. Der Längungsfaktor 1,6 entfällt, da so stark gelängte Reime möglicherweise als zögerliches Sprechen wahrgenommen werden und sich so die Bereiche der flüssigen und unflüssigen Phrasierung in diesem Experiment mischen würden.

Die resultierenden 9 Stimuli (drei melodische Muster und drei Längungsstufen) werden entweder ohne Pause oder mit einer Pause von 210 ms präsentiert. Diese Pausendauer entspricht etwa der mittleren Pausendauer in den untersuchten Korpusdaten (siehe 4.1.1, Tabelle 2). So ergibt sich eine Anzahl von 18 verschiedenen Stimuli. Die Testserie umfasst zehn Wiederholungen jedes Stimulus. Der sonstige Ablauf ist parallel zu Experiment 1a. Die Instruktionen für das Experiment sind identisch mit den Instruktionen für Experiment 1a und befinden sich im Anhang. An dem Experiment nahmen 24 Versuchspersonen in drei getrennten Datenerhebungssitzungen teil. Die Hypothesen

entsprechen mit Ausnahme der Hypothesen 2b und 4, die nicht noch einmal getestet werden, denen aus Experiment 1a.



Audiobeispiele für die Stimuli aus Experiment 1b: 4.1

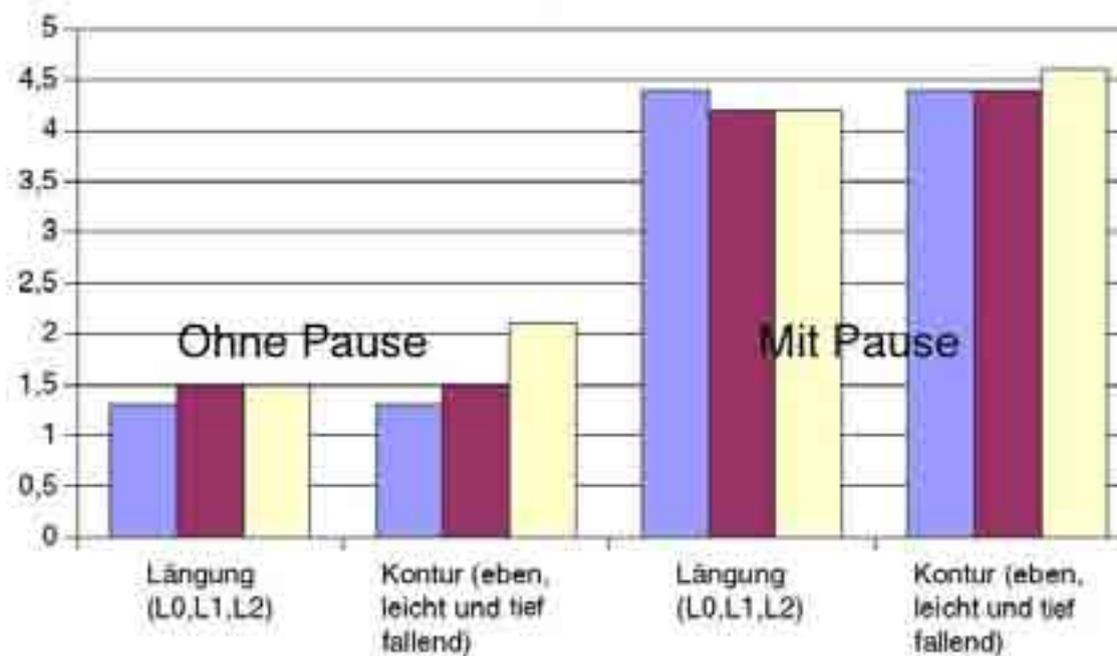
### 3.3.2.2.2 Ergebnisse

Die Varianzanalyse der Urteile aus Experiment 1b liefert die folgenden Ergebnisse.

**Tabelle 13:** Ergebnisse der Varianzanalyse für Experiment 1b.

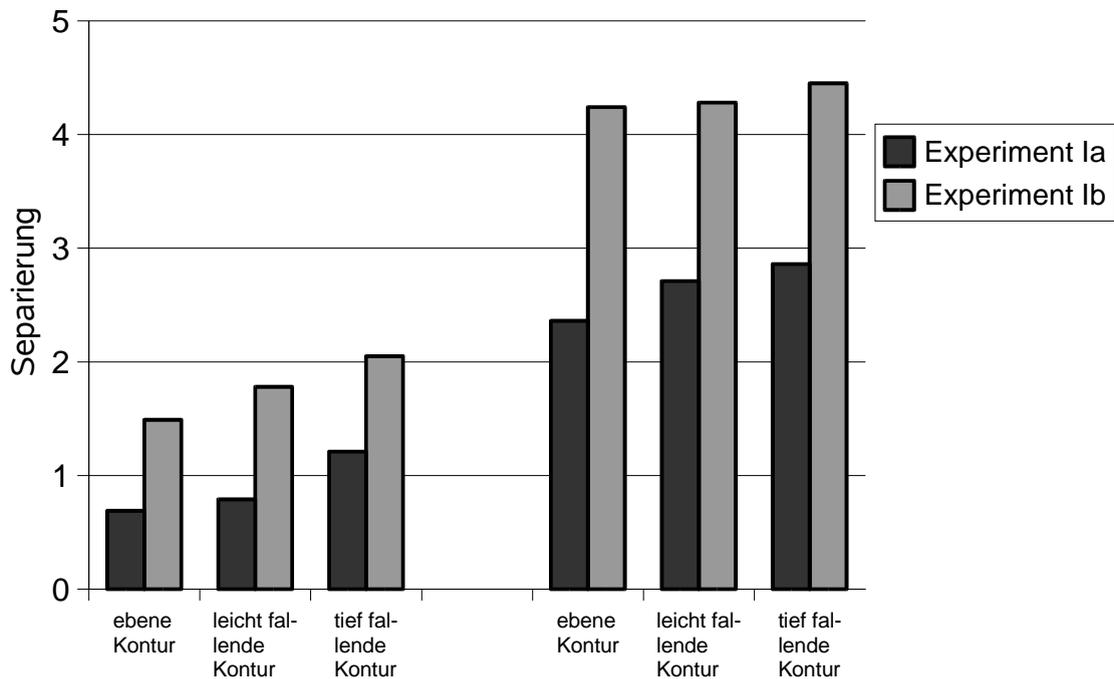
<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Längung des finalen Reims	1,53	2,69	0,095
Finale Kontur	1,49	22,4	0,000
Pause	1,0	303,58	0,000
Längung * Kontur	3,13	1,9	0,134
Längung * Pause	1,89	6,8	0,003
Kontur * Pause	1,59	7,14	0,004
Längung*Kontur*Pause	2,99	2,8	0,047

Abbildung 6 zeigt die perzeptorische Wirkung der Parameter 'Längung' und 'Kontur'; links in Stimuli ohne und rechts in Stimuli mit Pause. Um den Effekt der Einzelparameter möglichst isoliert zeigen zu können, wird bei der Darstellung des Effekts der Längung für alle drei Stufen von einem ebenen melodischen Verlauf am Phrasenende ausgegangen. Für die Darstellung der Auswirkung der phrasenfinalen Kontur wird für alle Konturen der Kontext der Längungsstufe L0 (keine Längung) gewählt. In der Abbildung entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteilen, ein Wert von 6 einer maximalen Trennung.



**Abbildung 6:** Wahrgenommene Separierung für die Längungsstufen und die Konturen für die Stimuli mit und ohne Pause in Experiment 1b.

Die Varianzanalyse (Tabelle 13) zeigt, dass im Gegensatz zu Experiment 1a hier die Konturen einen hochsignifikanten Effekt haben ( $F=22,4$ ,  $p=0,000$ ), während der Effekt der Längung nicht signifikant ist ( $F=2,69$ ,  $p=0,095$ ). Außerdem ist Interaktion zwischen dem Vorhandensein von Pausen und der Wahrnehmung der Konturen signifikant: Der Unterschied zwischen den drei Konturen in Stimuli ohne Pause ist stärker als in Stimuli mit Pausen ( $F=7,14$ ,  $p=0,04$ ). Dies deutet auf den Maskierungseffekt von Pausen hin.



**Abbildung 7:** Vergleich der wahrgenommenen Separierung zwischen zwei identischen Stimulusserien in den Experimenten 1a und 1b. Alle Stimuli haben eine Längung des finalen Reims mit Faktor 1,2. Links sind die Ergebnisse für die Serie ohne Pause dargestellt und rechts die Ergebnisse für die Serie mit einer Pause von 215 ms.

Im Vergleich zwischen den Bewertungen der identischen Stimuli in beiden Experimenten zeigt sich ein deutlicher Einfluss des unterschiedlich zusammengesetzten Stimulusmaterials. Die Urteile für alle Stimuli fallen in Experiment 1a deutlich niedriger aus. D.h., die Stimuli ohne Pause verschieben sich in Experiment 1b auf der Skala um etwa einen Skalenpunkt in Richtung der extremen Separierung, die Stimuli mit Pause um etwa zwei Skalenpunkte. Diese Befunde zeigen die Abhängigkeit der Hörerurteile von der Gesamtzusammensetzung des Stimulusmaterials (Abbildung 7). Da es in Experiment 1a sehr viele Stimuli gibt, die z.T. sehr lange Pausen zwischen den Äußerungsteilen enthalten, orientieren sich die Versuchspersonen an diesen extremen Trennungsgraden und setzen die kohäsiveren Realisierungen zu diesen stark separierten Äußerungen in Beziehung. So wandern die Urteile in Experiment 1b alle in Richtung stärkerer Separierung, da extrem separierte Stimuli in diesem Experiment fehlen. Die erhaltenen Urteile zu Kohäsion und Separierung sind also als relative Werte zu betrachten, die im Kontext der Experimente

eine Vergleichsmöglichkeit zwischen der Wahrnehmung verschiedener Merkmalsbündel bieten, nicht aber als absolute Messwerte auf natürliches Material übertragen werden können.

### **3.3.2.2.3 Diskussion zu Experiment 1a und 1b**

In allen Konturtripeln ohne Pausen in Experiment 1b gibt es einen hochsignifikanten Effekt des Konturverlaufs. Der Grad der wahrgenommenen Trennung steigt von der ebenen Kontur zur leicht fallenden Kontur und weiter zur tief fallenden Kontur. Dieser Effekt wird deutlich schwächer bei den Stimuli mit Pausen. Längung hat fast keinen systematischen Einfluss auf die wahrgenommene Separierung. Bei den Stimuli ohne Pausen lässt sich von Längungen mit Faktor 1 zu Längungen mit Faktor 1,4 bei den ebenen und leicht fallenden Konturen eine leichte Zunahme der Separierung feststellen. Bei den Stimuli mit Pausen ist keine Beziehung zwischen Längungsstufen und wahrgenommener Separierung zu erkennen. Es zeigt sich also ein leichter Maskierungseffekt der Pausen. Insgesamt sind jedoch die Interaktionen zwischen den getesteten Parametern weit geringer als erwartet.

Die Annahme, dass die Zusammensetzung des Stimulusmaterials in Experiment 1b im Vergleich zu Experiment 1a die separierende Wirkung unterschiedlicher Konturen deutlicher herausarbeiten kann, bestätigt sich, da der Effekt der Kontur in Experiment 1b hochsignifikant ist ( $F=22,4$ ,  $p=0,000$ ). Es muss allerdings beachtet werden, dass der Einfluss des melodischen Verlaufs auf die Urteile im Vergleich zu dem der Pausen nur gering ist (siehe Abbildung 6) und in den Experimenten 1a und 1c nicht signifikant wird. Trotzdem lässt sich Hypothese 3 durch das Experiment 1b prüfstatisch absichern.

In Experiment 1b ist der Effekt der Längung im Gegensatz zu Experiment 1a nicht signifikant ( $F=2,69$ ,  $p=0,095$ ). Die Ursache hierfür liegt wahrscheinlich darin, dass der stärkste Längungsfaktor (1,6) in Experiment 1a zu signifikanten Ergebnissen geführt hat und im Stimulusmaterial von Experiment 1b weggelassen wurde. Dieses Ergebnis legt nahe, die Längung des finalen Reims aus dem Klassifikationssystem für flüssige prosodische Grenzen herauszulassen. Dies gilt allerdings nur dann, wenn die Längung einen Faktor von etwa 1,6 nicht überschreitet, da stärkere Längungen des finalen Reims von Hörern als Hinweis auf eine unflüssige Sprechweise interpretiert werden. Das führt

im Rahmen der Klassifikation unterschiedlicher prosodischer Grenzen zu einer anderen funktionalen Einordnung (siehe 5.2).

### **3.3.2.4 Experiment 1c**

Experiment 1c erweitert die Fragestellung nach dem Zusammenhang zwischen der phonetischen Struktur prosodischer Grenzen und der wahrgenommenen Kohäsion zwischen Äußerungsteilen auf den Einfluss syntaktischer Strukturen. D.h. es wird untersucht, ob prosodische Grenzen an syntaktischen Grenzen anders beurteilt werden als prosodische Grenzen innerhalb syntaktischer Einheiten. Diese Fragestellung geht zurück auf Experimente von Boomer & Dittmann (1962) und Boomer (1965), in denen unterschiedliche Wahrnehmungsschwellen für Pausen in Abhängigkeit von ihrer Position in der syntaktischen Struktur festgestellt wurden. Pausen innerhalb syntaktischer Konstituenten wurden von den Versuchspersonen schon bei geringeren Pausendauern erkannt als Pausen am Ende syntaktischer Einheiten (siehe 3.3.1). Hieraus ergibt sich die Frage, ob nicht nur Pausen, sondern auch F0-Bewegungen und unterschiedliche segmentelle Dauerstrukturen bzw. Bündelungen dieser Merkmale in Abhängigkeit von ihrer Position in der syntaktischen Struktur unterschiedlich wahrgenommen werden.

Die zweite Erweiterung, die Experiment 1c vornimmt, ist die Untersuchung der wahrgenommenen Kohäsion in Abhängigkeit von Zögerungslängung am Phrasenende. Es werden in Experiment 1c Stimuli einbezogen, die eine so starke Längung des finalen Reims aufweisen, dass der Eindruck einer unflüssigen Sprechweise entsteht. Wie in Experiment 3a und 3b nachgewiesen werden kann, führt zunehmende Längung des phrasenfinalen Reims sukzessive zur Wahrnehmung einer unflüssigen Sprechweise (siehe 5.2). Besonders die Längung des finalen Reims mit Faktor 2,2 führt zu einer Beurteilung als stark unflüssig. Aus diesem Grund werden in das Stimulusmaterial auch Stimuli aufgenommen, deren finaler Reim mit den Faktoren 1,8 und 2,2 gelängt ist. Hierdurch soll die Wahrnehmung phonetischer Kohäsion bei Störungen im Sprechfluss untersucht werden.

### 3.3.2.4.1 Hypothesen

#### Hypothese 1

Der Grad der wahrgenommenen Separierung ist bei gleicher Merkmalsbündelung um so höher, je niedriger die syntaktische Grenze ist.

#### Hypothese 2

Leichte bis mittlere segmentelle Längung hat keinen deutlichen Einfluss auf die wahrgenommene Separierung. Starke segmentelle Längung führt zu deutlicher wahrgenommener Separierung.

### 3.3.2.4.2 Stimulusmaterial

Experiment 1c ist den Experimenten 1a und 1b ähnlich. Es wird wiederum die phonetische Struktur am Ende einer Teiläußerung systematisch variiert, indem die Parameter 'Längung', 'melodischer Verlauf' und 'Pause' verändert werden. Durch diese phonetische Variation der Merkmalsbündel am Ende des ersten Äußerungsteils entstehen 24 phonetische Varianten, die jeweils mit drei verschiedenen Äußerungsfortsetzungen gepaart werden. Je nach Art der Fortsetzung ist der syntaktische Einschnitt nach der ersten Teiläußerung unterschiedlich. Der erste Äußerungsteil ist für alle Paarungen *Ich hätte gern' 200 Gramm*. Die erste Fortsetzung ist *Marzipankugeln*, die zweite Fortsetzung *mit viel Nüssen* und die dritte Fortsetzung *Mehr ist schlecht für meine Diät*. Durch Paarung 1 (*Ich hätte gern' 200 Gramm - Marzipankugeln*) entsteht nach *Gramm* eine niedrige syntaktische Grenze innerhalb der Substantivgruppe *zweihundert Gramm Marzipankugeln*, zwischen der partitiven Apposition *zweihundert Gramm* und dem Bezugswort *Marzipankugeln*. Bei Paarung 2 (*Ich hätte gern' 200 Gramm - mit viel Nüssen*) entsteht eine höhere syntaktische Grenze zwischen *Gramm* und der Präpositionalphrase in attributiver Funktion *mit viel Nüssen*. Die dritte Paarung (*Ich hätte gern' 200 Gramm. - Mehr ist schlecht für meine Diät.*) führt nach *Gramm* zu einer noch höheren Grenze zwischen zwei Hauptsätzen.

Die Versuchspersonen beurteilen im Experiment für jede Paarung den Grad der wahrgenommenen Separierung zwischen den Äußerungsteilen. Dieser Aufbau bietet die

Möglichkeit zu prüfen, wie die Wahrnehmung identischer phonetischer Varianten durch den syntaktischen Kontext beeinflusst wird.

Das Stimulusmaterial wird vom Autor gesprochen und in einer schallgedämpften Aufnahmekabine mit 44,1 kHz digital aufgezeichnet. Jede der Äußerungsfortsetzungen wird zusammen mit dem Äußerungsbeginn, ohne eine intervenierende prosodische Grenze zwischen den Äußerungsteilen, gesprochen. Folgende Äußerungen werden aufgenommen:

- *Ich hätte gern' 200 Gramm Marzipankugeln.*
- *Ich hätte gern' 200 Gramm mit viel Nüssen.*
- *Ich hätte gern' 200 Gramm. Mehr ist schlecht für meine Diät.*

Bei der Weiterverarbeitung wird zuerst aus einer der Äußerungen die Sequenz *Ich hätte gern' 200 Gramm* herausgeschnitten und hinsichtlich der finalen Merkmalsbündel systematisch variiert. Dann wird aus den drei Äußerungen jeweils die zweite Teiläußerung isoliert und mit den Varianten des ersten Äußerungsteils gepaart. Folgende Kriterien gehen in die Auswahl des verwendeten Sprachmaterials ein:

- Durch das Zusammenfügen der Teiläußerungen sollen nach der ersten Teiläußerung unterschiedlich hohe syntaktische Grenzen entstehen.
- Die Schnittstelle zwischen dem Ende des ersten Äußerungsteils und dem Beginn des zweiten Äußerungsteils soll von der segmentellen Struktur her geeignet sein, die Teiläußerungen zusammenschneiden, ohne dass ein unnatürlicher Bruch an dieser Stelle entsteht, und sie soll ebenfalls geeignet sein, eine intervenierende Pause hineinzuschneiden, ohne dass dies unnatürlich klingt. Hierfür ist der Übergang zwischen den bilabialen Nasalen am Ende der ersten Teiläußerung und am Beginn aller drei Folgeäußerungen besonders gut geeignet, da aufgrund der Identität der finalen und initialen Laute keine deutlichen koartikulatorischen Veränderungen auftreten, die beim Schneiden an dieser Stelle unnatürlich klingen würden.
- Die Äußerungen sollen alle in einem vergleichbaren situativen Kontext geäußert werden können. Das macht es den Versuchspersonen vermutlich leichter, sich in die Aufgabenstellung hineinzusetzen. In diesem Fall könnte der situative Kontext ein Verkaufsgespräch an einem Stand mit Süßigkeiten sein.

### (1) Längung des finalen Reims

Die Dauer des finalen Reims von *Gramm* wird in vier Stufen vergrößert. Ausgehend von der Reimdauer der Originaläußerung ohne prosodische Grenze nach *Gramm* werden in PRAAT die Längungsfaktoren 1,4 und 1,8 und 2,2 angewendet. Es ergeben sich folgende Reimdauern:

- Keine Längung (Faktor 1, Originaläußerung), Reimdauer: 170 ms,
- Erste Längungsstufe (Faktor 1,4), Reimdauer: 240 ms,
- Zweite Längungsstufe (Faktor 1,8), Reimdauer: 305 ms,
- Dritte Längungsstufe (Faktor 2,2), Reimdauer: 375 ms.

Die erste Reimdauer (Faktor 1) kann als ungelängte Variante des finalen Reims angesehen werden, da in der Ausgangsäußerung, aus der sie entstammt, nach *Gramm* keine prosodische Grenze, sondern eine direkte Äußerungsfortsetzung folgt. Auch der auditive Eindruck deutet nicht auf finale Längung hin. Die zweite Reimdauer (Längungsstufe 1,4) entspricht der mittleren phrasenfinalen Längung, die in den Produktionsdaten an den etikettierten prosodischen Grenzen gemessen wurde. Die zwei höheren Längungsstufen (1,8 und 2,2) liegen deutlich oberhalb des Mittelwerts. Stimuli mit Längungen oberhalb des Faktors 1,4 verursachen zunehmend den Eindruck einer unflüssigen Sprechweise (siehe 5.2).

### (2) Verlauf des phrasenfinalen melodischen Musters

Die Ausgangsäußerung hat einen Satzakkzent auf *Zweihundert*, realisiert mit einem F0-Gipfel, der von 115 Hz auf 130 Hz ansteigt. Ausgehend von diesem Gipfel werden drei linear verlaufende melodische Verläufe erzeugt, die sich bis an das Ende von *Gramm* erstrecken. Die drei melodischen Muster sind mit denen der Experimente 1a und 1b vergleichbar. Es wird zum einen ein ebenes Muster generiert, das den Gipfel auf *Zweihundert* durch eine F0-Kontur ohne Einbuchtung mit dem Gipfel der zweiten Teiläußerung verbindet, der in allen drei Varianten bei 120 Hz liegt. Die Satzakkzentpositionen in den zweiten Teiläußerungen sind folgende: *Marzipankugeln, mit viel Nüssen, mehr ist schlecht für meine Diät*. Die zweite resynthetisierte Variante ist ein leicht fallendes Muster, das vom Gipfel auf *Zweihundert* bis zum Ende von *Gramm* eine

Differenz von 4 Halbtönen aufweist und bei 103 Hz endet. Die dritte Variante fällt über demselben Zeitraum um 8 Halbtöne und endet bei 82 Hz. Für die zweite und dritte Variante gilt, dass nach *Gramm*, also ab Beginn der zweiten Teiläußerungen, ein Aufstieg zur konstanten Gipfelposition bei 120 Hz folgt. Die zweite Teiläußerung endet in allen drei Varianten mit tief fallender Intonation (vgl. Abbildung 4).

### (3) Pausen

Durch die Kombinationen der Teiläußerungen und die Variationen der Parameter 'Längung' und 'Kontur' am Ende der ersten Teiläußerung entstehen 36 unterschiedliche Stimuli (4 Längungsstufen x 3 Konturen x 3 Äußerungsfortsetzungen). In jeden dieser Stimuli wird zwischen die erste und zweite Teiläußerung eine Pause von 210 ms geschritten. Diese Pausendauer liegt im Bereich der mittleren Pausendauer, die in den untersuchten Produktionsdaten gemessen wurde. Die Stimuli mit Pause werden separat abgespeichert. So entsteht neben den 36 Stimuli ohne Pause eine ansonsten identische Serie von 36 Stimuli mit interner Pause.

Der Schnittpunkt für das Einfügen der Pause wird in allen Fällen so gesetzt, dass der initiale Nasal der zweiten Teiläußerung vier Perioden lang ist. Da die Schnittstelle sich innerhalb eines langen bilabialen Nasals befindet, kann hier eine Pause hineingeschnitten werden, ohne dass koartikulatorische Abläufe durch die Pause unterbrochen werden und so ein unnatürlicher Gehörseindruck entsteht. Dies wird zusätzlich durch die geringe Intensität des Nasals begünstigt.



Audiobeispiele für die Stimuli aus Experiment 1c: 4.2

#### 3.3.2.4.3 Durchführung

Aus den 72 Stimuli wird eine randomisierte Abfolge generiert und wie in den Experimenten 1a und 1b präsentiert. Im Experiment 1c wird auf eine Stimuluswiederholung verzichtet, da die Experimente 1a und 1b bereits gezeigt haben, dass die Versuchspersonen bei Stimuluswiederholungen relativ konsistente Urteile abgeben. Die gesamte Serie ist 9 Minuten lang.

Vor der Präsentation und Beurteilung der Stimuli werden den Versuchspersonen die Instruktionen für das Experiment vorgespielt. Im Prinzip handelt es sich um dieselben

Anweisungen und Erläuterungen wie in den Experimenten 1a und 1b. Allerdings ist eine Passage hinzugefügt worden, in der auf die drei unterschiedlichen Sätze eingegangen wird, die durch das Zusammenfügen der ersten Teiläußerung mit den drei nachfolgenden Äußerungsfortsetzungen entstehen. Die Instruktionen befinden sich im Anhang.

Die Versuchspersonen beurteilen den Grad der Trennung zwischen den angrenzenden Äußerungsteilen wieder auf der 7-stufigen Skala. An dem Experiment nehmen 17 Versuchspersonen teil.

### 3.3.2.4.4 Ergebnisse

**Tabelle 14:** Ergebnisse der Varianzanalyse für Experiment 1c.

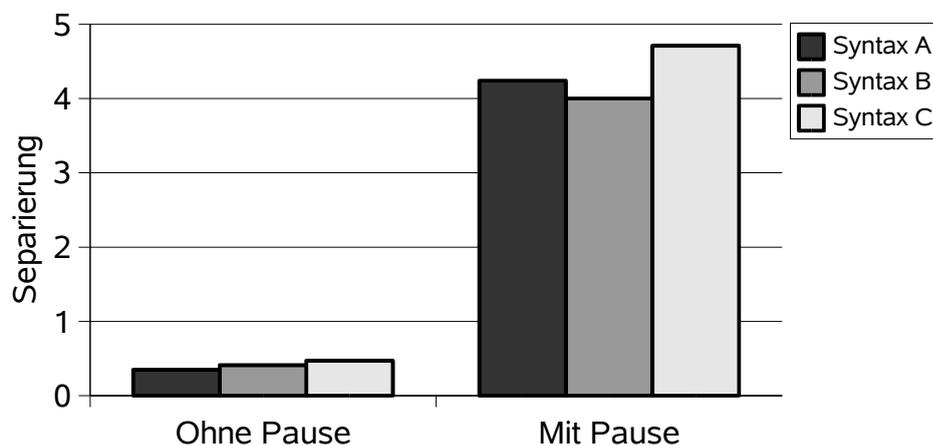
<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Syntax	1,91	0,03	0,976
Längung des finalen Reims	1,53	8,47	0,003
Finale Kontur	1,42	1,7	0,208
Pause	1	162,3	0,000
Syntax * Längung	4,69	2,71	0,029
Syntax * Kontur	3,26	1,49	0,227
Syntax * Pause	1,91	3,58	0,042
Längung * Kontur	3,95	0,82	0,519
Längung * Pause	1,9	29,16	0,000
Kontur * Pause	1,92	0,53	0,588
Syntax * Längung * Pause	3,8	0,72	0,577
Syntax * Längung * Kontur	5,79	2,6	0,024
Syntax * Pause * Kontur	3,26	0,88	0,464
Längung * Pause * Kontur	4,24	0,47	0,771
Syntax * Pause * Kontur * Längung	5,77	0,87	0,517

Die folgenden Abbildungen zeigen die Effekte der variierten phonetischen und syntaktischen Parameter auf die Urteile der Versuchspersonen. Der Einfluss der syntaktischen Struktur auf die wahrgenommene Separierung ist nicht signifikant ( $F=0,03$ ,  $p=0,976$ ; siehe Tabelle 14). Hypothese 1 wird deshalb verworfen.

Um den perzeptorischen Effekt der Einzelparameter möglichst isoliert zeigen zu können, ist jeder phonetische Parameter in einer separaten Abbildung dargestellt. Für die Darstellungen gilt, dass hier die Mittelwerte von Stimuli einbezogen werden, in denen die jeweils nicht relevanten Parameter die am wenigsten trennende Ausprägung haben. Das bedeutet, dass bei der Darstellung des Effekts der Pausen von Stimuli ohne finale Längung und mit einer ebenen Kontur am Phrasenende ausgegangen wird. Die abgebildeten Mittelwerte für die vier Längungsstufen kommen aus Stimuli ohne Pause und mit einem ebenen melodischen Verlauf am Phrasenende. Für die Darstellung der phrasenfinalen Konturen wird für alle melodischen Muster der phonetische Kontext ohne Pause und einer Längung mit Faktor 1,0 (keine Längung) gewählt.

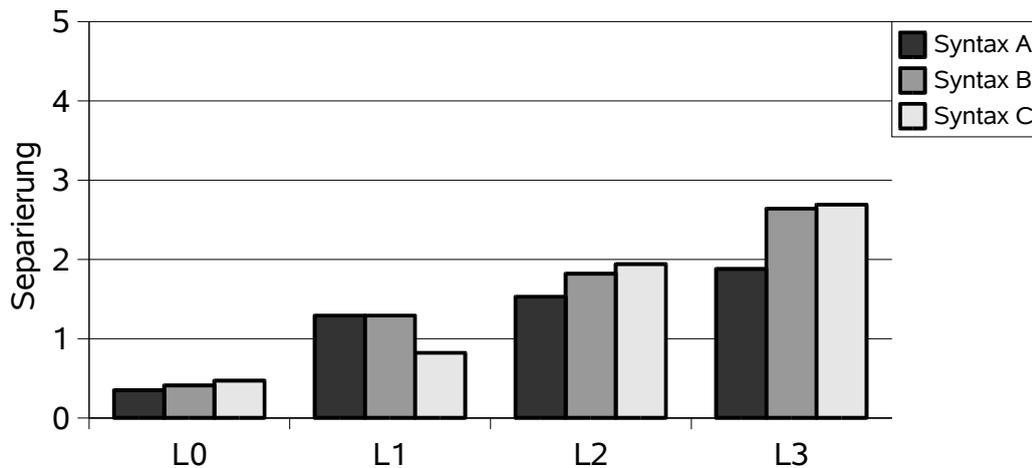
In den folgenden drei Abbildungen entspricht auf der y-Achse ein Wert von 0 einer maximalen Kohäsion zwischen den Äußerungsteile, ein Wert von 6 einer maximalen Trennung. Die syntaktischen Kontexte sind folgende:

- **Syntax A:** Ich hätte gern' 200 Gramm Marzipankugeln.
- **Syntax B:** Ich hätte gern' 200 Gramm mit viel Nüssen.
- **Syntax C:** Ich hätte gern' 200 Gramm. Mehr ist schlecht für meine Diät.



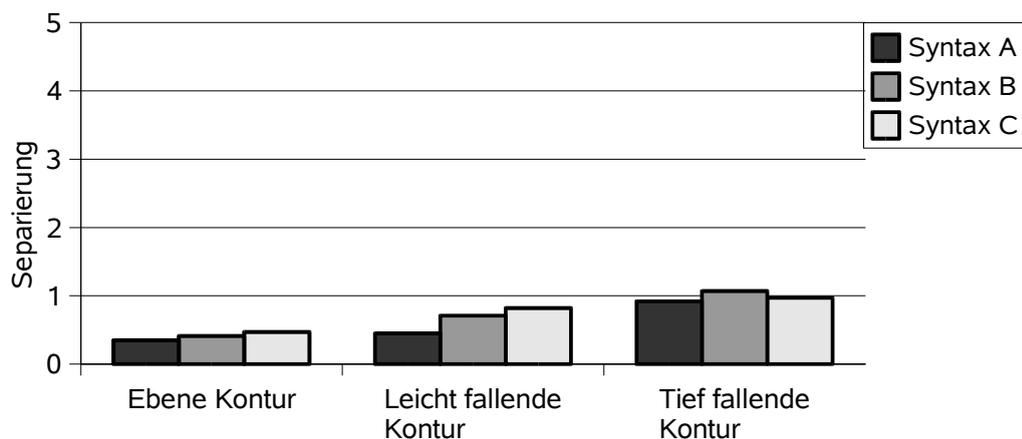
**Abbildung 8:** Wahrgenommene Separierung für Stimuli mit und ohne Pausen.

Abbildung 8 zeigt den wiederum hochsignifikanten Effekt der Pausen (siehe Tabelle 14).



**Abbildung 9:** Wahrgenommene Separierung für die vier Längungsstufen

Abbildung 9 zeigt die signifikante Auswirkung der vier Längungsstufen ( $F=8,47$ ,  $p=0,003$ , siehe Tabelle 14): Von Faktor 1 bis Faktor 2,2 ist ein kontinuierliches Ansteigen der wahrgenommenen Separierung zu beobachten. Die Unterschiede zwischen den Beurteilungen der vier Längungsstufen sind im Vergleich zu den Experimenten 1a und 1b deutlich. Der Grund hierfür ist, dass im Experiment 1c eine stärkere Variation der Längung des finalen Reims einbezogen wird und die einzelnen Längungsfaktoren hierbei weiter auseinander liegen als in 1a und 1b. Hypothese 2 wird verworfen; denn bei der gewählten Zusammensetzung des Stimulusmaterials zeigen alle Längungsstufen im Vergleich zur nächsthöheren Stufe eine deutliche Zunahmen der wahrgenommenen Separierung.



**Abbildung 10:** Wahrgenommene Separierung für die drei melodischen Muster

Die in Abbildung 10 gezeigte Auswirkung des Konturverlaufs ist gering und statistisch nicht signifikant (siehe Tabelle 14).

### 3.3.2.5 Diskussion der Ergebnisse aus den Experimenten 1a, 1b und 1c

Die Position der prosodischen zur syntaktischen Grenze hat keinen bedeutenden Einfluss auf die Urteile der Versuchspersonen. Die Versuchspersonen trennen also anscheinend zwischen der phonetischen Kohäsion und der syntaktischen Kohäsion und konzentrieren sich in ihrer Beurteilung auf die phonetischen Veränderungen in den Stimuli. Es zeigt sich vielmehr ein Trend in den Urteilen, der der Hypothese 1 gegenläufig ist. Die Urteile fallen an niedrigen syntaktischen Grenzen etwas niedriger aus als an hohen syntaktischen Grenzen. Eine enge syntaktische Bindung über die prosodische Grenze hinweg führt zu leicht geringerer wahrgenommener Separierung als dies bei einer identischen phonetischen Struktur und einer hohen syntaktischen Grenze an der prosodischen Grenze der Fall ist. Möglicherweise wird also unabhängig von der phonetischen Struktur auch die syntaktische Separierung beurteilt, die unter der Bedingung Syntax A geringer ist. Das Experiment kann also die Hypothese 1 und die Ergebnisse und Interpretationen von Boomer & Dittman (1962) und Boomer (1965) nicht stützen. Möglicherweise sind auch in deren Experimenten durch die Manipulation des Stimulusmaterials Artefakte aufgetreten, die bei der Interpretation der Ergebnisse nicht genügend Beachtung gefunden haben. Vor allem beim Herausschneiden von Pausen aus dem Stimulusmaterial sind vermutlich

präpausale Längung und melodische Merkmale erhalten geblieben, die weiterhin phonetische Zäsuren kodieren und so einen Einfluss auf die Urteile der Versuchspersonen gehabt haben können. Eine andere Erklärung für die Umkehrung des Effekts ist, dass die Versuchspersonen bei Boomer & Dittmann direkt nach dem Vorkommen einer Pause gefragt worden sind und nicht wie im vorliegenden Experiment nach der Stärke einer Zäsur.

Im Bereich der Perzeption der Kohäsion bzw. Separierung an flüssigen turn-internen prosodischen Grenzen spielt der Faktor Längung kaum eine Rolle, wie die Perzeptionsexperimente 1a und 1b zeigen. Dies ist überraschend, da phrasenfinale Längung in den Produktionsdaten eines der häufigsten Korrelate prosodischer Grenzen bildet und an zwei Dritteln aller untersuchten prosodischen Grenzen zu finden ist. Ferner liegt die Anzahl der prosodischen Grenzen, bei denen Längung das einzige automatisch ableitbare Merkmal ist, bei 8,9%. Es ist hierbei zu beachten, dass die automatische Merkmalsextraktion mögliche weitere Merkmale wie Intensitätsabfall und phonatorische Veränderungen an prosodischen Grenzen nicht erfasst. Insgesamt tritt Längung in beiden Korpora zusammen immerhin an 65,9% aller turninternen Grenzen auf. Die Daten deuten also darauf hin, dass hinsichtlich der finalen Längung die Mechanismen der Sprachproduktion und der Sprachperzeption auseinanderlaufen. In der Produktion ist Längung ein typisches Korrelat prosodischer Grenzen, während es in der Signalisierung flüssiger prosodischer Grenzen nur schwach funktionalisiert ist. Experiment 1c zeigt allerdings, dass sehr starke Längung des finalen Reims einen deutlichen Anstieg der wahrgenommenen Separierung hervorruft. Für diese starke Längung gilt aber, dass sie ebenfalls die Wahrnehmung einer unflüssigen Sprechweise hervorruft (siehe 5.2.1).

Auch der Effekt der melodischen Bewegung am Phrasenende ist bei der für die Experimente 1a, 1b und 1c gewählten Art der Befragung nur schwach ausgeprägt und bleibt weit hinter dem Einfluss von Pausen auf die Urteile zurück. Wie aber die interpretative Analyse im folgenden Kapitel und die Perzeptionsexperimente im Kapitel zu den turnfinalen prosodischen Grenzen zeigen, gibt es vor allem für die tief fallende Kontur eine klare funktionale Bindung an den Abschluss größerer Sinneinheiten und auch an den Abschluss von Gesprächsbeiträgen. Diese Ergebnisse deuten darauf hin, dass die Skala von Kohäsion zu Separierung diese funktionalen Differenzierungen durch Sprechmelodie nicht adäquat erfassen kann bzw. die Experimente 1a, b und c nicht geeignet sind,

um die funktionalen Unterschiede perzeptorisch zu validieren. Wenn starke melodische Bewegungen nicht als phonetisch separierend perzipiert werden, die tief fallende Kontur aber klar in Verbindung mit Terminalität steht, muss davon ausgegangen werden, dass es im Bereich der Sprechmelodie funktionalisierte melodische Muster gibt, die im phonologischen System verankert sind und als inhaltlich separierend rezipiert werden, ohne eine starke phonetische Trennung zu verursachen. Das phonetische Perzept der Separierung wird in diesen Fällen wohl meistens durch zusätzliche Merkmale unterstützt, denn nur 2,4% der prosodischen Grenzen in beiden Korpora weisen eine starke melodische Bewegung als einziges Merkmal auf.

Diese Befunde belegen, dass das vorrangige phonetische Prinzip bei der Signalisierung prosodischer Grenzen die Bündelung von Merkmalen ist. Die Bündel steuern sowohl die Wahrnehmung von Einschnitten als auch deren sprachliche Funktion. Pausen haben hierbei den stärksten Einfluss auf die wahrgenommene Separierung, Konturen steuern vor allem die funktionale Interpretation im Dialogkontext. Reset und Längung können als begleitende phonetische Phänomene bei der Signalisierung prosodischer Grenzen, nicht aber als eigenständige Merkmale angesehen werden.

### **3.4 Interpretative Analysen**

Die Beschreibung prosodischer Grenzen wird in diesem Abschnitt durch inhaltlich-pragmatische und syntaktische Klassifikationen ergänzt. Das Ziel hierbei ist, Verbindungen zwischen unterschiedlichen phonetischen Ausprägungen turninterner prosodischer Grenzen und der inhaltlichen, pragmatischen und syntaktischen Struktur von Äußerungen empirisch zu belegen. Als zugrunde liegende Annahme gilt, dass hohe inhaltlich-pragmatische Grenzen innerhalb von Redebeiträgen durch stark separierende Merkmalsbündel repräsentiert werden, während die phonetische Separierung an prosodischen Grenzen innerhalb inhaltlicher Blöcke schwächer ist bzw. durch weniger stark separierende Merkmale hervorgerufen wird. Die Anforderungen und Richtlinien nach denen die interpretative Analyse vorgenommen wird, werden nach einem kurzen Literaturüberblick erläutert.

### 3.4.1 Stand der Forschung

In der Literatur findet sich vor allem eine Forschungsrichtung, die interpretativ Verbindungen zwischen den phonetischen Ausprägungen und inhaltlichen und pragmatischen Eigenschaften von Äußerungen und Gesprächen herstellt: die *Conversation Analysis* (dt.: *Konversations-* oder *Gesprächsanalyse*). Obwohl sich die in dieser Arbeit erstellten und angewendeten Kriterien für die Interpretation von Gesprächen erheblich von denen der Conversation Analysis (CA) unterscheiden, sollen trotzdem kurz die methodischen Prinzipien beschrieben werden, die diese Forschungsrichtung verfolgt, denn aus dem Umfeld der Conversation Analysis stammen wichtige Forschungsergebnisse hinsichtlich der phonetischen Struktur natürlicher Gespräche. Folglich sind auch viele Anregungen sowohl methodischer als auch inhaltlicher Art in diese Arbeit eingegangen. Folgende Schriften sind grundlegend für das Forschungsparadigma der CA im Bereich linguistischer und phonetischer Forschung: Sacks (1963); Sacks, Schegloff & Jefferson (1974); Wilson (1978); Gumperz (1982). Für einen Überblick über die Konversationsanalyse vgl. Levinson (1983) und Bergmann (1988).

Ursprünglich stammt die CA aus einem relativ engen soziologischen Wirkungsbereich, fand aber Anfang der 1970er-Jahre großen Anklang bei einigen Nachbardisziplinen, so auch der Linguistik und später der Phonetik. Die CA widmet sich den formalen Strukturen, Prinzipien und Mechanismen von Alltagshandlungen und Gesprächen. Soziale Interaktion wird begriffen als ein fortwährender interaktiver Prozess der Hervorbringung und Absicherung kommunikativer Elemente. Es geht nicht um das Gesagte an sich, sondern um die Entstehung und Interpretation sprachlicher Handlungen im jeweiligen situativen Kontext. Hierbei geht die CA von folgenden zentralen Annahmen aus:

- Interaktionen liegen Ordnungssysteme zugrunde. Menschen setzen immer bestimmte Techniken und Verfahren ein, um ihre Handlungen für andere erkennbar, verstehbar und erklärbar zu machen.

- Die Handelnden versuchen,
  - den Kontext und die Situation der Handlung zu analysieren,
  - die Äußerungen der Handlungspartner zu interpretieren,
  - die situative Angemessenheit, Verständlichkeit und Wirksamkeit der eigenen Äußerungen herzustellen,
  - das eigene Tun mit dem der Anderen zu koordinieren.
- Eine Äußerung (Handlung) muss zu ihrem Verständnis in dem Zusammenhang betrachtet werden, in dem sie entstand.

Für die CA umfasst jeder Moment einer Interaktion viele begleitende Faktoren, welche als Interaktionskontext relevant sein könnten. Aufgabe der CA ist es, spezifische Faktoren zu finden und als Handlungskontext zu begreifen, der für den jeweils Handelnden handlungsrelevant ist. Die Handelnden werden als kontextsensitive Akteure gesehen, welche den Interaktionskontext permanent analysieren, interpretieren, sich in ihrer Handlung auf den Kontext einstellen und sich stets wechselseitig ihre Kontextorientierungen anzeigen. Von besonderer Bedeutung ist für die CA das Prinzip des sequentiellen Kontexts. Dies besagt, dass jede Äußerung zu einer Kontextveränderung, zu einer Abweichung vom Bisherigen, führt und somit einen neuen Kontext für alles Folgende darstellt. Die Sequenzanalyse ist eine typische Form der CA-Kontextanalyse.

Ziel der Methode ist es zunächst, das registrierte soziale Geschehen (z.B. einen Dialog) möglichst verlustarm zu bewahren. Zur Konservierung des Geschehens dienen daher als erster Schritt der Analyse auditive und/oder visuelle Aufzeichnungen realer Interaktionen. Viele Vertreter der CA fordern, nur Datenmaterial zu verwenden, welches auch ohne Aufzeichnung in dieser Form entstanden wäre. Der zweite Schritt ist meist eine detaillierte Transkription des Datenmaterials. Jedes transkribierte Textelement wird als nicht-zufällig angesehen und als Teil der interaktionalen Ordnung betrachtet. Als nächstes versucht die CA, für ein beobachtbares, gleichförmiges Phänomen (z.B. bestimmte Formulierungen oder phonetische Details) generative Prinzipien zu rekonstruieren, die den Handelnden als Lösung für interaktive Probleme dienen und deren Einsatz die beobachtbare Geordnetheit eines Interaktionsgeschehens hervorbringt. Diese Prinzipien werden *the machinery* genannt (Sacks et al. 1974).

Bei dieser Vorgehensweise ist der Sozialforscher auf seine Intuition und Kompetenz als Gesellschaftsmitglied angewiesen, muss aber auch versuchen, sein intuitives Verständnis mit einer definierten methodischen Vorgehensweise zu verknüpfen, um eine sinnhafte und konsistente Interpretation des Handlungsgeschehens möglich zu machen. Im Zentrum der frühen linguistisch ausgerichteten Arbeiten steht die Analyse natürlicher Konversation, allerdings sind diese Arbeiten eher von linguistischen Strukturüberlegungen geprägt als von phonetischen Beobachtungen. Mitte der 80er-Jahre erscheinen diverse Arbeiten, die einen stark phonetisch ausgerichteten Zweig der Conversation Analysis etablieren und die das Ziel verfolgen, die Produktions- und Interpretationsleistungen von interagierenden Gesprächspartnern zu beschreiben, mit denen die Interaktion hergestellt und koordiniert wird. John Local ist ein Vertreter der phonetisch ausgerichteten Conversation Analysis und hat mit seiner wissenschaftlichen Ausrichtung starken Einfluss auf das methodische Vorgehen vieler Linguisten und Phonetiker ausgeübt (z.B. Selting, Auer, Ogden). Local (2000-2002) formuliert die Grundsätze phonetischer Studien im Rahmen der CA folgendermaßen: Das zentrale Ziel der Conversation Analysis (CA) ist die Erforschung der Kompetenzen von Sprechern und Hörern im Dialog, bezüglich der Produktion, der Rezeption und der Koordination interaktionalen Verhaltens. Die CA untersucht hierfür die phonetischen Ressourcen, die zum Einsatz kommen, um die Interaktion in Gesprächen zu formen. Zentrale Untersuchungsbereiche sind die Strukturierung von Dialogen in Turns, die Korrektur von Missverständnissen und die Signalisierung von Korrekturen und Fehlern. Das Ziel der Analyse ist es, spezielle Bündelungen phonetischer Merkmale (Cluster) in einen funktionalen Zusammenhang mit den genannten Untersuchungsbereichen zu stellen und hierdurch eine (sprachspezifische) Phonologie der Konversation zu erarbeiten. Diese Phonologie soll Verbindung zu morphosyntaktischen Strukturen schaffen und die Beziehung der Merkmalsbündel mit der Turnstruktur herstellen. Die phonetischen Merkmale, die in die Bündelungen eingehen, sind Rhythmus, Dauerstruktur, Tempo, Lautheit, Tonhöhenverläufe, Stimmqualität und artikulatorische Strukturen - also praktisch alle Mittel, die auf der phonetischen Seite von Lautsprache zur Verfügung stehen.

Als Datengrundlage soll nur Sprachmaterial in die Analysen eingehen, das aus natürlich aufgetretener Konversation stammt (z.B. heimliche Telefonmitschnitte). Diese Daten werden einem signal- und ohrenphonetischen Analyseprozess unterzogen, der

interaktionale Aktivitäten herausarbeitet und Verbindungen mit phonetischen Parametern offenlegt. Hierbei gelten folgende Maximen:

- Jedes Detail kann für den Kommunikationsprozess wichtig sein, kein Detail darf a priori ausgeschlossen werden.
- Es sollen so wenig wie möglich Vorhersagen über die phonetischen Parameter a priori gemacht werden.
- Es muss sich in den Daten zeigen, dass die analysierten Kategorien für die Sprecher/Hörer relevant sind. Das analysierte Verhalten muss also in den Daten beobachtbar sein.

Diese Maximen erfordern eine feine phonetische Analyse und kontextuelle Interpretation vieler Fälle jeder beobachteten Kategorie. Die Interpretation folgt der sequentiellen Sichtweise, indem jede sprachliche Handlung als eine Folge der vorherigen Handlung angesehen wird, die sich wiederum auf die nachfolgende Handlung bezieht. Jede sprachliche Handlung ist also in den Kommunikationsprozess eingebettet. Local geht davon aus, dass Produzenten und Rezipienten durch ihr sprachliches Verhalten erkennbar werden lassen, wie sie den Kommunikationsprozess auffassen und dass diese Erkennbarkeit den Untersucher davon befreit, sich auf die eigene Intuition verlassen zu müssen. So kann die Aktivität eines Sprechers als konversationelle Frage interpretiert werden, wenn der Rezipient im Gespräch eine relevante Antwort auf die potentielle Frage gibt oder erklärt, warum er diese Antwort nicht abgibt.

Hier wird auch das Konzept der Kontextualisierung relevant (Cook & Gumperz 1976; Auer 1986; Selting 1995), das angewendet wird um die Rahmenbedingungen für die sequenziell ablaufenden interaktionalen Elemente zu analysieren. Es wird davon ausgegangen, dass die Gesprächspartner in einer emotionalen und sozialen Position zueinander stehen (z.B. Arzt/Patient) und dass sie über gemeinsames Hintergrundwissen verfügen. Der eigentliche Kontext der Kommunikation wird aber von den Interaktionspartnern durch die Art und Weise ihrer Interaktion fortschreitend selber konstruiert. Auer (1986:23):

*Im Gegensatz zu dieser Vorstellung von Kontext als eine Determinante gehen nun Cook & Gumperz 1976 von einem interaktiven Interaktionsteilnehmer aus, der nicht nur auf Kontext reagiert, sondern auch Kontext aufbaut: Sprecher bilden nicht nur Sätze, um ('referentielle') Bedeutung oder Informationen zu übermitteln, sie stellen ihre Äußerungen zugleich in einen Kontext und ermöglichen so dem Referenten das Verstehen. Anders gesagt: für die Interaktionsteilnehmer besteht die Aufgabe darin, (sprachliche) Handlungen auszuführen und zugleich interpretierbar zu machen, indem ein Kontext konstruiert wird, in den sie sich einbetten.*

Zur Konstruktion von Kontexten verwenden die Interaktionspartner die sog. Kontextualisierungshinweise - linguistische, paralinguistische und phonetische Strukturen - mit denen sie sich wechselseitig die für die Interaktion notwendigen Interpretationsrahmen verfügbar machen bzw. nahe legen. Eingebettet in diesen kontextuellen Rahmen funktionieren alle anderen Regelwerke, wie z.B. das Turntaking-System als formale Apparate, die lokal von den Interaktionspartnern verwendet und auf die spezifische Situation zugeschnitten werden (Sacks et al. 1974). Die funktionale Analyse der Regelwerke muss folglich immer über die Kontextualisierung - die Analyse des umgebenden Kontextes und des *settings* des Gesprächs - erfasst werden und darf nicht aus einzelnen Sätzen oder konstruierten Dialogen abgeleitet werden.

Einige Forschungsergebnisse aus Arbeiten der CA, die mit der vorliegenden Arbeit in direktem Zusammenhang stehen, finden sich in den Literaturüberblicken der Kapitel zu flüssigen, unflüssigen und turnfinalen prosodischen Grenzen.

Die Methoden der CA sind aus verschiedenen Gründen angreifbar und für die Dateninterpretation in dieser Arbeit nicht anwendbar. Vor allem der Hinweis von Local (2000-2002), dass Produzenten und Rezipienten durch ihr sprachliches Verhalten erkennen lassen, wie sie den Kommunikationsprozess auffassen und dass diese Erkennbarkeit den Untersucher davon befreit, sich auf die eigene Intuition verlassen zu müssen, ist sehr problematisch und spiegelt eine Objektivität des Untersuchungsverfahrens vor, die m.E. nicht gegeben ist. Denn auch wenn die sprachlichen Reaktionen

innerhalb einer Sequenz von Redebeiträgen betrachtet werden, ist deren Beurteilung durch einen außenstehenden Bearbeiter immer auch vom Wissen und der Intuition desselben abhängig; auch die Antwortreaktion eines beobachteten Hörers im Gespräch erklärt nicht aus sich heraus die Absichten und Ziele des Sprechers. So kann z.B. ein Hörer aus verschiedenen Gründen nicht mit einer Antwort reagieren, obwohl im Vorangehenden eine Frage gestellt wurde. Der Hörer kann sich z.B. um eine Antwort drücken, wenn ihm diese unangenehm ist, und die Frage einfach übergehen. Dieses Verhalten ist zwar u.U. auch im Rahmen einer Interpretation herauszuarbeiten, aber wohl nicht im Rahmen eines wirklich objektiven Verfahrens. Hierfür müssten Phonetiker und Linguisten tatsächlichen Einblick in die Gedanken und Planungsprozesse von Sprecher und Hörer haben, als dessen Abbild das Sprachsignal bei weitem nicht ausreicht.

Neben dieser allgemeinen Kritik am methodischen Vorgehen der CA gibt es auch zwei Punkte im Zusammenhang mit den Anforderungen der vorliegenden Arbeit, weshalb hier ein anderes methodisches Verfahren für die Interpretation von Korpusdaten entwickelt und angewendet werden muss. Es ist das Ziel dieser Arbeit, interpretative inhaltliche und pragmatische Kategorien mit den automatisch abgeleiteten Merkmalsbündeln der statistischen Auswertung des Korpus zu korrelieren. Das hat verschiedene Konsequenzen, die mit den Maximen der CA nicht zu vereinbaren sind. Schon in der Auswahl des Datenmaterials ist es hier nicht möglich, Sprachdaten zu verwenden, die aus vollkommen natürlichen Gesprächen stammen, denn für die Aufbereitung der Daten (speziell der segmentellen Etikettierung) ist eine hohe akustische Qualität der Daten und eine Kanaltrennung zwischen den Sprechern notwendig. Es ist derzeit kein Verfahren bekannt, das einen von den Sprechern unbemerkten Mitschnitt von Gesprächen unter diesen Bedingungen ermöglicht.

Weitere Unvereinbarkeiten liegen in dem methodischen Ansatz der CA, alle phonetischen Details in die Analysen einzubeziehen und keinerlei Vorhersagen über die phonetische Struktur in die Analysen einzubeziehen. Dies ist eine gute Herangehensweise, um sich mit weitgehend unbekanntem Phänomenen in einem nicht aufbereiteten Datenkorpus auseinanderzusetzen und so einen Phänomenbereich einzukreisen, nach phonetischen Exponenten übergeordneter linguistischer Muster zu suchen und diese an einer Reihe einzelner Beispiele zu exemplifizieren. In der Bearbeitung der aufbereiteten LINDENSTRASSE-Dialoge geht es aber darum, hypothesengesteuert über eine große Zahl von

Fällen statistische Aussagen über die Zusammenhänge zwischen phonetischen und inhaltlich-pragmatischen Strukturen zu treffen. Hierfür ist es notwendig, inhaltliche, pragmatische und syntaktische Muster zu erfassen und in einem klar strukturierten Kategoriensystem festzuhalten, so dass später eine Beziehung zwischen den phonetischen Merkmalen und den interpretativen Kategorien festgestellt werden kann. Da die phonetischen Analysen bereits vor dem Beginn der Interpretation automatisch vorgenommen wurden, können sie in der Interpretation der Daten nicht explizit herangezogen werden, weil so Zirkularität bei der Prüfung der Hypothesen entstünde. Der Ansatz, jedes phonetische Detail in die Analyse einzubeziehen, birgt ein weiteres Problem. Durch die entstehende Komplexität der phonetischen Beschreibungen gelangen die Autoren kaum zu einer Systematisierung und funktionalen Zuordnung phonetischer Parameter und bleiben in Einzelfallbeschreibungen verhaftet. Hierbei werden oft bestimmte Beispiele zur Exemplifizierung eines Phänomens herangezogen und (über-)generalisiert, wobei ebenfalls vorhandene Gegenbeispiele nicht ins Blickfeld geraten.

Ein ganz anderes Verfahren für die Bewertung des inhaltlichen Aufbaus von Dialogen kommt in einer Untersuchung schwedischer Lesesprache und *Map Task*-Daten zur Anwendung (Megyesi & Gustavson-Capkova 2002). Diese Untersuchung versucht einen Zusammenhang zwischen Vorkommen und Dauer von Pausen und deren Beziehung zu inhaltlichen Grenzen zu belegen. Das angewendete Verfahren entspricht in einigen Punkten den inhaltlichen Zielen und den methodischen Anforderungen der Analyse meiner Datenbasis, denn auch hier werden die Ergebnisse einer automatischen Merkmalsdetektion in Verbindung mit übergeordneten linguistischen Strukturen gebracht. Die Ergebnisse von Megyesi und Gustavson-Chapkova sind im Abschnitt 3.3.1 beschrieben. Hier soll nur die verwendete Methode erläutert werden: Im ersten Schritt findet eine automatische Messung von Pausenpositionen und -dauern im Korpus statt. Im zweiten Schritt wird überprüft, ob die Pausen an Positionen vorkommen, an denen eine thematische Weiterführung (*theme continuation*) oder ein thematischer Wechsel (*theme shift*) vorliegt. Hierfür ist es notwendig Aussagen über die Kategorien *theme shift* und *theme continuation* an den Positionen zu treffen, wo Pausen detektiert und gemessen wurden. Diese Beurteilung wird im Rahmen eines Experiments durch fünf Versuchspersonen getroffen, die an jeder Pausenposition über das Auftreten eines thematischen Wechsels urteilen. Die Beurteilung findet ausschließlich auf der Basis der Verschriftung

der *Map Task*-Dialoge, bzw. der schriftlichen Vorlage bei der Lesesprache statt. So soll eine Beeinflussung der Urteile durch phonetische Parameter verhindert werden. In der Auswertung der Urteile werden vier interpretative Kategorien aufgestellt:

- *No boundary* wenn keine der VP an der Position ein Urteil für *theme shift* abgibt.
- *Weak boundary* wenn eine oder zwei Versuchspersonen für *theme shift* urteilen
- *Strong boundary* wenn drei oder vier Versuchspersonen für *theme shift* urteilen
- *Extra strong boundary* wenn alle fünf Versuchspersonen für *theme shift* urteilen

Dieses Verfahren liefert unabhängig von der phonetischen Struktur des untersuchten Materials eine grobe Einschätzung des inhaltlichen Aufbaus von Äußerungen. Die herausgearbeiteten Kategorien von 'No boundary' bis 'Extra strong boundary' können durch nachfolgende phonetische Analysen ohne die Gefahr zirkulären Vorgehens mit verschiedenen phonetischen Merkmalen in Verbindung gebracht werden. Das Verfahren hat allerdings den Nachteil, dass die Klassifikation nur hinsichtlich eines Parameters vorgenommen wird und dass die von den Versuchspersonen angewendeten Kriterien der Beurteilung unbekannt sind und sich von Person zu Person unterscheiden können.

### **3.4.2 Ergebnisse der Interpretation von Korpusbeispielen**

Weder die labelbasierte Korpusanalyse noch die Perzeptionsexperimente geben Auskunft über die Beziehung zwischen der semantisch-pragmatischen Struktur komplexer Redebeiträge und deren phonetischer Manifestation durch unterschiedliche prosodische Grenzen. Um Erkenntnisse in diesem Bereich zu gewinnen, wird eine große Anzahl von Redebeiträgen einem interpretativen Verfahren unterzogen, in dem unter anderem beurteilt wird, ob sich an einer vorgefundenen prosodischen Grenze ein starker inhaltlicher Einschnitt, ein schwacher Einschnitt oder gar kein inhaltlicher Einschnitt befindet. Diese grobe Klassifikation dient bei der Validierung des Klassifikationssystems prosodischer Grenzen dazu, Divergenzen und Konvergenzen zwischen inhaltlicher und phonetischer Separierung zu belegen und das Klassifikationssystem entsprechend zu optimieren.

Folgende Kriterien gelten bei der interpretativen Beurteilung der inhaltlichen Struktur. Auf der niedrigsten Stufe inhaltlicher Separierung (IB-, 'keine inhaltliche Blockgrenze') kommt es am Phrasenende nicht zum Abschluss einer inhaltlichen Einheit. Die Phrase ist kein mögliches geplantes Äußerungsende, sondern ist auf inhaltliche Komplettierung ausgerichtet. Die größte vergebene Grenze (IB+) wird als Abschluss eines inhaltlichen Blocks bezeichnet. Hier ist der Abschluss einer inhaltlichen Einheit zu beobachten. Eine inhaltliche Einheit kann als abgeschlossen gelten, wenn sich die Funktion eines Phrasenblocks im Dialog (z.B. Aussage vs. Frage) ändert oder dem Gesprächsbeitrag eine inhaltliche Wendung durch neue Information gegeben wird. Auch der Abschluss einer pragmatischen Einheit, etwa einer Begrüßung, wird als hohe Grenze innerhalb von Äußerungen angesehen. IB+ repräsentiert also einen hohen Grad inhaltlicher Separierung. Die mittlere der drei Kategorien (IB0) ist eine negativ definierte Kategorie, die Fälle beinhaltet, welche weder IB- noch IB+ zugeordnet werden können. Dies betrifft eine in spontaner Sprache recht große Anzahl inhaltlicher Einschnitte.

Hinter dieser Einteilung steht das Problem, dass die inhaltliche Struktur spontaner Sprache oft nur sehr schwer erfassbar und klassifizierbar ist. Die vorgenommene Klassifikation hat den Vorteil, dass die Kategorien IB- und IB+ klar voneinander abgegrenzt sind und in den Untersuchungen zum Klassifikationssystem prosodischer Grenzen zumindest über diese beiden extremen Ausprägungen klare Aussagen getroffen werden können. Die Fälle in der mittleren Kategorie IB0 bilden ein Kontinuum zwischen den Extremen der fehlenden inhaltlichen Separierung (IB-) und dem Ende von inhaltlichen Blöcken (IB+).



Audiobeispiele: hohe inhaltliche Grenze: 1.3, niedrige inhaltliche Grenze: 1.7.

In der Literatur wird die Stärke inhaltlicher Einschnitte seit langem diskutiert ohne, dass eindeutige Kriterien gefunden wurden. Folgende Terminologien werden verwendet, um die beiden Pole im Grad der inhaltlichen und phonetischen Verbindung zwischen Äußerungsteilen zu bezeichnen: Weiterweisung/Abschluss, Progredienz/Terminalität, theme-continuation/theme-shift, Kohäsion/Separierung, new topic/old topic (v. Essen 1964; Couper-Kuhlen 1983; Local & Kelly 1986; Local et al. 1986; Selting 1995; Caspers 2000; Gilles 2000; Shriberg et al. 2000; Wichmann & Caspers 2001; Local 2000-2002; Megyesi & Gustafson-Capkova 2002; Chen 2003).

Bei der Zuordnung der prosodischen Grenzen wird sowohl das akustische Signal als auch die orthographische Verschriftung des jeweiligen Gesprächsabschnitts herangezogen. In die interpretative Analyse gehen 60 längere Gesprächsbeiträge von verschiedenen Sprechern aus dem Terminabsprachekorpus ein. Bei der Auswahl der Gesprächsbeiträge wird darauf geachtet, dass die Beiträge mindestens eine hohe inhaltliche Grenze enthalten und mit einer weitgehend flüssigen Sprechweise produziert sind. So entsteht für jede etikettierte prosodische Grenze in den interpretierten Gesprächsabschnitten eine Klassifikation hinsichtlich der Stärke der inhaltlichen Grenze. Diese Klassen inhaltlicher Kohäsion bzw. Separierung werden in der Auswertung der Daten auf die jeweiligen Merkmalskombinationen an den prosodischen Grenzen bezogen.

Die bisherigen Untersuchungen zeigen einen klaren Zusammenhang zwischen dem Auftreten von Pausen, der wahrgenommenen Separierung zwischen benachbarten prosodischen Phrasen und der inhaltlichen Struktur von Äußerungen.

Die Perzeptionsexperimente 1a, 1b und 1c (3.3.2) belegen, dass Pausen zu einer stärkeren perzeptorischen Separierung führen als alle anderen untersuchten Merkmale. Eine Erhöhung der Pausendauern in 7 Schritten von 50 auf 890 ms führte im Perzeptionsexperiment in jedem Fall zu einer hochsignifikanten Zunahme der wahrgenommenen Separierung. Aus diesen Ergebnissen lässt sich keine Kategoriengrenze zwischen einer kurzen und einer langen Pause ableiten. Auch die Untersuchung der Pausendauern im gesamten Korpus deutet nicht auf separate Kategorien hin (z.B. durch eine Bimodalität der gemessenen Pausendauern).

Die Auswertung der interpretativen Beurteilungen der 60 Gesprächsbeiträge aus dem Terminabsprachekorpus belegt einen Zusammenhang zwischen dem Auftreten von Pausen und der inhaltlichen Struktur von Äußerungen. Folgende Zahlen zeigen, dass mit zunehmender Stärke inhaltlicher Einschnitte das Vorkommen von Pausen stark ansteigt.

- IB-: n=30 davon mit Pause/Atmen: 8 (26,6%),
- IB0: n=85 davon mit Pause/Atmen: 34 (40,0%),
- IB+: n= 64 davon mit Pause/Atmen: 44 (68,8%).

Pausen verhalten sich somit entsprechend der Hypothese, dass wahrgenommene phonetische Separierung mit inhaltlicher Separierung einhergeht.

Weitere Untersuchungsschritte ziehen auch die Dauer der Pausen an verschiedenen hohen inhaltlichen Einschnitten in Betracht und belegen eine Tendenz zu größeren Pausendauern an starken inhaltlichen Einschnitten. Diese Tendenz ist allerdings so schwach ausgeprägt, dass sie keine Bedeutung für die Neukonzeption eines funktionsorientierten Klassifikationssystems hat. Ebenso spricht dieser Befund (wie auch die Perzeptionsexperimente 1a, 1b und 1c) gegen einen festen Pausengrenzwert und die darauf basierenden Einteilung in lange vs. kurze Pausen.

Die Untersuchung der Merkmalsbündel zeigt weiterhin, dass das Auftreten von Pausen fast immer an andere Merkmale gebunden ist. Nur bei 3,8% aller Pausen in beiden Korpora (Terminabsprachen und LINDENSTRASSE) können keine weiteren Merkmale automatisch extrahiert werden.

### **3.4.2.1 Starke tonale Bewegungen**

Die Interpretation der Korpusdaten belegt einen klaren Zusammenhang zwischen dem Auftreten starker tonaler Bewegungen am Phrasenende und hoher inhaltlicher Grenze und somit die Signalwirkung der Intonationsmuster bei der phonetischen Markierung von inhaltlichem Abschluss vs. inhaltlicher Kontinuität, wie folgende Verteilung zeigt.

- IB-: n=30 davon mit starker tonaler Bewegung: 6 (20%),
- IB0: n=85 davon mit starker tonaler Bewegung: 39 (45,9%),
- IB+: n=64 davon mit starker tonaler Bewegung: 48 (75%).

Unter den Konturen mit starker tonaler Bewegung finden sich vor allem melodische Muster, die bis an die untere Grenzen der Sprechstimme fallen (terminale Intonationsmuster). Besonders diese Kontur signalisiert meist das Ende einer inhaltlichen Einheit. Wie die Perzeptionstests zeigen, wirken starke tonale Bewegungen allerdings in deutlich geringerem Maße perzeptorisch separierend als Pausen. Insofern ist der Zusammenhang zwischen phonetischer und funktionaler Separierung im Bereich der Sprechmelodie weniger eindeutig als bei den Pausen.

### 3.4.2.2 F0-Reset

F0-Reset hat im Gegensatz zu den anderen untersuchten Merkmalen, eine andere Struktur. Reset ist eine Eigenschaft der nachfolgenden Phrase, während alle übrigen Merkmale in den Merkmalsbündeln Eigenschaften entweder der vorangehenden Phrase (Längung, Kontur) oder der Fuge (Pause/Atmen) zwischen den Phrasen beschreiben. Die Verteilung von prosodischen Grenzen mit und ohne Reset bezogen auf die inhaltlichen Kategorien zeigt folgendes Bild:

- IB-: n=30 davon mit Reset : 20 (66,7%),
- IB0: n=85 davon mit Reset : 51 (60,0%),
- IB+: n=64 davon mit Reset : 48 (75,0%).

Reset ist demzufolge in allen Kategorien ähnlich stark vertreten und ist somit kein inhaltlich funktionalisiertes Mittel in der syntagmatischen Strukturierung von Äußerungen in prosodische Phrasen. Die Ergebnisse von Pilotexperimenten zu den Experimenten 1a und 1b haben ausserdem gezeigt, dass die separierende Wirkung von F0-Reset, wenn überhaupt vorhanden, dann sehr gering ist. Diese Überlegungen führen dazu, den Reset aus dem funktional orientierten Klassifikationssystem prosodischer Grenzen auszuschließen.

### 3.4.2.3 Längung des finalen Reims

In den Perzeptionsexperimenten 1a, 1b und 1c konnte gezeigt werden, dass Versuchspersonen bei zunehmender Längung des phrasenfinalen Reims nur in sehr geringem Maße eine zunehmende Separierung zwischen benachbarten Phrasen wahrnehmen. Weiterhin ist Längung in allen IB-Klassen ähnlich stark vertreten und somit keiner inhaltlichen Funktion im Rahmen der flüssigen turninternen Phrasierung zuzuordnen:

- IB-: n=30 davon mit Längung: 24 (80%),
- IB0: n=85 davon mit Längung: 59 (69,4%),
- IB+: n=64 davon mit Längung: 47 (73,4%).

Auf der Basis dieser Daten muss die Hypothese verworfen werden, dass eine Längung des phrasenfinalen Reims eine perzeptorische Separierung bewirkt, die wiederum eine inhaltliche Separierung signalisiert. Diese Überlegungen führen dazu, auch das Merkmal 'Längung' aus dem Klassifikationssystem auszuschließen.

#### **3.4.2.4 Zusammenfassung**

Phrasenfinale Konturverläufe und Pausen sind sprachlich klar funktional angebunden. Vor allem die tief fallenden Konturen stehen im Zusammenhang mit inhaltlichen Einschnitten in der Äußerung. Ebenso treten Pausen vor allem an hohen inhaltlichen Grenzen auf. Längung und F0-Reset zeigen hingegen im Bereich flüssiger prosodischer Grenzen keine erkennbare inhaltliche Funktion. Trotz der deutlichen funktionalen Bindung vor allem der tief fallenden Konturen steht aber die perzeptorische Separierung, die diese melodischen Verläufe verursachen, hinter der der Pausen zurück. Hier wird das Prinzip der Konvergenz semantisch-pragmatischer und phonetischer Kohäsion bzw. Separierung durchbrochen.

Ein funktional orientiertes Klassifikationssystem für flüssige prosodische Grenzen sollte also auf den Merkmalen F0-Kontur und Pause beruhen. So können für die extremen Kategorien der inhaltlichen Grenzen (IB- und IB+) jeweils prototypische Merkmalsbündel angesetzt werden:

- Für prosodische Grenzen ohne inhaltlichen Einschnitt (IB-): keine Pause, keine starke tonale Bewegung. Die perzeptorische Separierung ist durch phrasenfinale Längung und/oder Reset bedingt.
- Für prosodische Grenzen an hohen inhaltlichen Grenzen (IB+): separierende Kontur und Pause. Üblicherweise werden prosodische Grenzen mit Pausen und/oder separierenden Konturen durch F0-Reset und/oder segmentelle Längung verstärkt. Dieses Ergebnis deckt sich weitgehend mit Shriberg et al. (2000), die beobachten, dass sich im Englischen an prosodischen Grenzen mit nachfolgendem Themenwechsel meist eine Pause ('silent pause'), ein tiefer Grenzton und ein F0-Reset findet.
- Der Kategorie der mittelstarken inhaltlichen Einschnitte (IB0) ist weniger eindeutig ein bestimmtes Merkmalsbündel zuzuordnen.

### 3.5 Klassifikationssystem I - Flüssige turninterne prosodische Grenzen

Aus den Resultaten der Untersuchungen in Kapitel 3 wird das in Tabelle 15 zusammengefasste Klassifikationssystem für flüssige turninterne prosodische Grenzen abgeleitet. Die PG-Klassen sind mit PGi1-PGi4 bezeichnet, wobei i für 'intern' steht, da mit diesem System die turn**internen** Grenzen klassifiziert werden. Für die turnfinalen und die unflüssigen prosodischen Grenzen werden zwei weitere Klassifikationssysteme entwickelt und als System II und System III bezeichnet.

**Tabelle 15:** Klassifikationssystem I.

<b>Nomenklatur</b>	<b>Phonetische Definition</b>	<b>Funktion</b>
PGi1	Kontur: - Pause: -	Phrasierung innerhalb inhaltlicher Blöcke (IB-)
PGi2	Kontur: - Pause: +	Phrasierung innerhalb inhaltlicher Blöcke (IB-) und an schwachen inhaltlichen Grenzen (IB0)
PGi3	Kontur: + Pause: -	Phrasierung an schwachen oder starken inhaltlichen Grenzen (IB0 oder IB+)
PGi4	Kontur: + Pause: +	Phrasierung an starken inhaltlichen Grenzen (IB+)

Alle Merkmalsbündel des Systems I können von Längung des finalen Reims und/oder F0-Reset begleitet sein.

### 3.5.1 Verteilungen der prosodischen Grenzen über die inhaltlichen Klassen für das Klassifikationssystem I

Die folgenden Tabellen zeigen die empirischen Verteilungen der PGi-Klassen im Terminabsprachekorpus. So soll die Adäquatheit des Systems I überprüft werden. Tabelle 15 zeigt die vorgefundenen Anteile der vier PGi-Klassen über die drei Klassen inhaltlicher Grenzen in den Terminabsprachedaten. Um eine bessere Vergleichbarkeit zwischen den inhaltlichen Kategorien zu schaffen, werden die Klassen IB-, IB0 und IB+ auf eine Klassengröße von n=100 normiert.

**Tabelle 16:** Zuordnung der PG-Typen aus System I zu den inhaltlichen Grenzen (Terminabsprachen, auf n=100 normiert).

	<b>IB- (n=100)</b>	<b>IB0 (n=100)</b>	<b>IB+ (n=100)</b>
<b>PGi1 (n=106,6)</b>	60,0 (56,3%)	34,1 (32,0%)	12,5 (11,7%)
<b>PGi2 (n=52,5)</b>	20,0 (38,1%)	20,0 (38,1%)	12,5 (23,8%)
<b>PGi3 (n=51,9)</b>	10,0 (19,3%)	24,7 (47,6%)	17,2 (33,1%)
<b>PGi4 (n=89)</b>	10,0 (11,2%)	21,2 (23,8%)	57,8 (64,9%)

Das System I kann die Kategorien PGi1 und PGi4 mit einer Wahrscheinlichkeit von 56,3% bzw. 64,9% den Kategorien IB- bzw. IB+ zuordnen. Oder: Mit einer Wahrscheinlichkeit von 56,3% ist die Kategorie IB- durch das Merkmalsbündel PGi1 repräsentiert, mit einer Wahrscheinlichkeit von 64,9% die Kategorie IB+ durch das Merkmalsbündel PGi4. Weiterhin wird deutlich, dass die inhaltliche Kategorie IB+ nur selten durch die prosodische Klasse PGi1 repräsentiert wird (11,7%) und die inhaltliche Klasse IB- nur zu einem geringen Anteil durch die Klasse PGi4 (11,2%). Die Ergebnisse sind also in beiden Richtungen relativ gut. Trotzdem zeigt die Verteilung, dass eine eindeutige Form-Funktions-Beziehung nicht gegeben ist. Dies wird vor allem durch die breite Streuung der Grenztypen PGi2 und PGi3 über die drei inhaltlichen Klassen deutlich.

Tabelle 17 zeigt einen Vergleich zwischen einer zufälligen Verteilung der PG-Klassen über die inhaltlichen Klassen und der gefundenen Verteilung. Es wird deutlich, dass vor allem die Verbindungen von PGi1 mit niedrigen inhaltlichen Grenzen und von PGi4 mit hohen inhaltlichen Grenzen signifikant von der zufälligen Verteilung abweicht.

Für die Klassen PGi2 und PGi3 zeigen sich hingegen keine eindeutigen Abweichungen von der zufälligen Verteilung, außer dass die Klasse PGi3 deutlich seltener an niedrigen inhaltlichen Grenzen vorkommt als bei zufälliger Verteilung. Diese Befunde lassen sich auch in einem Chi-Quadrat-Test bestätigen.

**Tabelle 17:** Wahrscheinlichkeitstabelle für das System I (Terminabsprachen). Jede Zelle enthält die folgenden vier Werte: (1) Zellenwahrscheinlichkeiten bei zufälliger Verteilung vs. gemessene Zellenwahrscheinlichkeiten und (2) erwartete Häufigkeiten bei zufälliger Verteilung (normiert) vs. gemessene Häufigkeiten (normiert).

	<b>IB- (n=100)</b>	<b>IB0 (n=100)</b>	<b>IB+ (n=100)</b>
<b>PGi1 (n=106,6)</b>	0.118 vs. 0,2 35,4 vs. 60	0.118 vs. 0,113 35,4 vs. 34,1	0.118 vs. 0,042 35,4 vs. 12,5
<b>PGi2 (n=52,5)</b>	0.058 vs. 0.067 17,5 vs. 20	0.058 vs. 0,67 17,5 vs. 20	0.058 vs. 0.042 17,5 vs. 12,5
<b>PGi3 (n=51,9)</b>	0.058 vs. 0,033 17,5 vs. 10	0.058 vs 0,082 17,5 vs. 24,7	0.058 vs 0.057 17,5 vs. 17,2
<b>PGi4 (n=89)</b>	0.1 vs. 0,033 30 vs. 10	0.1 vs. 0.071 30 vs. 21,2	0.1 vs. 0,193 30 vs. 57,8

### 3.5.2 Die nicht-spezifizierten Merkmale Längung und Reset in System I

Die folgenden Auswertungen der Merkmalsbündel zeigen die Verteilungen der Merkmale Längung und Reset über die Klassen PGi1-PGi4 des Systems I.

**Tabelle 18:** Verteilung von Reset und Längung über die PG-Klassen des Systems I (Terminabsprachen). Prosodische Grenzen mit Resetentscheidung, n=2008. In jeder Zelle sind drei Zahlen angegeben: (1) absolute Häufigkeit, (2) relative Häufigkeit in Bezug auf n der PG-Klasse (Prozentangaben der Zeile addieren sich auf 100%), (3) relative Häufigkeit in Bezug auf die Gesamtstichprobe (n=2008, alle Zellen zusammen addieren sich auf 100%).

	<b>-L -R</b>	<b>-L +R</b>	<b>+L -R</b>	<b>+L +R</b>
<b>PGi1 (n= 761 (37,9%)) (-K -P)</b>	96 (12,6%) 4,8%	255 (33,5%) 12,7%	124 (16,3%) 6,2%	286 (37,6%) 14,2%
<b>PGi2 (n= 357 (17,8%)) (-K +P)</b>	13 (3,6%) 0,7%	44 (12,3%) 2,2%	65 (18,2%) 3,2%	235 (65,8%) 11,7%
<b>PGi3 (n= 524 (26,1%)) (+K -P)</b>	41 (7,8%) 2,0%	202 (38,6%) 10,1%	71 (13,6%) 3,5%	210 (40,1%) 10,5%
<b>PGi4 (n= 366 (18,2%)) (+K +P)</b>	8 (2,2%) 0,4%	60 (16,4%) 3,0%	33 (9%) 1,6%	265 (72,4%) 13,2%

Die Zahlen in Tabelle 18 zeigen für das Auftreten von Längung und/oder Reset keine eindeutige Bindung an die Merkmale Pause und/oder separierende Kontur. Ein Zusammenhang der Form "Wenn eine Pause auftritt, liegt auch Reset vor, und wenn keine Pause auftritt, liegt auch kein Reset vor" wäre ein gutes (weiteres) Argument für das Weglassen von Reset aus dem Klassifikationssystem. Die einzigen deutlichen Zusammenhänge, die die Tabelle zeigt, sind, dass das Vorkommen einer Pause in über 80% aller Fälle mit Längung des finalen Reims einhergeht und dass ein kombiniertes Auftreten von separierender Kontur und Pause in 72,4% mit Längung und Reset zusammengeht. Allerdings gibt es auch viele Fälle, bei denen Reset und/oder Längung ohne Pause und/oder separierende Kontur vorkommen (vgl. Tabellen 7 und 8). Die hohe Auftretenshäufigkeit von Längung und Reset in der Klasse P*Gi*4 deutet darauf hin, dass ein wirkliches "zum Ende kommen", das mit der Kategorie P*Gi*4 verbunden ist, üblicherweise neben einer Pause und einer separierenden Kontur auch finale Längung und Reset impliziert.

Weiterhin ist auffällig, dass prosodische Grenzen ohne Reset und ohne Längung selten sind (7,9% = 158 von 2008 Fällen). Von diesen 158 Fällen gibt es nur 21 Fälle, die auch eine Pause aufweisen. Wie bereits gezeigt werden konnte, stehen vor allem die prosodischen Grenzen mit Pause und/oder separierender Kontur in Verbindung mit hohen inhaltlichen Einschnitten (IB+). In diesen Fällen scheint eine Ergänzung des Merkmalsbündels durch Längung und/oder Reset üblich zu sein.

Hiermit sind Pausen und melodische Merkmale am Phrasenende relativ unabhängige, funktional gebundene Gliederungsmerkmale, während Reset und Längung nur eine schwache funktionale Bindung im Bereich der prosodischen Phrasierung zeigen.

## 4 Prosodische Grenzen und Sprecherwechsel

Wie im vorangegangenen Kapitel gezeigt wurde, sind die Positionierung und die phonetische Ausprägung prosodischer Grenzen mit der Informationsstruktur von Äußerungen verknüpft. Durch diese Verbindung zwischen inhaltlicher und phonetischer Struktur entsteht auch eine zentrale Bedeutung prosodischer Grenzen für die Organisation des Sprecherwechsels im Gespräch, denn Gesprächspartner orientieren sich bei Turnwechseln an der inhaltlichen Struktur des Gesprächs. In diesem Kapitel werden prosodische Muster hinsichtlich ihrer interaktionalen Funktionen in Gesprächen eingeordnet.

Gespräche werden während des Sprechens geplant und strukturiert, und sowohl Inhalt als auch phonetische Struktur ergeben sich zu einem großen Teil spontan aus der Interaktion der Gesprächsteilnehmer. Die Fragestellungen, Hypothesen und Experimente dieses Kapitels basieren u.a. auf der Annahme, dass Hörer in Gesprächen fortlaufend Erwartungen über den weiteren Äußerungs- und Gesprächsverlauf bilden, also einschätzen, ob der derzeitige Sprecher weitersprechen wird oder ob eine Übernahme des Rederechts sinnvoll ist. Hierbei orientieren sich die Hörer unter anderem an den phonetischen Eigenschaften der gehörten Äußerung (speziell im Bereich der prosodischen Grenzen). Diese Einschätzungen geschehen unbewusst und folgen dem Fluss des Gehörten in seiner phonetischen, morpho-syntaktischen und semantisch-pragmatischen Ausprägung. Die von Hörern gebildeten Erwartungen über den weiteren Äußerungs- bzw. Gesprächsverlauf werden während des Zuhörens fortlaufend modifiziert und evtl. verworfen. Diese Erwartungsbildung ist für den Aufbau von Dialogen entscheidend, denn sie ermöglicht schnelle Reaktionen seitens des Hörers auf das Gehörte und ist so eine Voraussetzung für die Strukturierung von Gesprächen in Turns der einzelnen Sprecher. Vor allem, wenn an einer prosodischen Grenze eine Pause auftritt, muss der Hörer richtig einschätzen, ob der Sprecher in seiner Rede fortfahren oder das Rederecht abgeben möchte, ob er eine inhaltliche Ergänzung des bereits Gesagten geben möchte oder ob er zum Abschluss einer thematischen Einheit gekommen ist und evtl. eine Reaktion des Gesprächspartners erwartet (z.B. durch ein Rezeptionssignal). Wenn die Einschätzung des weiteren Gesprächsverlaufs durch den Hörer nicht mit der Erwartung und der Planung des

Sprechers übereinstimmt, steigt die Gefahr von Missverständnissen und unerwünschten Unterbrechungen durch den Dialogpartner.

Als empirische Belege für diesen Standpunkt lassen sich verschiedene Beobachtungen aus den untersuchten Produktionsdaten anführen. Die drei im Folgenden angeführten Phänomene zeigen die schnelle und zeitlich exakte Koordination zwischen den Redebeiträgen der Dialogpartner, die m.E. nur durch die vorausschauende Erwartungsbildung als grundlegenden Mechanismus in der Gesprächsorganisation erklärt werden kann.

- Unter den 377 untersuchten Sprecherwechseln im LINDENSTRASSE-Korpus gibt es 66 Turnübergänge (17,5%), an denen es weder zu einer Überlagerung von mehr als 100 ms, noch zu einer Pause von mehr als 100 ms zwischen den Turns kommt; vielmehr findet der vormalige Hörer äußerst exakt den Anschluss an den vorangehenden Turn. Diese präzise Koordination vieler Turnübergänge zeigt die große Sensibilität der Hörer für den Aufbau der Rede des Partners.



Audiobeispiele: exakt synchronisierte Sprecherwechsel: 2.6, 2.7, 2.9

- Auch die zeitliche Synchronisierung von Rezeptionssignalen mit der syntagmatischen Struktur der fortlaufenden Äußerung zeigt, dass die Hörer Gesprächsverläufe genau mitverfolgen und sehr präzise, vor allem am Ende prosodischer Phrasen, durch kurze Äußerungen ihre Aufmerksamkeit bekunden. Der größte Teil der vorkommenden Rezeptionssignale befindet sich in kurzen Redepausen des jeweiligen Turninhabers.



Audiobeispiel: Position von Rezeptionssignalen: 3.1

- In den Dialogen kommt es zu insgesamt 20 sog. *collaborative completions*. Das sind vom Hörer gegebene inhaltliche Ergänzungen dort, wo seitens des Sprechers Wortfindungsprobleme auftreten und der Hörer bereits weiß, was der Sprecher sagen will (Szczeppek 2000). *Collaborative completions* zeigen ebenfalls, dass Hörer sowohl inhaltlich als auch phonetisch den Ablauf der Äußerung mitverfolgen, hierbei Hypothesen über den weiteren Ablauf bilden und unter Umständen, wiederum mit sehr exaktem Timing, inhaltliche Ergänzungen liefern. Das Phänomen der *collaborative*

*completions* deutet vor allem darauf hin, dass Hörer Planungsprobleme seitens des Sprechers schnell erkennen und sich hierbei vermutlich auch auf typische phonetische Exponenten unflüssiger Sprechweise stützen.



Audiobeispiel: *Collaborative Completion*: 2.14

Natürlich gibt es das "perfekte Gespräch" nicht. Ein Gespräch beinhaltet fast immer auch Missverständnisse, Verständigungsschwierigkeiten und konkurrierendes Verhalten um das Rederecht. Jeder kennt auch Situationen und Gespräche, in denen Hörer und Sprecher nur wenig aufmerksam aufeinander reagieren. Wenn ein Hörer z.B. desinteressiert, müde oder geistig abwesend ist, kann es vorkommen, dass eine Äußerung einfach an ihm "vorbeirauscht". Trotzdem zeichnen sich Gespräche oft durch ein hohes Maß an Kooperationsbereitschaft und "interaktiver Feinfühligkeit" aus.

So ergeben sich folgende zentrale Fragen:

- Welche interaktionalen Funktionen können angesetzt werden, um die Koordination von Sprecher und Hörer zu erklären?
- Welche phonetischen Parameter prosodischer Grenzen stehen in Verbindung mit den angesetzten interaktionalen Funktionen?

Unter diesen Gesichtspunkten wird als erstes eine Zusammenfassung der Forschungsergebnisse aus der Literatur gegeben. Im Anschluss daran sind die eigenen Arbeiten beschrieben.

## **4.1 Stand der Forschung**

Die Frage nach der interaktionalen Bedeutung prosodischer Strukturen führt in ein verzweigtes Forschungsgebiet, in dem die Frage nach den grundlegenden Organisationsprinzipien sprachlicher Kommunikation aufgeworfen wird. Hier ist die Phonetik nur eine Teildisziplin unter vielen, die herangezogen werden, um einen so komplexen Vorgang wie ein Gespräch zu beschreiben und zu erklären. Es ist aber nicht das Ziel dieser Arbeit, die Dialoge, die hier untersucht werden in ihrer ganzen Komplexität zu erfassen. Es soll

vielmehr der Beitrag der Phonetik für die interaktive Strukturierung von Gesprächen herausgearbeitet und systematisiert werden. Deshalb konzentriert sich der folgende Literaturüberblick weniger auf linguistische, psychologische und soziologische Ansätze der Forschung, sondern mehr auf phonetisch orientierten Untersuchungen.

In der Literatur zum Thema Gesprächsorganisation gibt es ein zentrales Problem: Sowohl die untersuchten interaktionalen Strukturen als auch ihre formalen Exponenten sind komplex und zeigen ein hohes Maß an Variabilität. Dies macht die Definition der zu untersuchenden Kategorien problematisch. Auf der Ebene der Wortphonetik und -phonologie ist dies anders. Hier kann traditionell z.B. mit dem Konzept der Bedeutungsunterscheidung durch verschiedene phonetische Einheiten gearbeitet werden. So kann ein Wortpaar wie *Paar* vs. *Bar* durch die *voice onset time* im initialen Konsonanten differenziert werden und es ergibt sich ein relativ klarer Form-Funktions-Zusammenhang. Auf der Ebene der Satzphonetik beginnt dieser Zusammenhang durch satzphonetische Reduktionserscheinungen zu verschwimmen und kontextuelle, morpho-syntaktische und semantisch-pragmatische Faktoren müssen in Form-Funktions-Beziehungen integriert werden, um zu einer adäquaten Erklärung z.B. der weitgehenden Verschleifungen zu gelangen, die in spontaner Sprache auftreten (siehe Kohler 1995a: 201ff).

Die Analyse interaktionaler Strukturen im Gespräch stellt den Bearbeiter vor ein noch komplexeres Netzwerk beobachtbarer Formen und Funktionen. Hier müssen kommunikativ relevante Formen, vom phonetischen Detail, über syntaktische und pragmatische Strukturen, bis hin zum gesamten Gesprächskontext einbezogen und in ihren Wechselwirkungen untersucht werden. Auf der funktionalen Seite ergibt sich das Problem, dass es keine einfachen Funktionen gibt, wie z.B. die Übermittlung eines lexikalischen Eintrags, sondern immer eine Kopplung aus semantischer Information und interaktionaler Bedeutung vorliegt, die kaum zu trennen und zu systematisieren ist. Diese Komplexität und das Hinausgehen der Fragestellung über den Bereich der traditionellen Phonetik und deren Methoden hat möglicherweise dazu geführt, dass systematische phonetische Erkenntnisse über die Gesprächsorganisation erst verhältnismäßig spät mit der aufkommenden *Conversation Analysis* und der Adaption ihrer Methoden in Untersuchungen mit phonetischen Schwerpunkten erzielt wurden. Mit dem Aufsatz *A simplest systematics for the organisation of turn-taking for conversation*. legten Sacks et al. 1974 den Grundstein der Anwendung des ursprünglich soziologisch ausgerichteten

interpretativen Paradigmas auf die phonetische Untersuchung von Gesprächsstrukturen und somit auch für die *Conversation Analysis*.

Wie bereits in Abschnitt 3.4.1 beschrieben, ist die Vorgehensweise der CA nicht ohne Probleme und Widersprüche. Es kann ihr die Übergeneralisierung von Einzelbeobachtungen ebenso vorgehalten werden wie mangelnde statistische und messphonetische Absicherung ihrer Ergebnisse. Auch Kritik an der postulierten Objektivierbarkeit von Beobachtungen durch die sequentielle Analyse ist angebracht. Trotzdem haben die Analysen der CA viele neue Perspektiven für die Untersuchung sprachlicher Interaktion geliefert. Sie haben die Komplexität des Gesprächs beleuchtet und gezeigt, dass wirkliche Erklärungen von sprachlicher Interaktion viele funktional relevante Ebenen integrieren müssen. Und sie hat nicht zuletzt gezeigt, dass die systematische Interpretation und phonetische Detailanalyse auch ohne eine befriedigende Quantifizierung phonetischer Strukturen entscheidende Konzepte für den Zusammenhang phonetischer und interaktionaler Muster liefern kann.

Die Konzepte, die im Rahmen der CA erarbeitet wurden, sind in diesem Kapitel die Basis für die systematische Untersuchung von Sprachkorpora und für die Konzeption eines Perzeptionsexperiments, denn diese Konzepte zeigen, wonach in den Datenbasen gesucht werden muss und welche Hypothesen weiteren Prüfungen zu unterziehen sind. Insofern legt die CA für viele Ansätze dieser Arbeit den Grundstein, ohne dass deren Prüfung mit den Methoden der CA unternommen wird.

#### **4.1.1 Der Turn als Einheit**

Local & Kelly (1986) definieren die Einheit *Turn* als alles, was ein Sprecher sagt, nachdem er das Wort übernommen hat und bis ein Sprecherwechsel folgt. Die interaktive Delimitation von Turns kommt dadurch zustande, dass der Hörer die Rede des Sprechers als beendet auffasst und dieses demonstriert, indem er anfängt zu sprechen. Ein Sprecher plant also, sowohl wann als auch wie er anfängt zu reden in Relation zur fortlaufenden Rede des anderen Sprechers. Diese Sichtweise spiegelt die Prinzipien der sequentiellen Analyse der CA wider und benutzt das Hörerverhalten als "objektivierendes" Kriterium, die Intentionen von Sprechern festzustellen. Der Ansatz ist zum einen problematisch, da Beobachtungen in spontanen Dialogen zeigen, dass es auch viele Turnübernahmen im

Gespräch gibt, an denen der Sprecher seine Rede an der Position der Übernahme mit großer Wahrscheinlichkeit nicht als beendet auffasst (z.B. wenn der Hörer das Wort mitten in einer syntaktischen Einheit des Sprechers übernimmt).



Audiobeispiele: Turnübernahme innerhalb einer syntaktischen Einheit: 2.8, 2.10.

Zum anderen bietet der Ansatz für die Forschung ein einfaches externes Kriterium, Turns abzugrenzen: Ein Turn ist alles, was ein Sprecher sagt, bis ein anderer Sprecher das Wort übernimmt. Diese Definition wird auch im Rahmen der vorliegenden Studie angewendet. Der Terminus *Redebeitrag* wird synonym zu *Turn* benutzt.

An den hinteren Grenzen von Turns sind eine Reihe prosodischer Veränderungen zu beobachten, die nach Local & Kelly (1986) vor allem dort auftreten, wo Satzgrenzen mit Turngrenzen zusammenfallen. Diese Veränderungen werden als phonetische Delimitationssignale von Turns angesehen und treten meist gebündelt auf. An syntaktischen Grenzen, die nicht in turnfinaler Position stehen, sind hingegen andere (oder keine) Delimitationssignale zu beobachten. Bei der phonetischen Kodierung von Turngrenzen wird also eine Wechselwirkung phonetischer Eigenschaften mit syntaktischen und pragmatischen Strukturen angenommen, d.h. die phonetischen Merkmale zur Signalisierung der Turnübernahme werden vor allem am Ende syntaktischer Einheiten eingesetzt, und das Fehlen der Merkmale für einen Turnabschluss an hohen syntaktischen Grenzen wird im Gegenzug als Signal für den Hörer verstanden, das Wort nicht zu übernehmen (Local & Kelly 1986). Hiermit werden zwei wichtige Konzepte eingeführt. Erstens die Bündelung von Delimitationssignalen an prosodischen Grenzen bzw. an Turngrenzen und zweitens die funktionale Differenzierung zwischen dem Halten eines Turns (*turn-holding*) und dem Übergeben des Rederechts (*turn-yielding*). Diesen Konzepten wird im Laufe dieses Kapitels nachgegangen. Vorerst soll aber weiter die Einheit *Turn* eingeführt werden. Nach der gegebenen Definition eines Turns stellt sich nun die Frage: Woraus bestehen Turns?

Betrachtet man die Turnstruktur eines Gesprächs, wird klar, dass sich einzelne Turns deutlich voneinander unterscheiden. Es gibt Beiträge, die nur aus einem einzigen Wort bestehen, und auch solche, die komplexe Schilderungen enthalten und mehrere Minuten dauern. Diese unterschiedliche Komplexität kann erfasst werden, indem man Turns in *Turnkonstruktionseinheiten* unterteilt (*turn constructional units, TCU*, Sacks et

al. 1974). Ein Turn kann aus einer oder mehreren Turnkonstruktionseinheiten bestehen. Die Definition der linguistischen Form der Einheit TCU bleibt bei Sacks et al. allerdings vage und wird erst in späteren Untersuchungen präzisiert. Ursprünglich wird vor allem eine syntaktisch basierte Beschreibung der Einheit TCU angestrebt. Da sich aber in der Analyse von Gesprächen gezeigt hat, dass syntaktische Strukturen für die Vorhersage von TCU-Grenzen keine geeigneten Prädiktoren sind, wird die traditionelle Auffassung der syntaktisch gesteuerten Einheitenbildung vor allem durch prosodische Einheitenkonstruktionsmechanismen ergänzt. Selting (1998:30):

*A 'TCU' is thus a unit that is constituted and delimited by the interplay of syntax and prosody: it is constituted as a cohesive whole by the deployment of syntactic and prosodic construction schemata, and it ends with the co-occurrence of a possible syntactic and a possible prosodic unit completion in its sequential context.*

Hier wird also neben der syntaktischen Struktur das Prinzip der prosodischen Kohäsion durch globale intonatorische Muster in die Definition der TCU einbezogen. Solche Muster können z.B. Hutmuster sein, aber auch eine Reihe von Gipfelkonturen mit Downstep (vgl. *Vortagsintonation*, v. Essen, 1964) oder eine Aneinanderreihung von mäßig steigenden Talkonturen (vgl. *Zickzackmelodie*, v. Essen, 1964). Das Ende von Turnkonstruktionseinheiten wird hingegen durch den Abschluss einer syntaktischen Konstruktion und variierende Bündel phonetischer Merkmale realisiert, die antagonistisch zum Mechanismus der internen Kohäsion der TCU zu Separierung zwischen benachbarten TCU führen. Ford & Thompson (1996) weisen zusätzlich darauf hin, dass auch die Komplettierung pragmatischer Einheiten zur Einheitenbildung beiträgt.

Werden alle diese Faktoren zusammengenommen, ergibt sich eine ziemlich unklare Mischung aus verschiedensten definitorischen Kriterien, deren konsistente Anwendung auf natürliche Sprache m.E. problematisch ist bzw. nur mit einem hohen Maß an Interpretation die Segmentation in Turnkonstruktionseinheiten erlaubt. Aus diesem Grund wird in der vorliegenden Arbeit keine Einteilung von Turns in Turnkonstruktionseinheiten im Sinne der CA vorgenommen. Statt dessen wird eine rein prosodisch basierte Beschreibung von Redebeiträgen gewählt, die das Problem der

syntaktischen und inhaltlich-pragmatischen Interpretation umgeht. Turns werden als eine Aneinanderreihung prosodisch kohäsiver Abschnitte angesehen, die durch prosodische Grenzen voneinander getrennt sind und die in flüssiger Sprache als *prosodische Phrasen* bezeichnet werden. Somit kann eine prosodische Phrase durchaus einer TCU entsprechen und eine TCU wird immer mindestens eine prosodische Phrase enthalten. Aber es kann durchaus der Fall eintreten, dass sich die Einheiten nicht decken. Selting (1995) führt z.B. an, dass Sätze, die durch Pausen unterbrochen werden, trotzdem eine zusammenhängende TCU bilden. Hiermit wird die prosodische Separierung innerhalb von TCU der syntaktischen Kohäsion untergeordnet. Diese Problematik stellt sich bei der Untersuchung der von mir gewählten Analyseeinheit nicht, da hier lediglich prosodische Einschnitte betrachtet werden, die ohne Bezug zur Syntax auf perzeptorischer Basis etikettiert wurden (siehe 2.1.3.4).

Im Rahmen der CA spielen die Endpunkte von Turnkonstruktionseinheiten eine wichtige Rolle für die Steuerung von Sprecherwechseln, da diese Positionen in Äußerungen von den Gesprächsteilnehmern als *turnübergaberelevante Stellen* interpretiert werden (*transition relevant places (TRP)*; Sacks et al. (1974)). An diesen TRP kann ein Sprecherwechsel vorkommen, muss aber nicht. Schegloff (1996:55):

*By 'turn constructional unit', it may be recalled, we meant to register that these units can constitute possibly complete turns; on their possible completion, transition to a next speaker becomes relevant (although not necessarily accomplished).*

In der Sichtweise der CA ermöglichen vor allem zwei Merkmale dem Hörer zu entscheiden, wann eine TCU endet und eine turnübergaberelevante Stelle und somit ein mögliches Turnende erreicht wird: die Syntax und die Prosodie. Der zentrale Terminus der CA ist hierbei *Projektierung*. Sprecher *projektieren* durch syntaktische und prosodische Muster das Ende von Turnkonstruktionseinheiten und damit das mögliche Ende von Turns. In älteren Veröffentlichungen wird als zentrales Merkmal die syntaktische Projektierung angesetzt, also davon ausgegangen, dass die Projektierung turnübergaberelevanter Stellen vor allem durch syntaktische Komplettierung gesteuert wird. Aktuelle Studien heben hingegen vor allem die prosodisch basierte Projektierung

hervor oder gehen von komplexen Merkmalsbündeln aus, die ebenfalls prosodische und pragmatische Projektierungen einbeziehen. Ford & Thomson (1996) nennen eine Position im Turn, wo prosodische, syntaktische und pragmatische Komplettierung gemeinsam auftreten *complex transition relevant place (CTRP)*. An diesen Positionen finden bevorzugt Sprecherwechsel statt. Ford & Thompson (1996:172):

*A major finding of this study is the fact that speaker change correlates with CTRPs. This is evidence that the units defined by the convergence of syntactic, intonational, and pragmatic completion are real for conversationalists; speakers and hearers orient to, and design their own turns in response to, these units.*

Dieses Zitat und das damit verbundene Konzept des *möglichen Turnendes* führt zu einer zentralen funktionalen Fragestellung der CA wie auch der vorliegenden Arbeit. Eine TCU (oder auch eine prosodische Grenze) kann sowohl turnintern oder turnfinal stehen und jeder Turn ist prinzipiell nach rechts expandierbar. Selbst wenn ein syntaktischer und prosodischer Abschlusspunkt erreicht ist, kommt es nicht immer zum Sprecherwechsel. Wie signalisieren Sprecher also, dass eine Einheit (TCU bzw. prosodische Phrase) das Ende bzw. nicht das Ende des Redebeitrags ist? Oder anders formuliert: Aufgrund welcher Faktoren reagieren Sprecher mit Turnübernahmen? Diese Fragestellung soll hier unter dem Begriffspaar *Abschluss vs. Weiterweisung* diskutiert werden. Gilles (2005:21):

*Bei der Gliederung von Gesprächsbeiträgen ('Turns') können zwei antagonistische Funktionskomplexe unterschieden werden: Einerseits werden spezifische linguistische Merkmale eingesetzt, um das Ende eines Gesprächsbeitrags zu signalisieren. Andererseits können spezifische Merkmale darauf hindeuten, dass der Sprecher seinen Beitrag noch nicht beendet hat. Diese abstrakten Funktionskomplexe werden im Folgenden mit den Begriffen 'Abschluss' und 'Weiterweisung' bezeichnet.*

Prototypische Abschlüsse sind am Ende von Gesprächsbeiträgen zu beobachten. Im Rahmen der CA wird davon ausgegangen, dass an einem Turnende zur Signalisierung der Abschlussfunktion sowohl syntaktische als auch pragmatische und prosodische Projektierungen eingelöst werden. Das bedeutet für die syntaktische Ebene, dass die durch den Valenzrahmen vorgegebenen obligatorischen syntaktischen Strukturpositionen besetzt sein müssen (*syntactic completion*), also die syntaktische Struktur zu einem möglichen Endpunkt kommt, an dem keine weiteren Ergänzungen **notwendig** sind (syntaktische Ergänzungen sind aber immer **möglich**, da jede syntaktische Konstruktion nach rechts erweiterbar ist). Ford & Thompson (1996:155) gehen davon aus, dass die projektierende Kraft (möglicher) syntaktischer Abschlüsse nicht besonders groß ist, auch wenn der syntaktische Abschluss ein notwendiges Kriterium für die Abschlussfunktion ist:

*Syntactic completion points alone are the least reliable indicators of any other sort of completion.*

Also liegt die funktionale Last für die Differenzierung von Abschluss und Weiterweisung wieder auf der inhaltlich-pragmatischen und/oder prosodischen Ebene. Für die Definition eines semantisch-pragmatischen Abschlusses lassen sich in der phonetisch ausgerichteten Literatur kaum Kriterien finden. Die Analyse semantisch-pragmatischer Projektierungen kann nur interpretativ vorgenommen werden und verläuft sich weitgehend in Spekulationen.

Weiterweisung als Antagonist in der Funktionsopposition von Abschluss und Weiterweisung bedeutet, dass eine Fortsetzung des Turns über die erreichte Grenze (TCU oder prosodische Grenze) angestrebt wird. Der derzeitige Sprecher ist also nicht bereit, das Rederecht abzugeben, obwohl ein gewisser Einschnitt im Turn vorliegt, und baut eine Fortsetzungsprojektion auf. Aus der Perspektive des Hörers werden solche weiterweisenden prosodischen Grenzen so interpretiert, dass dort kein Sprecherwechsel erwartet oder erwünscht ist. Das Konzept der *Weiterweisung* findet sich in vergleichbarer Form auch bei v. Essen (1964) als *Progredienz*, bei Delattre, Poenack & Olsen (1965) und Pierrehumbert & Hirschberg (1990) als *continuation* und bei Selting (1995) als *Turnhalten*.

Die CA orientiert sich in der Bestimmung von weiterweisenden vs. abschließenden Grenzen innerhalb von Mehrheitenturns am Verhalten des Hörers. Weiterweisung wird dort angenommen, wo ein Sprecher seine Rede fortsetzt. Abschluss liegt vor, wenn der Hörer nachfolgend das Wort übernimmt. Das bedeutet, dass Weiterweisung nicht an einem Turnende vorkommen kann und konversationelle Abschlüsse nicht im Turninneren. Hier zeigt sich eine grundlegende Problematik der CA, die in ihren Interpretationen davon ausgeht, dass die Intention des Sprechers durch das Verhalten des Hörers erkennbar wird. Wenn also ein Sprecher das Wort übernimmt, muss dies in der Sichtweise der CA durch die Projektierung eines konversationellen Abschlusses hervorgerufen sein bzw. wenn ein Sprecher nach einer Grenze widerspricht, geht dies mit der Projektierung einer Folgeäußerung einher. Die Analyse natürlicher Gespräche zeigt aber, dass diese Sichtweise nur einen Teil der Turnabschlüsse bzw. Folgeäußerungen erklären kann, denn an vielen Turnübergängen in den von mir untersuchten Daten findet sich weder ein syntaktischer noch ein pragmatischer Abschluss und auch die melodische Bewegung liefert keinen Hinweis auf einen Abschluss. Es scheint an diesen Positionen zu unterbrechenden Turnübernahmen zu kommen, die rein auf die Initiative des Hörers zurückgehen und nicht durch die Intention des Sprechers erklärt werden können. Auf der anderen Seite finden sich viele prosodische Grenzen im Korpus, an denen sämtliche Projektionen eingelöst sind und es trotzdem nicht zu einem Sprecherwechsel kommt.



Audiobeispiele: unterbrechende Turnübernahmen: 2.8, 2.10, Turnfortsetzung trotz

Finalitätsmerkmalen: 1.4

Wenn also dem Signalisierungssystem des Sprechers eine Regelmäßigkeit in der Form-Funktions-Beziehung zugrunde liegt, kann diese nicht allein aus dem Hörerverhalten abgeleitet werden. Im folgenden Abschnitt des Literaturüberblicks wird zusammengefasst, welche Rolle der Prosodie in der Signalisierung von Abschluss und Weiterweisung beigemessen wird.

In der Literatur finden sich viele Hinweise auf interaktionale Funktionen bestimmter Melodiemuster. Zahlreiche Untersuchungen belegen eine universelle Tendenz, Abschlüsse mit fallenden Intonationsmustern zu realisieren (Delattre et al. 1965; Lieberman 1967; Bolinger 1978; Cruttenden 1981, 1986; Wichmann 2000). Weiterweisung wird oft in Zusammenhang mit final steigenden oder fallend-steigenden

Intonationsverläufen gebracht (z.B. Bolinger 1978; Cruttenden 1981). Tendenzen dieser Art lassen sich auch in Untersuchungen zum Deutschen finden. So zeigen Peters et al. (2005b), dass am Ende von über 1000 untersuchten Turns tief fallende Intonationsmuster mit einer Häufigkeit von fast 50% vorkommen. Eine eindeutige Beziehung zwischen melodischen Mustern und Sprecherwechseln im Gespräch lässt sich hieraus aber nicht ableiten. Im Gegenteil – Peters et al. belegen eine starke Überschneidungen zwischen den Verteilungen melodischer Muster an prosodischen Grenzen im Inneren von Turns und am Ende von Turns, denn ein großer Teil der turninternen prosodischen Grenzen im untersuchten Korpus (29,6%) wird ebenfalls mit tief fallender Intonation realisiert.

Befunde wie diese werfen die Frage auf, wie eng der Zusammenhang zwischen Form und Funktion melodischer Muster im Dialog ist und welche Systematisierungen die große Variabilität intonatorischer Muster in vergleichbaren Positionen im Turn adäquat erfassen können. Einen extremen Standpunkt in der Beantwortung dieser Frage nimmt Jefferson (1986) ein, die die Intonation nur als Begleiterin der ausschlaggebenden syntaktischen und semantisch-pragmatischen Strukturen ansieht und in der intonatorischen Signalisierung von Dialogstrukturen kein Ursache–Wirkung-Prinzip sieht. Diese Ansicht wird durch das folgende Zitat deutlich. Es muss allerdings gesagt werden, dass Jefferson im weiteren Verlauf der Publikation Einschränkungen dieser rigorosen Sichtweise vornimmt. (Jefferson 1986:179, f):

*When I talk about how recipients monitor an utterance in progress, I talk in terms of possibly complete actions, and/or syntactic possible completion. Now and again people ask if I've considered intonation, and if so, does it matter? I say I have, and it doesn't seem to.*

*At the start of this project on overlapping talk I did a mass of transcribing, and was careful to catch a range of intonational features, assuming that would have a lot to do with the phenomenon. But while I was working with the materials, it didn't seem to matter. Speakers would, for example, produce what I heard as 'full stop' intonation and then keep going (...). Or, for example, speakers would produce 'continuing' or indeterminate intonation, a recipient would start up, and there would be a clean transition (...). What held across*

*cases of clean transition and recipient onset resulting in overlap alike was observable completed actions or syntactic possible completion.*

Einen ganz anderen Standpunkt vertritt Local in verschiedenen Publikationen. Locals Ansatz ist es, eine Phonologie der Konversation (*Conversational Phonology*) zu erarbeiten, die vor allem auf prosodische Merkmale gegründet ist. Local et al. (1985) belegen für das London-Jamaikanische einen klaren Zusammenhang zwischen melodischen Mustern und Turnübergaben. Allerdings gehen Local und seine Mitarbeiter davon aus, dass neben dem melodischen Verlauf auch jedes phonetische Detail von interaktionaler Bedeutung sein kann und in die Untersuchungen einbezogen werden muss. Insofern misst Local intonatorischen Mustern einen hohen funktionalen Stellenwert bei, nimmt aber gleichzeitig an, dass Intonationsmuster immer in komplexe prosodische Merkmalsbündel eingebunden sind, die in ihrer gesamten phonetischen Ausprägung eine bestimmte Funktion im Gespräch repräsentieren. Local, Kelly & Wells (1986) untersuchen weiterhin nicht-überlagernde Turnübergänge im Tyneside Englisch und stellen neben bestimmten wiederkehrenden Intonationsmustern folgende phonetische Merkmale an Turnenden fest:

- allgemeine Verlangsamung des Tempos
- plötzlicher Anstieg oder Abnahme der Lautstärke
- auffallende Dauer der ictischen Silbe im letzten Fuß
- Zentralisierung der Vokale

Diese Merkmalsbündel können optional phonatorische Veränderungen sowie segmentelle Merkmale wie die Lösung und Aspiration finaler Plosive enthalten.



Audiobeispiele: komplexe phonetische Veränderungen am Turnende: 2.1, 2.2

Local et al. (1986) kommen zu der Schlussfolgerung, dass Sprecher des Tyneside Englisch die oben genannten Merkmale als Elemente einer Phonologie der Konversation innerhalb ihres sprachlichen Systems nutzen, um Hörern den Wunsch nach einer Turnübergabe anzuzeigen und dass die Hörer diese Signale dekodieren und dadurch unmittelbar nach Abschluss der Äußerung den Turn übernehmen können. Allerdings wird

hierbei, wie bereits unter 4.1.1 beschrieben, eine Wechselwirkung mit semantischen und syntaktischen Strukturen angenommen. Phonetische Merkmale zur Signalisierung der Turnübernahme werden vor allem am Ende syntaktischer Einheiten eingesetzt, und das Fehlen von Abschlussmerkmalen an hohen syntaktischen Grenzen wird im Gegenzug als Signal verstanden, das Wort nicht zu übernehmen.

Selting (1995) beschreibt die funktionale Relevanz der Intonation unter Bezugnahme auf syntaktische Projektierungen und argumentiert ebenfalls im Sinne komplexer Signalisierungsmechanismen, die Informationen verschiedener Ebenen einbeziehen. Sie nimmt an, dass die Syntax eine globale projektierende Kraft hat und lokal durch Prosodie unterstützt wird, z.B. indem eine geplante Turnfortsetzung an einem vorläufigen syntaktischen Abschluss lokal durch ein ebenes melodisches Muster signalisiert wird (s.206). Selting (1995:75): Der global projektieren Kraft der Syntax gegenüber

*...haben prosodische Signale eher lokal kontextualisierende Kraft: Spezielle Haltesignale projektieren prospektiv eine Fortsetzung der Einheit bzw. des Turns und damit auch eine noch ausstehende Beendigung, und kontextualisieren retrospektiv die Art des Anschlusses der Fortsetzung (dies setzt/setzt nicht die projizierte Einheit fort). An möglichen Einheitenenden kontextualisieren sowohl fallende als auch steigende lokale Tonhöhenbewegungen mögliche Einheitenenden, nach denen jedoch prinzipiell ebenfalls noch Einheitenverlängerungen kohäsiv angeschlossen und die Einheiten fortgesetzt werden können. Dem spezifischen Einheiten- und Turnhaltesignal lokal gleichbleibender Tonhöhenbewegung stehen offenbar keine ähnlich spezifischen prosodischen Einheitenendtonhöhenbewegungen gegenüber.*

Die globale syntaktische Projektierung dient dem Hörer beim "Vorausblicken", also der in die Zukunft gerichteten Einschätzung des Äußerungs- und Gesprächsverlaufs und wird als wichtiges Orientierungsprinzip für den Hörer eingeordnet, das es dem Rezipienten erlaubt, Turnübernahmen mit möglichst minimaler Verzögerung durchzuführen. (Selting, 1995:75):

*Auch für den Rezipienten ist relativ gut einschätzbar, wann eine Einheit 'Satz' potentiell einem Ende zugeht, nach der dann im Prinzip [...] das Rederecht zur Disposition steht. Damit haben syntaktische Konstruktionen wie der 'mögliche Satz' eine global projektierende Kraft, indem sie während ihrer Produktion ihr potentielles Ende projektieren können. Mit solchen Sätzen kookkurriert eine holistische Intonationskontur mit lokal fallender oder steigender letzter Tonhöhenbewegung, die diese 'Sätze' zusätzlich als Ganzheiten konfiguriert.“*

In diesem Sinne äußern sich auch Ford, Fox & Thompson (1996:428):

*Syntactic units allow for the prediction of the possible completion points in advance of their arrival and thus contribute to the precise exchange of speakership.*

Zusammenfassend wird also von einer Reihe phonetischer und grammatischer Finalitätsmerkmale ausgegangen, die das Ende von Redebeiträgen im Gespräch kodieren und so im Rahmen komplexer Merkmalsbündel die Abschlussfunktion von der Weiterweisungsfunktion trennen. In der Analyse ergibt sich eine Parallele zwischen syntaktischen Einheiten und der Einheit Turn: Sowohl das Ende einer syntaktischen Einheit als auch ein Turnende ergibt sich in der Interaktion und kann nicht durch den Untersucher allein bestimmt werden, denn sowohl ein Turn als auch eine syntaktische Konstruktion ist immer nach rechts expandierbar. Hierbei ist m.E. eine geplante Rechtsexpansion funktional gesehen der Weiterweisungsfunktion gleichzusetzen. Auf der formalen Ebene ist die Weiterweisung eher durch das Fehlen von Finalitätsmerkmalen als durch ein eigenständiges System weiterweisender Merkmale kodiert.

#### 4.1.2 Typen von Sprecherwechseln und deren phonetische Beschreibung

Den bisher zusammengefassten Studien liegt die gemeinsame Sichtweise zugrunde, dass Sprecher im Gespräch durch komplexe Merkmalsbündel anzeigen, wann ihr Turn abgeschlossen ist und Hörer an diesen Positionen das Wort übernehmen. Die Analyse der LINDENSTRASSE-Daten zeigt aber, dass Sprecherwechsel auch an Positionen vorkommen, an denen weder ein pragmatischer noch ein syntaktischer oder intonatorischer Abschluss vorliegt. Sprecherwechsel dieses Typs folgen also nicht dem Prinzip der Signalisierung von Turnenden durch den Sprecher und der nachfolgenden Turnübernahme als Reaktion des Hörers. Vielmehr scheint der Hörer oftmals mitten in der Rede des Sprechers zu unterbrechen und einen Redebeitrag zu beginnen, ohne dass er vom Sprecher dazu "aufgefordert" wird. Solche unterbrechenden Turnübernahmen sind in ungezwungenen und lebhaften Gesprächen häufig und werden vom unterbrochenen Sprecher meist sofort akzeptiert, indem er unmittelbar nach dem unterbrechenden Redebeginn des Gesprächspartners aufhört zu sprechen. Es sind aber auch Turnübernahmen zu beobachten, bei denen über mehrere Sekunden gleichzeitig gesprochen wird und dadurch der Eindruck entsteht, dass die Turnübernahme nicht nur unterbrechend, sondern auch strittig ist. Im folgenden Abschnitt werden Untersuchungen zusammengefasst, die sich mit interaktional verschiedenen Arten von Sprecherwechseln und deren phonetischen Korrelaten beschäftigen.

French und Local (1983) arbeiten eine Klassifikation aus, die zwischen *competitive incomings* und *non-competitive incomings* trennt. *Non-competitive incomings* sind die oben bereits beschriebenen Turnübergänge, die Merkmale der Abschlussfunktion tragen und keine deutlichen Überlagerungen der angrenzenden Turn aufweisen. *Competitive incomings* finden an Positionen im Turn statt, die nicht als TRP (*transition relevant place*) klassifiziert werden, da sie keinen pragmatischen Abschluss bilden und auch syntaktisch oft nicht komplett sind. Die Gesprächsbeiträge der beteiligten Sprecher zeigen hierbei häufig Überlagerungen. Aber weder das Auftreten von Überlagerungen noch die thematische Relation zwischen bisherigem Turn und Folgeturn reichen aus, um eine Turnübernahme als *competitive* vs. *non-competitive* zu klassifizieren. Denn es gehört zu French & Locals Konzept, eine Turnübernahme nur dann als *competitive* zu klassifizieren, wenn sie den Eindruck haben, dass der Sprecher, dem

das Rederecht abgenommen wurde, dies als unerwünscht oder unangemessen empfindet. Und es kann durchaus vorkommen, dass ein Turn überlagernd ohne vorangehenden syntaktischen und/oder pragmatischen Abschluss übernommen wird, der Turninhaber dies aber nicht als unangemessen empfindet. Die Klassifizierung als *competitive* nimmt deshalb Bezug auf phonetische Eigenschaften von Turnübergängen, da sich in diesen Eigenschaften die Einstellung der Sprecher hinsichtlich der Verteilung des Rederechts widerspiegelt. French und Local analysieren hierfür Turnübergänge, an denen der unterbrechende Gesprächsteilnehmer in einen klaren Wettbewerb um das Rederecht mit dem in diesem Moment Sprechenden tritt, ohne dass dieser seinen Übergabewillen durch prosodische Mittel signalisiert hat. Folgende phonetische Eigenschaften werden als Belege eines umstrittenen Sprecherwechsels angesehen: Der Übernehmende beginnt mit angehobener Lautstärke und hohem F0 zu sprechen (im Vergleich zu nicht überlagernden Äußerungen) und senkt beide Parameter sofort ab, nachdem der andere Sprecher aufgehört hat zu reden. Der Sprecher, dem das Rederecht abgenommen werden soll, reagiert auf den Übernahmeversuch ebenfalls mit angehobener Lautstärke und oft mit einer Verlangsamung der Sprechgeschwindigkeit. French und Local gehen davon aus, dass diese Merkmalsbündel systematisch von Gesprächsteilnehmern produziert werden und konstitutive Merkmale von umstrittenen Turnübernahmen darstellen.

Die oben beschriebene zweifache Gliederung wird u.a. von Selting (1995) auf eine dreifache Unterteilung erweitert, die auch im weiteren Verlauf dieser Arbeit von Bedeutung ist. Hierbei trennt sie die turnkompetitiven (*competitive*) Übernahmen in zwei Kategorien auf. Zum einen die *Unterbrechung* als *nicht-legitime* Turnübernahme und zum anderen die Fortsetzung bzw. Übernahme eines *legitimen Turns*. Diese Unterscheidung beruht auf der Beobachtung, dass viele Turnübernahmen, die nicht an TRPs lokalisiert sind, nicht die von French und Local beschriebene Erhöhung der Tonhöhe zeigen und in der sequentiellen Analyse auch nicht als strittig empfunden werden. Selting geht davon aus, dass Sprecher am Beginn ihres Redebeitrags andere phonetische Mittel einsetzen, wenn sie die Übernahme als legitim empfinden, als wenn sie ohne jeden Anspruch auf das Rederecht den Turn übernehmen. Einen ihnen zustehenden Redebeitrag beginnen die Sprecher zwar mit angehobener Lautstärke, aber ohne angehobene Tonhöhe. Als *legitim* empfinden Sprecher nach Selting Einwürfe, die als direkte Entgegnung auf das unmittelbar Gehörte konzipiert sind. Auch scherzhafte Entgegnungen oder Korrekturen

des Vorgängerturns können als Gründe für eine legitime Turnübernahme angesehen werden. Nicht als legitim empfundene Übernahmen hingegen werden neben der angehobenen Lautstärke auch mit angehobenem F0 produziert. Diese Argumentation im Rahmen der CA ist nicht unproblematisch, da aufgrund phonetischer Kriterien auf die Intention der Sprechenden geschlossen wird und die interpretierte Intention später als Begründung für die vorgefundenen phonetischen Strukturen herangezogen wird. Dennoch ist Seltings Klassifikation m.E. geeignet, um verschiedene Arten von Sprecherwechseln zu klassifizieren und die zugrunde liegenden Einstellungen der Sprecher zu erklären.

In der vorliegenden Arbeit wird deshalb an einer Klassifikation gearbeitet, die folgende Arten von Sprecherwechseln trennt: *gezielte Turnübergaben*, *erlaubte Turnübernahmen* und *umstrittene Turnübernahmen*. Bei dieser Klassifikation werden zwei Perspektiven berücksichtigt: die des Turninhabers und die des Turnübernehmenden. *Gezielt* bedeutet, dass das Rederecht durch die Initiative des Sprechers an einer bestimmten Stelle zum Hörer abgegeben wird. Mit gezielten Turnübergaben sind in der Regel bestimmte Aktivitätstypen verbunden z.B. Fragen oder abgeschlossene Antworten auf Fragen. Hier ist der Inhaber des Rederechts der aktive Partner und reicht das Rederecht an den Partner weiter. Bei der *erlaubten* Turnübernahme verschiebt sich die Initiative für die Turnübernahme zum momentanen Hörer. Dieser ergreift das Rederecht, ohne dass dies durch die typischen Merkmale der Abschlussfunktion vom vormaligen Sprecher eingeleitet wird. Oft kommt es zu einem kurzen Abschnitt überlagerndem Sprechens. Das Einverständnis des momentanen Sprechers mit der Übernahme wird dadurch signalisiert, dass dieser sich schnell in die Hörerrolle begibt und keine Versuche macht, das Wort zu halten. In einigen Fällen ist zu beobachten, dass Hörer das Wort dann ergreifen, wenn der Sprecher in seiner Rede ins Stocken gerät. Im Falle der *erzwungenen* Turnübernahme fällt der Hörer dem Sprecher ins Wort, ohne dass dieser phonetisch oder inhaltlich einen Übergabewunsch signalisiert hat. Es kommt zu einer längeren Zeitspanne mit überlappendem Sprechen - dem Kampf ums Rederecht. Im Falle der erzwungenen Turnübernahme ist nicht jeder Versuch, das Rederecht zu bekommen, erfolgreich, deshalb müssen die erfolgreichen Übernahmeversuche getrennt von den gescheiterten Versuchen betrachtet werden.



Audiobeispiele zu den Kategorien *gezielte*, *erlaubte* und *umstrittene* Sprecherwechsel: 2.1 – 2.13

### 4.1.3 Perzeptionsexperimente

Die bisher zusammengefassten Ergebnisse aus der Literatur gehen zum größten Teil auf die gesprächsanalytische Interpretation von Produktionsdaten zurück. Untersuchungen dieser Art führen in der Regel nicht zu quantifizierbaren Ergebnissen und aufgrund der methodischen Konzeption der CA werden in diesem Umfeld keine Perzeptionsexperimente durchgeführt, die systematisch Hörerurteile mit phonetischen Strukturen verbinden. Aber auch außerhalb der gesprächsanalytischen Forschung sind experimentelle Untersuchungen zur Dialogsteuerung selten. Die Ursache hierfür liegt möglicherweise darin, dass neben phonetischen Faktoren auch der linguistische und situative Kontext entscheidenden Einfluss auf den Dialogverlauf hat und eine systematische Erfassung dieser Variablen in einem Experiment kaum zu leisten ist.

Trotz dieser generellen Problematik hat Caspers (2000) ein Experiment zum Turntaking im Holländischen konzipiert und durchgeführt, das eine Kombination aus der Verwendung natürlichsprachlicher Stimuli und deren funktionaler Beurteilungen im Experiment ermöglicht. Dieses Experiment wird von Wichmann & Caspers (2001) für das Englische reproduziert. Da in den Experimenten zu den beiden Sprachen sowohl das generelle Vorgehen als auch die Ergebnisse übereinstimmen, wird hier nur auf das Experiment zum Englischen im Detail eingegangen. Der Ausgangspunkt für die Experimente ist, dass nach Ford & Thompson (1996) ein Sprecherwechsel im Dialog dann eintreten kann, wenn melodische, syntaktische und pragmatische Komplettierung zusammenfallen. Wichmann & Caspers (2001) überprüfen diese Hypothese, indem sie Dialogfragmente unterschiedlicher phonetischer und syntaktischer Ausprägung in zwei Experimenten hinsichtlich ihrer Funktion für die Dialogsteuerung beurteilen lassen. Für den ersten Teil des Experiments isolieren die Autorinnen eine Reihe von 8-13 Sekunden langen Fragmenten aus südeinglischen Maptask-Dialogen (Anderson et al. 1991). Alle Fragmente haben die Eigenschaft, dass vor deren Beginn und nach deren Ende im Turn eine Pause auftrat. Die Fragmente sind also sog. 'interpausal units' (Koiso et al. 1998). Weiterhin werden die Stimuli so ausgewählt, dass folgende Parameter in systematischer Variation im Stimulusmaterial vertreten sind:

- syntaktische Abgeschlossenheit vs. syntaktische Unabgeschlossenheit an der Grenze des Fragments,
- nachfolgender Sprecherwechsel vs. Weitersprechen des aktuellen Sprechers im Ursprungsdialog.

Aus der Kombination dieser Kriterien ergeben sich vier Bedingungen hinsichtlich der syntaktischen Abgeschlossenheit und des weiteren Dialogverlaufs. Für jede der vier Bedingungen werden (soweit vorhanden) aus dem Korpus jeweils zwei Ausschnitte herausgesucht, die mit einem der folgenden fünf Intonationsmuster enden:

- 'high rise' (H\* H%)
- 'high level' (H\* %)
- 'fall-rise' (H\*L H%)
- 'fall' (H\*L L%)
- 'truncated fall' (H\*L %)

Bei der Auswahl der Stimuli wird darauf geachtet, dass die Fragmente im Ursprungsdialog keine syntaktisch markierte Frage an den Gesprächspartner darstellen. Im Fall der hoch steigenden Kontur gibt es im Korpus nicht genügend Fälle, die den genannten Kriterien entsprechen. Hier wird deshalb nur mit sechs Ausschnitten gearbeitet (zwei in deklarativer, zwei in interrogativer Funktion und zwei syntaktische Fragmente). Insgesamt kommen so 38 Stimuli zusammen (statt 40 Stimuli bei kompletter Symmetrie).

Die Hörer hören jedes Fragment zweimal und beurteilen dann, wie der Dialog, aus dem das Fragment stammt, nach dem vorgespielten Ausschnitt vermutlich weitergehen wird. Es gibt drei Antwortkategorien:

- Der Sprecher spricht weiter.
- Der Sprecher spricht nach einer kurzen Bestätigung durch den Hörer weiter ('Rezeptionssignal').
- Der Turn wird vom Dialogpartner übernommen.

Die Ergebnisse des Hörtests zeigen, dass die Beurteilung der Funktion einer prosodischen Grenze stark von der syntaktischen Abgeschlossenheit an der Stelle des prosodischen Einschnitts abhängig ist. Wenn die präsentierte Äußerung syntaktisch nicht vollständig ist, vermuten die Versuchspersonen in insgesamt 98% der Urteile, dass der Sprecher weitersprechen wird. In 24% der Stimuli wird hierbei angenommen, dass dem Weitersprechen ein Rezeptionssignal vorangehen könnte. Eine systematische Auswirkung der unterschiedlichen Intonationskonturen auf die Urteile lässt sich nicht erkennen.

Bei syntaktischer Abgeschlossenheit wird mit einer relativen Häufigkeit von 62% geurteilt, dass derselbe Sprecher weitersprechen wird (hiervon 30% mit vorangehendem Rezeptionssignal). Die Urteile zeigen hierbei eine klare Präferenz des Urteils 'Sprecher spricht weiter' wenn das Fragment mit H\* % endet (89%). Weiterhin ist interessant, dass bei syntaktischer Vollständigkeit für die Muster H\* H%, H\*L L% und H\*L H\* jeweils etwa die Hälfte der Urteile für einen unmittelbaren Sprecherwechsel abgegeben werden und die andere Hälfte sich auf die anderen beiden Kategorien verteilt. Das Experiment zeigt also erstens die primäre Wichtigkeit der syntaktischen Struktur bei der Beurteilung des Dialogfortgangs und zweitens die fehlende Eindeutigkeit bei der Form-Funktions-Zuordnung melodischer Muster durch Hörer. Die einzige Ausnahme bildet der Verlauf H\* %, der als klares Signal des Turnhaltens aufgefasst wird. Vor allem ist auffällig, dass auch das tief fallende Muster H\*L L% nur in etwa der Hälfte der Urteile als Turnabschluss beurteilt wird.

Im zweiten Teil des Experiments werden alle Stimuli aus dem ersten Experiment einbezogen, die einen nachfolgenden Sprecherwechsel haben. Die Stimuli werden den Versuchspersonen zusammen mit dem Anfang der Folgeäußerung zusammen präsentiert, und es ist Aufgabe der Probanden zu beurteilen, ob der gehörte Sprecherwechsel vom ersten Sprecher erwartet war. Drei Antwortkategorien stehen zur Auswahl:

- Sprecher erwartet den Wechsel,
- Sprecher erwartet weiterzusprechen,
- Unklarheit über die Erwartung des Sprechers.

Auch bei dieser Frage orientieren sich die Versuchspersonen vor allem an der syntaktischen Struktur. Bei nicht abgeschlossener Syntax werden nur 15% der Turn-

übergänge als erwartet eingestuft, während es bei den syntaktisch kompletten Teil-äußerungen 68% sind. Unter den Turnübergängen mit syntaktischem Abschluss sticht wieder der H\* % Verlauf heraus, der nur mit 4% als Ende eines erwarteten Übergangs und mit 74% als Ende eines nicht erwarteten Übergangs beurteilt wird. H\* % zeigt also auch hier eine relativ klare Funktion als Turnhaltesignal. Ein überraschendes Ergebnis ist, dass bei den Übergängen mit syntaktischem Abschluss bei allen anderen Mustern als H\*% mindestens 70% der Urteile für den erwarteten Übergang abgegeben werden. Das bedeutet, dass jede dieser Konturen als adäquate turnfinale Kontur für erwartete Turnübergänge angesehen wird. Diese Ergebnisse zeigen deutlich die Abhängigkeit der Urteile von der syntaktischen Struktur am Äußerungsende und die weitgehende Uneindeutigkeit der funktionalen Einordnung der Intonationsmuster.

Die Experimente haben allerdings einige Nachteile. Zum einen kann aufgrund der Stimulusgenerierung der semantisch-pragmatische Kontext der getesteten Äußerungen nicht konstant gehalten werden, obwohl er möglicherweise in der Beurteilung durch die Versuchspersonen eine Rolle spielt. Zweitens fließen in das Urteilsverhalten der Versuchspersonen unter Umständen auch noch andere phonetische Parameter als die Intonationskontur ein, ohne dass diese Parameter kontrolliert werden können (z.B. Phonationstyp, segmentelle Dauerstruktur und Intensitätsverlauf).

Im Abschnitt 4.3 wird die Konzeption und Durchführung eines Experiments zum Deutschen beschrieben, das auf den Untersuchungen von Wichmann und Caspers aufbaut, aber den beschriebenen methodischen Nachteilen durch verschiedene Modifikationen in Aufbau und Durchführung entgegenwirkt.

Die folgenden Abschnitte beschreiben die unterschiedlichen Arbeitsschritte, die im Rahmen dieser Arbeit unternommen werden, um korpusbasiert und durch Perzeptionsexperimente einerseits die phonetischen Exponenten der Abschlussfunktion von denen der Weiterweisungsfunktion zu trennen und andererseits funktional unterschiedliche Typen von Sprecherwechseln zu differenzieren.

## 4.2 Symbolbasierte Analyse der Produktionsdaten

In diesen Abschnitt werden die Ergebnisse der automatischen labelbasierten Korpusabfrage der prosodischen Merkmale turnfinaler prosodischer Grenze dargestellt. Die meisten Analysen beziehen sich auf die Daten aus dem LINDENSTRASSE-Korpus, da hier im Gegensatz zum Terminabsprachekorpus aufgrund des technischen Aufbaus die Aufnahme von überlagerndem Sprechen möglich war und dadurch hinsichtlich der Turnübergänge realitätsnahe Daten erhoben werden konnten.

Die automatisierten Korpusabfragen prosodischer Merkmale an turnfinalen prosodischen Grenzen umfassen die Parameter 'melodischer Verlauf', 'Längung des phrasenfinalen Reims', 'Pause' bzw. 'Überlagerung zwischen aufeinanderfolgenden Turns' und 'phonatorische Veränderungen'. Das Merkmal 'Reset' spielt in den Analysen keine Rolle, da nach einer turnfinalen prosodischen Grenze der aktuelle Sprecher aufhört zu sprechen und somit kein Bezugspunkt für einen nachfolgenden Reset gegeben ist. Folgende globale Hypothesen werden überprüft:

### **Hypothese 1**

Turnfinale Merkmalskomplexe an Äußerungsgrenzen unterscheiden sich von den Merkmalen turninterner prosodischer Grenzen.

### **Hypothese 2**

Die phonetischen Merkmale turnfinaler prosodischer Grenzen unterscheiden sich, je nachdem, ob es sich um Turnübergänge mit oder ohne überlagerndes Sprechen handelt.

Diese Hypothesen werden für die unterschiedlichen Analysebereiche in spezifischere Teilhypothesen untergliedert.

## **4.2.1 Einzelmerkmale an turnfinalen prosodischen Grenzen**

### **4.2.1.1 Pausen und Atmungsphasen**

Pausen (und Atmungsphasen) zwischen den Redebeiträgen verschiedener Dialogpartner sind etwas prinzipiell anderes als Pausen an prosodischen Grenzen im Inneren von Redebeiträgen. Ob eine Pause auftritt und wie lang diese Pause ist, wird im Inneren von Turns vom derzeitigen Sprecher bestimmt und steht in Zusammenhang mit der inhaltlichen Gliederung des Gesprochenen, mit Störungen im Sprechfluss und mit der physiologischen Notwendigkeit des Luftholens. Pausen zwischen den Redebeiträgen verschiedener Sprecher im Dialog sind hingegen eine Folge der Interaktion zwischen den Gesprächspartnern. Sie können nur entstehen, wenn Sprecher A aufhört zu sprechen und Sprecher B nach einer Pause das Rederecht übernimmt. Unter diesem Gesichtspunkt können drei Arten von Turnübergängen unterschieden werden. Erstens diejenigen mit einer Pause zwischen den Redebeiträgen der Sprecher. Zweitens Übergänge zwischen Redebeiträgen, in denen Sprecher B beginnt zu sprechen, bevor Sprecher A seine Äußerung beendet hat und es zu gleichzeitigem (überlagerndem) Sprechen kommt. Der dritte zu unterscheidende Fall liegt in der Mitte zwischen den Turnübergängen mit Pause und den Turnübergängen mit Überlagerung. Hier kommt es zu einem unmittelbaren Anschluss der Rede von Sprecher B an den Redebeitrag von Sprecher A, so dass weder eine Pause noch eine Überlagerung entsteht. Da sich bei einer präzisen akustischen Analyse der Übergänge zwischen den Sprachsignalen der Sprecher jedoch praktisch immer entweder eine Pause oder eine Überlagerung zeigt, wird für die dritte Kategorie ein oberer Grenzwert von 100 ms Pausen- bzw. Überlagerungsdauer gewählt. Dieser Wert liegt im Bereich der Dauer von einzelnen Segmenten, und eine informelle auditive Analyse zeigt, dass Pausen und Überlagerungen unter 100 ms als unmittelbarer Anschluss wahrgenommen werden. Mit einem Grenzwert von 100 ms (allerdings für Pausen innerhalb von Redebeiträgen) arbeiten auch Koiso et al. (1998) und Wichmann & Caspers (2001).

Um die unterschiedlichen Synchronisationen der Redebeiträge im Korpus messen zu können, werden in einer manuellen Etikettierung sämtliche Überlagerungen und Pausen an Turnübergängen durch Zeitmarken abgegrenzt. In der späteren Verarbeitung

der Etikettierdateien kann aus diesen Etiketten die Dauer von Pausen bzw. Überlagerungen berechnet und in die automatisch generierten Merkmalsbündel integriert werden (siehe 2.1.3.5 (11) und 2.1.3.5 (12)).

In den Daten des LINDENSTRASSE-Korpus können für insgesamt 377 turnfinale prosodische Grenzen Merkmalsbündel abgeleitet werden, die Informationen über die Längung des finalen Reims, das melodische Muster am Phrasenende und die Dauer von Überlagerungen oder Pausen zwischen den Turns enthalten. Die Verteilung der Sprecherwechsel über die drei oben beschriebenen Kategorien von Turnübergängen im LINDENSTRASSE-Korpus sehen folgendermaßen aus:

- Sprecherwechsel mit Pause oder Atmen zwischen den Redebeiträgen: 183 Fälle (48,5%). Die durchschnittliche Pausendauer beträgt 704 ms mit einer Standardabweichung von 702 ms.
- Sprecherwechsel mit unmittelbarem Anschluss (+-100 ms): 66 Fälle (17,5%).
- Sprecherwechsel mit Überlagerung: 128 Fälle (34%). Die durchschnittliche Dauer der Überlagerungen liegt bei 480 ms mit einer Standardabweichung von 375 ms.

Diese drei akustisch definierten Klassen, die sich in der temporalen Koordination der Turnübergänge unterscheiden, geben einen ersten Ansatz für die korpusbasierte Differenzierung funktional verschiedener Arten von Sprecherwechseln. Es wird angenommen, dass sich in der Klasse der überlagernden Übergänge vor allem Sprecherwechsel finden, die durch den Hörer eingeleitet werden und den Sprecher unterbrechen, ohne dass dies von ihm intendiert ist. Bei den Übergängen mit Pausen und mit unmittelbarem Anschluss hingegen kann eher davon ausgegangen werden, dass der Sprecher zu einem Abschluss gekommen ist, dies phonetisch signalisiert und dem Hörer das Wort überlässt. Unter diesem Gesichtspunkt wird in den folgenden Darstellungen der Merkmale und Merkmalsbündel an turnfinalen prosodischen Grenzen zwischen diesen drei Arten von Sprecherwechseln unterschieden, um zu untersuchen, ob sich andere phonetische Merkmale in Abhängigkeit von der Art der Synchronisierung der aufeinander folgenden Turns unterschiedlich ausprägen. Ist dies der Fall, können Hinweise auf interaktional unterschiedliche Funktionen der drei Arten von Turnübergängen abgeleitet werden.



Audiobeispiele: Sprecherwechsel mit Pause: 2.1, Sprecherwechsel mit unmittelbarem Anschluss: 2.6, Sprecherwechsel mit Überlagerung: 2.9

Ein weiterer Aspekt in der Analyse von Atmungsaktivität ist der interaktional funktionalisierte Einsatz von deutlich hörbaren Einatmungsgeräuschen. Während der interpretativen Analyse der LINDENSTRASSE-Dialoge und der Aufbereitung der Audio-beispiele wurde deutlich, dass viele Sprecher ihren Wunsch zu Sprechen durch "lautes" Einatmen signalisieren. Dies gilt sowohl für Turninhaber, die ihren Turnhalteanspruch an prosodischen Grenzen anzeigen wollen, als auch für Hörer, die von der Hörer- in die Sprecherrolle wechseln möchten. Im Rahmen der automatischen Merkmalsabfrage kann allerdings nicht zwischen "lautem" und "leisem" Atmen getrennt werden. Es wurde an jeder Position im Dialog das Vorhandensein von Atmen etikettiert, wo es für den Etikettierer über Kopfhörer wahrnehmbar war. Welche dieser hörbaren Atmungsphasen eine dialogsteuernde Funktion haben, kann also nur über eine spezifische interpretative Bearbeitung der Daten geklärt werden. Eine solche Analyse soll hier nicht vorgenommen werden.



Audiobeispiele: Atmen als Turnübernahmesignal: 1.1, 2.9, Atmen als Turhaltesignal: 1.1, 1.3

#### **4.2.1.2 Melodische Muster**

Funktionale Unterschiede zwischen den drei unterschiedlich synchronisierten Turn-übergängen müssten sich vor allem durch die Verteilungen der melodischen Muster an den Turnenden zeigen. Wenn sich in der Klasse der überlagernden Sprecherwechsel tatsächlich mehr unterbrechende und damit nicht vom Sprecher gezielt eingeleitete Sprecherwechsel befinden als in den anderen beiden Klassen, ist in dieser Klasse mit einer geringeren Anzahl starker melodischer Bewegungen am Phrasenende zu rechnen. Insbesondere ist ein niedrigerer Anteil tief fallender Konturen zu erwarten, da diese auf einen gezielten Abschluss hindeuten. Beobachtungen dieser Art finden sich bei Jefferson (1986) zum amerikanischen Englisch. Hier wird ebenfalls ein Vergleich zwischen den melodischen Mustern bei überlagernden und nicht-überlagernden Sprecherwechseln

vorgenommen. Es werden aus beiden Klassen je 800 Sprecherwechsel analysiert. Die Autorin nimmt hierbei eine Unterteilung der Intonationsmuster in *completion intonation* und *non-completion intonation* vor. Die gefundene Verteilung zeigt, dass 70% der Sprecherwechsel ohne Überlagerung mit *completion intonation* einhergehen, während es bei den überlagernden Sprecherwechseln nur 25% sind. Leider gibt die Autorin weder an, welches Korpus der Untersuchung zugrunde liegt, noch werden Angaben über die melodischen Muster gemacht, die als *completion intonation* bzw. als *non-completion intonation* angesehen werden.

Tabelle 19 zeigt die phrasenfinalen melodischen Muster turnfinaler prosodischer Grenzen im LINDENSTRASSE-Korpus. Alle hier untersuchten Phrasengrenzen sind syntaktisch abgeschlossen und zeigen keine Zögerungslängung am Ende der Phrase.

**Tabelle 19:** Melodische Verläufe am Turnende (LINDENSTRASSE).

	<b>Pause (n=183)</b>	<b>Unmittelbarer Anschluss (n=66)</b>	<b>Überlagerung (n=128)</b>	<b>Gesamt (n=377)</b>
<b>Starke melodische Bewegung</b>				
Tief fallend	53 (29,0%)	23 (34,9%)	35 (27,3%)	111 (29,4%)
Tief fallend + sehr leicht steigend	12 (6,6%)	4 (6,1%)	5 (3,9%)	21 (5,6%)
Fallend-steigend	12 (6,6%)	6 (9,1%)	5 (3,9%)	23 (6,1%)
Hoch steigend	13 (7,1%)	4 (6,1%)	3 (2,3%)	20 (5,3%)
<b>Gesamt starke melodi- sche Bewegung</b>	<b>90 (49,2%)</b>	<b>37 (56,1%)</b>	<b>48 (37,5%)</b>	<b>175 (46,4%)</b>
<b>Schwache melodische Bewegung</b>				
Eben	25 (13,7%)	7 (10,6%)	19 (14,8%)	51 (13,5%)
Leicht fallend	20 (10,9%)	9 (13,6%)	34 (26,6%)	63 (16,7%)
Leicht bis mittelhoch steigend	37 (20,2%)	9 (13,6%)	11 (8,6%)	57 (15,1%)
Eben + sehr leicht stei- gend	3 (1,6%)	2 (3,0%)	5 (3,9%)	10 (2,7%)
Leicht fallend + sehr leicht steigend	8 (4,4%)	2 (3,0%)	11 (8,6%)	21 (5,6%)
<b>Gesamt schwache mel. Bewegung</b>	<b>93 (50,8%)</b>	<b>29 (43,9%)</b>	<b>80 (62,5%)</b>	<b>202 (53,6%)</b>
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>377 (100%)</b>

In allen drei Klassen ist die tief fallende Kontur am häufigsten vertreten. Die Rangliste der häufigsten Konturen in den drei Klassen zeigt aber einen deutlichen Unterschied der Verteilungen: Während in den Klassen mit Pause und unmittelbarem Anschluss neben den tief fallenden Konturen meistens leicht bis mittelhoch steigende und ebene Konturen vorkommen, treten bei überlagernden Übergängen neben den tief fallenden Konturen vor allem leicht fallende Konturen auf. Alle starken melodischen Bewegungen zusammengekommen zeigen eine Abnahme von Turnübergängen mit unmittelbarem Anschluss (56,1% starke melodische Bewegungen), über Übergänge mit Pause (49,2%), hin zu den überlagernden Übergängen mit nur 37,5% starken melodischen Bewegungen. Ein Chi-Quadrat-Test zeigt, dass die Verteilung zwischen den drei Klassen signifikant voneinander abweicht ( $F=7,12$ ,  $p=0,02$ )

Tabelle 20 zeigt zum Vergleich die turnfinalen melodischen Muster im Terminabsprachekorpus. Dieser Vergleich ist insofern interessant, als jeder Turnübergang in diesen Daten als gezielter Turnabschluss durch den aktuellen Sprecher eingeleitet wird – also als intendiert angesehen werden kann. Dies ist eine Folge des technischen Aufbaus der Aufnahmen: Die Sprecher konnten hierbei nur wechselseitig sprechen, solange sie einen Knopf gedrückt hielten. Das Halten des Knopfs blockierte den Kanal des Partners. Wenn ein Sprecher also seinen Turn für beendet hielt, wurde der Knopf losgelassen und der Partner konnte seinen Knopf drücken und seinen Redebeitrag beginnen. Dieses technische Szenario ist zwar unnatürlich, bietet aber durch die Aufnahmesituation ein externes Kriterium, das es zulässt, jedes Turnende als gezielten Abschluss zu klassifizieren. In den Daten mit überlagerndem Sprechen ist dies anders, denn der Hörer kann hier jederzeit das Wort übernehmen (z.B. in einer kurzen Sprechpause des Sprechers). Die Entscheidung, ob eine Übernahme vom Sprecher erwünscht war, ist damit immer eine Frage der Interpretation.

**Tabelle 20:** Melodische Verläufe am Turnende (Terminabsprachen).

<b>Starke melodische Bewegung</b>	
Tief fallend	536 (49,1%)
Tief fallend + sehr leicht steigend	161 (14,7%)
Fallend-steigend	46 (4,2%)
Hoch steigend	112 (10,3%)
<b>Gesamt starke melodische Bewegung</b>	<b>855 (78,3%)</b>
<b>Schwache melodische Bewegung</b>	
Eben	48 (4,4%)
Leicht fallend	45 (4,1%)
Leicht bis mittelhoch steigend	117 (10,7%)
Eben + sehr leicht steigend	6 (0,5%)
Leicht fallend + sehr leicht steigend	21 (1,9%)
<b>Gesamt schwache melodische Bewegung</b>	<b>237 (21,6%)</b>
<b>Gesamt</b>	<b>1092 (100%)</b>

Der Vergleich von Tabelle 19 und Tabelle 20 zeigt, dass die Verteilungen der melodischen Muster in den beiden Korpora starke Unterschiede aufweisen. Etwa die Hälfte aller turnfinalen Muster in den Terminabsprachedaten sind tief fallend und fast 80% haben starke melodische Bewegungen. In den LINDENSTRASSE-Daten gibt es über alle Kategorien knapp 30% fallende Muster und insgesamt 46,4% starke melodische Bewegungen. Die schwachen melodischen Bewegungen spielen bis auf die leicht bis mittelhoch steigenden Konturen in den Terminabsprachedaten praktisch keine Rolle, während sie in den LINDENSTRASSE-Daten zusammen fast 30% aller gefundenen Verläufe ausmachen. Die Zahlen zeigen somit deutlich, dass bei Sprecherwechseln, die vom Sprecher intendiert sind, in den meisten Fällen starke melodische Bewegungen am Turnende eingesetzt werden. Da in den Daten aus dem Videotaskkorpus in allen drei akustisch definierten Klassen von Turnübergängen wesentlich weniger starke melodische Bewegungen als in den Terminabsprachedaten vorkommen, ist anzunehmen, dass jede der drei Klassen sowohl intendierte als auch nicht intendierte Sprecherwechsel enthält. Das bedeutet, dass das Kriterium von Pause vs. Überlagerung nicht hinreichend ist, eine funktionale Differenzierung intendierter vs. nicht intendierter Sprecherwechsel zu tragen, obwohl sich eine Tendenz in diese Richtung zeigt.



Audiobeispiele zu melodischen Mustern an Turnübergängen: 2.1 – 2.13

### 4.2.1.3 Segmentelle Längung

Das unter 3.2.1.2 beschriebene Messverfahren erlaubt Aussagen über das Vorhandensein von segmenteller Längung an prosodischen Grenzen. Tabelle 21 zeigt wie häufig die turnfinalen Reime in den drei Klassen gelängt sind.

**Tabelle 21:** Segmentelle Längung des turnfinalen Reims (LINDENSTRASSE).

	<b>Pause (n=183)</b>	<b>Unmittelbarer Anschluss (n=66)</b>	<b>Überlagerung (n=128)</b>
<b>Kürzung</b>	1 (0,5%)	-	2 (1,5%)
<b>Weder Längung noch Kürzung</b>	47 (25,7%)	18 (27,3%)	40 (31,3%)
<b>Längung</b>	135 (73,8%)	48 (72,7%)	86 (67,2%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>

Im Gegensatz zu den melodischen Konturen gibt es keine klare Abhängigkeit zwischen den drei differenzierten Klassen von Turnübergängen und dem Auftreten von finaler Längung. In allen drei Klassen liegt der Anteil von gelängten Reimen bei rund 70%. Die Daten zeigen lediglich einen leichten Trend zu weniger Längung im Falle der Übergänge mit Überlagerung. Die gemessenen Werte sind den Ergebnissen der turnfinalen prosodischen Grenzen im Terminabsprachekorpus sehr ähnlich. Hier wurde finale Längung in 74,3% der Fälle gemessen (Peters et al. 2005b).

Die Zahlen zeigen somit, dass auch bei den überlagernden Übergängen üblicherweise eine Längung des finalen Reims auftritt, also Sprecher, selbst wenn sie unterbrochen werden, am Ende ihrer Äußerungen üblicherweise finale Längung produzieren. Somit ist der Faktor Längung im Rahmen einer funktionalen Differenzierung kein geeignetes Klassifikationsmerkmal.

Eine Erklärung, warum in allen drei Klassen in rund einem Drittel der Fälle keine Längung des finalen Reims gemessen werden kann, liegt bisher nicht vor.

#### 4.2.1.4 Phonation

Neben den oben untersuchten Merkmalen finden sich in der Literatur Hinweise auf phonatorische Veränderungen am Ende prosodischer Phrasen. Die vorliegende Untersuchung soll zeigen, welche Rolle Knarrstimme an turnfinalen prosodischen Grenzen im LINDENSTRASSE-Korpus spielt.

Aus Untersuchungen zum Englischen (Local & Kelly 1986; Nakatani & Hirschberg 1994) ist bekannt, dass Knarrstimme an prosodischen Grenzen auftreten kann. Die Autoren beschreiben zwei funktional unterschiedliche Bereiche, in denen dieser Phonationstyp eingesetzt wird:

- Unvollständige, vom Sprecher abgebrochene Äußerungen enden z.T. mit Knarrstimme. In solchen Fällen folgt häufig ein Glottalverschluss, nach dessen Ende der Sprecher seine Äußerung fortsetzt. In diesem Kontext wird die Glottalisierung als phonetisches Signal eingesetzt, um anzuzeigen, dass der Sprecher mit seinem Redebeitrag noch nicht zum Ende gekommen ist und beabsichtigt, die abgebrochene Äußerung fortzusetzen. Die Funktion der nicht-modalen Phonation ist hier also, den Anspruch auf das weitere Rederecht anzuzeigen. Local und Kelly bezeichnen diese Art der Glottalisierung deshalb als *turn-holding-device*, Nakatani und Hirschberg als *interruption glottalization*, die als *edit-signal* das Ende eines Reparandums markiert.
- Der zweite Bereich, in dem oft Knarrstimme vorkommt, ist am Ende kompletter prosodischer Phrasen, die mit terminaler Intonation (Absinken der Grundfrequenz in den untersten Bereich der Sprechstimme) produziert werden. Hier ist die Funktion gerade umgekehrt. Der Sprecher zeigen durch dieses sprachliche Verhalten an, dass sie zum Ende einer Äußerung gekommen sind und der Gesprächspartner das Rederecht bekommen soll. Folglich wird diese Verwendung der Knarrstimme von Local und Kelly als *turn-yielding-device* bezeichnet.



Audiobeispiele: Knarrstimme und Glottalverschluss an syntaktischem Abbruch: 3.8,

Knarrstimme an Turnabschluss: 2.2, 2.7

Weder Local und Kelly noch Nakatani und Hirschberg treten allerdings den Nachweis über die perzeptorische Relevanz der phonatorischen Veränderungen an. Sie leiten die postulierten Funktionen lediglich daraus ab, dass die unterschiedlichen Typen von Knarrstimme in den untersuchten Produktionsdaten komplementär verteilt auftreten. Aus unterschiedlichen Verteilungseigenschaften allein kann m.E. aber noch keine perzeptorische und funktionale Relevanz abgeleitet werden. Aus diesem Grund wird in den Perzeptionsexperimenten 2 und 3a gezielt die funktionale Zuordnung phonatorischer Veränderungen geprüft (siehe 4.3 und 5.2.1).

Kohler, Peters & Wesener (2005) untersuchen 22 Sprecher und Sprecherinnen des *Kiel Corpus of Spontaneous Speech* Vol. I und II auf vergleichbare phonetische Strukturen. Es werden durch den Vergleich von zwei Datensätzen die von Local und Kelly bzw. Nakatani und Hirschberg beschriebenen Verwendungen von Knarrstimme auch in den deutschen Daten nachgewiesen. Der erste Datensatz enthält etwa 300 prosodische Phrasen, die mit einem syntaktischen Abbruch enden, der nicht durch eine Unterbrechung durch den Dialogpartner hervorgerufen wurde. Der zweite Datensatz beinhaltet über tausend prosodische Grenzen ohne syntaktische Abbrüche, die mit terminaler Intonation realisiert sind. Die Sprachsignale beider Datensätze werden anhand der segmentellen und prosodischen Etikettierung aus dem Gesamtkorpus extrahiert. Für diese zwei Datensätze wird das Ende jeder extrahierten prosodischen Phrase hinsichtlich des Vorkommens von Knarrstimme und/oder Glottalverschlüssen manuell etikettiert.

Die Auswertung der Etikettierung zeigt, dass an prosodischen Grenzen mit syntaktischen Abbrüchen in 27% der Fälle Knarrstimme vorkommt, wobei diese in etwa einem Drittel der Fälle in Verbindung mit einem nachfolgenden Glottalverschluss steht. Die durchschnittliche Dauer des glottalisierten Signalabschnitts liegt bei 76 ms. Die phonetische Realisierung der prosodischen Grenzen unterliegt hierbei starken individuellen Schwankungen. Der Datensatz mit den tief fallenden Intonationsmustern zeigt eine Kopplung dieses melodischen Musters mit Knarrstimme in insgesamt 32% der Fälle, hier allerdings in weniger als einem Prozent der Fälle in Verbindung mit einem Glottalverschluss. Die durchschnittliche Dauer der nicht-modalen Phonation beträgt 164 ms. Auch bei der Auftretenshäufigkeit von Laryngalisierung gibt es eine deutliche Sprecherabhängigkeit. Eine weitere Differenzierung der tief fallenden Intonationsmuster mit Knarrstimme bezüglich ihrer Position in der Äußerung liefert ein sehr deutliches

Ergebnis: In turnfinaler Position findet sich geknarrte Stimme in 57% der Fälle mit tief fallender Intonation, in turninterner Position hingegen werden nur 27% der tief fallenden melodischen Muster von Knarrstimme begleitet.

Zusammenfassend lassen sich also zwei Unterschiede zwischen Knarrstimme an Abbruchstellen und an prosodischen Grenzen mit terminaler Intonation feststellen. An Abbruchstellen alterniert die Knarrstimme mit Glottalverschlüssen, was bei terminaler Intonation kaum vorkommt. Außerdem ist die Dauer der Knarrstimme an Abbrüchen deutlich kürzer. Kohler et al. (2005) beschreiben weiterhin, dass die auditive Analyse vieler Fälle den Eindruck vermittelt, dass die Knarrstimme an Abbruchstellen eher fest und angespannt klingt, während sie am Ende terminaler Intonationsmuster weniger fest und entspannt klingt. Diese Beobachtungen stützen die Trennung in zwei Phänomenbereiche, erstens die *Abbruchglottalisierung (interruption glottalisation)* und zweitens die phrasenfinale bzw. turnfinale *Laryngalisierung (final laryngealisation)*.

Da Glottalisierung und Laryngalisierung vermutlich relevante phonetische Signale für die Dialogsteuerung sind, soll nun auch das LINDENSTRASSE-KORPUS auf das Vorkommen dieser phonatorischen Signale untersucht werden. Hier allerdings weniger in Bezug zu syntaktischen Abbrüchen, als vielmehr hinsichtlich der Funktion von Knarrstimme als phonetische Markierung von Turnabschlüssen und somit als Korrelat der Abschlussfunktion. Da über die akustischen Daten nicht zuverlässig das Vorkommen von Knarrstimme bestimmt werden kann, ist für diesen Teil der Korpusanalyse eine weitere manuelle Etikettierung notwendig, die nach folgenden Kriterien vorgeht: Für den Phänomenbereich der Knarrstimme in Verbindung mit tief fallenden Intonationskonturen wird im ersten Arbeitsschritt jedes Vorkommen des tief fallenden Musters anhand der prosodischen Etikettierung aus dem Korpus automatisch herausgesucht. Die Verbindung zwischen dem Auftreten von Laryngalisierung und tief fallenden melodischen Verläufen wird im Vorfeld durch das gezielte Abhören der Dialoge bestätigt. Hierbei zeigt sich, dass Laryngalisierungen fast nie in Verbindung mit anderen melodischen Konturen stehen.

Jedes Vorkommen tief fallender Intonation wird dann nach den unter 3.2.11 eingeführten Kriterien beurteilt.

**Tabelle 22:** Phonationstyp bei tief fallendem melodischen Verlauf in turnfinaler Position (LINDENSTRASSE).

	<b>Pause (n=53)</b>	<b>Unmittelbarer Anschluss (n=23)</b>	<b>Überlagerung (n=35)</b>
<b>Modalstimme</b>	30 (56,6%)	14 (60,9%)	27 (77,1%)
<b>Knarrstimme (unregelmäßig)</b>	15 (28,3%)	7 (30,4%)	4 (11,4%)
<b>Knarrstimme (regelmäßig)</b>	2 (3,8%)	0 (0%)	2 (5,7%)
<b>Interferenz oder Unsicherheit</b>	6 (11,3%)	2 (8,7%)	2 (5,7%)
<b>Glottalverschluss</b>	0 (0%)	0 (0%)	0 (0%)
<b>Gesamt</b>	<b>53 (100%)</b>	<b>23 (100%)</b>	<b>35 (100%)</b>

Tabelle 22 zeigt trotz der relativ geringen Anzahl von Fällen eine Tendenz zum verstärkten Vorkommen von Knarrstimme an Turnübergängen mit Pausen und mit unmittelbarem Anschluss, gegenüber Übergängen mit Überlagerungen. Möglicherweise ist dieses gehäufte Vorkommen darauf zurückzuführen, dass in den Klassen ohne Überlagerung mehr gezielte Turnübergaben vorkommen, wobei die Laryngalisierung eines der phonetischen Merkmale der Abschlussfunktion ist. Auch der Vergleich mit Laryngalisierungen in turninterner Position belegt diese Deutung, denn hier tritt unregelmäßige Knarrstimme nur in knapp 11% der Fälle auf (siehe 3.2.11, Tabelle 9).

Die durchschnittliche Dauer von Knarrstimme beträgt über die drei Klassen von Synchronisierungen gemittelt 152 ms und ist damit etwas länger als mit 121 ms in turninterner Position. Aus dieser durchschnittlichen Dauer kann abgeleitet werden, dass die phonatorische Veränderung meist auf den finalen Reim beschränkt ist. In keinem Fall wurde ein Glottalverschluss in turnfinaler Position festgestellt.

In den Daten aus dem Terminabspracheszenario wurden allerdings an 57% aller Turnabschlüsse Laryngalisierungen gefunden. Dies legt nahe (wie schon die unterschiedlichen Verteilungen der melodischen Muster), dass viele der Turnübergänge in den LINDENSTRASSE-Daten nicht explizit durch den Sprecher eingeleitet werden.

#### **4.2.1.5 Merkmalsbündel an turnfinalen prosodischen Grenzen**

Tabelle 23 fasst die Bündelungen der Merkmale 'starke tonale Bewegung' und 'Längung' für die drei Klassen von Übergängen zusammen.

**Tabelle 23:** Merkmalsbündel an turnfinalen prosodischen Grenzen (LINDENSTRASSE).

	<b>Mit Pause (n=183)</b>	<b>Ohne Pause (n=66)</b>	<b>Mit Überlagerung (n=128)</b>
<b>Kein extrahierbares Merkmal</b>	26 (14,2%)	7 (10,6%)	28 (21,9%)
<b>Nur Längung</b>	67 (36,6%)	22 (33,3%)	52 (40,6%)
<b>Nur starke tonale Bewegung</b>	22 (12,0%)	11 (16,7%)	14 (10,9%)
<b>Längung und starke tonale Bewegung</b>	68 (37,2%)	26 (39,4%)	34 (26,6%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>

Die gefundenen Merkmalsbündel zeigen nur wenig weitere Unterschiede zwischen den drei Klassen von Turnübergängen. Wie auch in den Tabellen 19 und 20, ist der deutlichste Unterschied zwischen den Klassen die geringere Häufigkeit starker melodischer Bewegungen in der Klasse von Turnübergängen mit Überlagerungen. Neben diesem Ergebnis fällt vor allem auf, dass die Anzahl von Fällen, in denen die automatische Auswertung der Etikettierung kein Merkmal extrahieren kann, für die Klasse der überlagernden Übergänge mit 21,9% mit Abstand am größten ist.

#### 4.2.1.6 Diskussion

Ob Hypothese 1 ("Turnfinale Merkmalskomplexe an Äußerungsgrenzen unterscheiden sich von den Merkmalen turninterner prosodischer Grenzen.") als bestätigt angesehen werden kann, ist maßgeblich davon abhängig, welche Vergleiche vorgenommen werden. Die Verteilung der melodischen Muster an turninternen prosodischen Grenzen (Tabelle 3) unterscheidet sich stark von der Verteilung der melodischen Muster an turnfinalen Grenzen im Terminabsprachekorpus (Tabelle 20). Dies würde für die Annahme der Hypothese sprechen. Vergleicht man allerdings die Verteilung in Tabelle 3 mit den Häufigkeiten melodischer Muster am Turnende in den LINDENSTRASSE-Daten (Tabelle 19), zeigen sich keine klaren Unterschiede. Auch die weiteren Merkmale (segmentelle Längung und Phonationstyp) lassen keine eindeutige Bestätigung von Hypothese 1 zu.

Die Auswertung der phonetischen Merkmale in den Labeldateien zeigt lediglich Tendenzen, die auf funktionale Klassen hinweisen. Die drei Klassen von Turnübergängen (mit und ohne Überlagerung bzw. mit unmittelbarem Anschluss) weisen so starke Überschneidungen der Merkmalsbündel an den Turnenden auf, dass die durchgeführte

Analyse allein keine zuverlässige Zuordnung bestimmter phonetischer Merkmalskomplexe zur Abschluss- bzw. Weiterweisungsfunktion erlaubt. Hypothese 2 ("Die phonetischen Merkmale turnfinaler prosodischer Grenzen unterscheiden sich, je nachdem, ob es sich um Turnübergänge mit oder ohne überlagerndes Sprechen handelt.") kann also nur im Ansatz bestätigt werden. Folglich kann durch die Ergebnisse auch keine Trennung zwischen gezielt vom Sprecher eingeleiteten Turnübergaben und vom Hörer herbeigeführten Turnübernahmen abgeleitet werden. Die Daten deuten eher darauf hin, dass in allen drei Klassen von Übergängen sowohl Turnübergänge enthalten sind, die vom Sprecher eingeleitet sind, als auch solche, die der Hörer forciert, indem er "unaufgefordert" zu sprechen beginnt (z.B. in einer kurzen Redepause).

Eine Erweiterung der Korpusabfragen könnte die Ergebnisse entscheidend verändern. Die bisherige labelbasierte Analyse der finalen melodischen Muster bezieht sich nur auf den melodischen Verlauf im Anschluss an die letzte satzakzentuierte Silbe vor der jeweiligen prosodischen Grenze. Es ist aber bekannt, dass die Synchronisierung des F0-Maximums nuklearer Gipfelkonturen bzw. des F0-Minimums nuklearer Talconturen unabhängig von der nachfolgenden tonalen Bewegung zu semantischen Differenzierungen führt (Kohler 1991b). Besonders deutlich wird dies bei einer Positionierung des Gipfels vor der satzakzentuierten Silbe (*früher Gipfel*), im Gegensatz zu einer Positionierung nach der satzakzentuierten Silbe (*später Gipfel*). Der frühe Gipfel bringt eine retrospektive, zusammenfassende und abschließende semantisch-pragmatische Bedeutungskomponente in die Äußerung, während der späte Gipfel eröffnend wirkt, die gegebene Information als "neue Information" charakterisiert und den Eindruck von Überraschtheit hereinbringen kann. Die hierzu durchgeführten Experimente beziehen sich allerdings nicht auf dialogbezogene interaktionale Funktionen. Es ist aber anzunehmen, dass verschiedenen Gipfeltypen auch in der Signalisierung von interaktionalen Abschlüssen bzw. von Weiterweisung funktional relevant sind. In weiteren Arbeiten sollte also der nukleare Akzent und der nachfolgende melodische Verlauf als globales Muster betrachtet werden, dessen Funktion sich aus verschiedenen melodischen und temporalen Parametern der Gesamtkontur ableitet.

Weiterhin fehlt in der phonetischen Analyse das Merkmal 'Intensitätsverlauf', das evtl. weitere wichtige Information für die Steuerung von Turnübergaben trägt (French & Local 1983; Local & Kelly 1986). Auch eine durchgehende Analyse sämtlicher

phonatorischer Merkmale wie Verhauchung und Registerwechsel könnte zu klareren Ergebnissen der Produktionsdatenanalyse führen.



Audiobeispiele: Intensitätsabnahme an Turnabschluss: 1.1, 2.2, 2.7, Verhauchung an Turnabschluss: 1.1, 2.2, Registerwechsel an Turnabschluss: 1,1, 2.7

Ein anderes Problem der durchgeführten Korpusabfrage ist, dass nicht systematisch zwischen Turnenden, die mit einer konversationellen Frage an den Gesprächspartner enden und solchen, die dies nicht tun, unterschieden wird. Diese Trennung in unterschiedliche Aktivitätstypen kann anhand der Labeldateien nicht zuverlässig vorgenommen werden. Zwar sind in der orthographischen Verschriftung der Dialoge an den Positionen Fragezeichen gesetzt worden, wo vollständige Sätze enden, die eine konversationelle Frage abschließen, jedoch ist in den Dialogen häufig zu beobachten, dass nach einer grammatikalisch abgeschlossenen Fragekonstruktion noch weitere Ergänzungen gegeben werden, bevor der Sprecherwechsel stattfindet. Solche Turnenden sind als konversationelle Fragen einzuordnen, ohne dass sie direkt mit einer syntaktischen Fragekonstruktion oder einer prosodisch markierten Frage enden. Eine zuverlässige Trennung in Turns, die mit einer konversationellen Frage enden, kann also nur durch ein interpretatives Verfahren gewährleistet werden. In Abschnitt 4.4 wird anhand der interpretativen Untersuchung eines Teilkorpus auf die Trennung in Frage vs. Nicht-Frage und deren phonetische Signalisierung eingegangen.

### **4.3 Perzeptionsexperiment zur Erwartung von Turnübergaben an prosodischen Grenzen (Experiment 2)**

Im hier durchgeführten Perzeptionsexperiment werden verschieden ausgeprägte prosodische Grenzen in Bezug auf ihre interaktive Funktion im Dialog überprüft. Hierbei wird festgestellt, inwieweit die phonetische Struktur von Hörern einbezogen wird, wenn sie Erwartungen darüber bilden, ob ein Sprecherwechsel intendiert ist. Besonders der finale melodische Verlauf scheint eine wichtige Rolle bei der Signalisierung von Abschluss vs. Weiterweisung zu spielen, denn wie bereits in der Korpusanalyse gezeigt wurde, unterscheiden sich turninterne und turnfinale prosodische Grenzen in der

Verteilung der melodischen Muster deutlich (siehe 3.2.5 und 4.2.1.2). Unterschiede zeigen sich auch im verstärkten Auftreten von Laryngalisierung am Ende von Äußerungen. Diese Parameter und finale Längung werden experimentell hinsichtlich ihrer interaktionalen Funktion überprüft.

Um zu systematischen Untersuchungsergebnissen zu kommen, wird nicht wie bei Wichmann & Caspers (2001) mit Fragmenten aus aufgenommenen Dialogen gearbeitet, sondern ein lexikalisch konstanter Trägersatz systematisch und hypothesengesteuert in seinen phonetischen Eigenschaften variiert. Hierbei werden verschiedene phonetische Kontinua in mehrere Stufen unterteilt und über Resynthesen in die Stimuli eingearbeitet. So entsteht eine Vielzahl von Stimuli, die sowohl in der Kombination der Merkmale als auch in der Stärke der Einzelmerkmale variieren und so das Spektrum der im Korpus beobachteten phonetischen Variation repräsentieren. Die beiden wichtigsten Kontinua in diesem Experiment sind segmentelle Längung und der melodische Verlauf nach der letzten Akzentposition bis zum Phrasenende. Weiterhin werden phonatorische Eigenschaften am Ende der Stimuli gezielt verändert. Stimuli ohne syntaktischen Abschluss werden nicht in das Experiment einbezogen, da in dieser Hinsicht die Ergebnisse von Wichmann und Caspers eindeutig ausfallen und auch für das Deutsche keine Abweichung hiervon zu erwarten ist.

Die Versuchspersonen bekommen die phonetischen Varianten einzeln vorgespielt und bewerten auf 7-stufigen bipolaren Skalen jeweils die Wahrscheinlichkeit eines nachfolgenden Sprecherwechsels. Die Skala ist eine Messskala, die den Zusammenhang zwischen phonetischen Merkmalsbündelungen und der dadurch ausgelösten Erwartung von Hörern hinsichtlich der Äußerungsfortsetzung experimentell messbar macht. Sie ist folgendermaßen aufgebaut:

**Wird der Sprecher nach dem Satz noch weitersprechen?**

Ja							Nein
0	1	2	3	4	5	6	

Durch diese Vorgehensweise werden im Vergleich zu den Experimenten von Wichmann & Caspers (2001) drei Verbesserungen erreicht:

- Die lexikalische Struktur der verwendeten Stimuli bleibt konstant. So kann ein Einfluss dieser Informationsebene auf die Urteile der Versuchspersonen ausgeschlossen werden.
- Die phonetische Struktur der Stimuli ist vollständig kontrollierbar und weist systematische Variationen verschiedener Parameter an der finalen prosodischen Grenze auf.
- Graduelle Unterschiede in der funktionalen Einordnung der Stimuli werden durch die Verwendung von Messskalen erfassbar.

Das letztendliche Ziel des Experiments ist die funktionale Zuordnung phonetisch unterschiedlicher prosodischer Grenzen, die in der Kombination mit der Interpretation von Korpusbeispielen folgende (oder ähnliche) Kategorien herausarbeiten soll:

- gezielte Turnübergabe,
- akzeptierte Turnübergabe,
- erzwungene Turnübergabe.

Natürlich bestimmt in einem Gespräch, neben der phonetischen Gestaltung des Gesprochenen, auch dessen syntaktische und lexikalische Struktur sowie der situative Kontext die Interpretation durch den Hörer. Es ist davon auszugehen, dass in einem natürlichen Gespräch alle verfügbaren kommunikativen Ressourcen, also der gesamte situative und formale Kontext, vom Hörer in die Einschätzung des weiteren Äußerungsverlauf einbezogen werden. Diese Variablen in ein Experiment einzubeziehen ist aber kaum zu leisten und zum derzeitigen Forschungsstand nicht sinnvoll. In den Experimenten soll deshalb speziell der Beitrag der phonetischen Ausprägung auf die Interpretation durch den Hörer untersucht werden.

### **4.3.1 Hypothesen**

Folgende Hypothesen werden überprüft:

#### **Hypothese 1**

Melodische Muster an prosodischen Grenzen werden als Indikatoren für den vom Sprecher geplanten weiteren Verlauf des Dialogs interpretiert.

#### **Hypothese 2**

Tief fallende tonale Bewegungen am Phrasenende werden als Indikatoren eines nachfolgenden Sprecherwechsels interpretiert.

#### **Hypothese 3**

Eben verlaufende und leicht bis mittelhoch steigende Konturen werden als Indikatoren einer geplanten Äußerungsfortsetzung interpretiert.

#### **Hypothese 4**

Hoch steigende Muster signalisieren eine Frage an den Gesprächspartner und werden somit als Hinweis auf einen nachfolgenden Sprecherwechsel interpretiert.

#### **Hypothese 5**

Starke finale Längungen, die auf Planungsschwierigkeiten bzw. eine unflüssige Sprechweise hindeuten, senken die Erwartung eines Sprecherwechsels im Vergleich zu schwachen Längungen.

#### **Hypothese 6**

Endet der Stimulus mit einer tief fallenden melodischen Bewegung und Laryngalisierung, verstärkt sich die Erwartung eines nachfolgenden Sprecherwechsels gegenüber dem Stimulus mit tief fallender melodischer Bewegung ohne Knarrstimme.

#### **Hypothese 7**

Endet die Variante mit einem Glottalverschluss, führt dies eher zur Erwartung einer Äußerungsfortsetzung als bei vergleichbaren Varianten ohne finalen Glottalverschluss.

### 4.3.2 Stimulusmaterial

#### (1) Generierung des Ausgangsstimulus

Als Trägersatz für die Resynthesen wird folgender Satz vom Autor gesprochen, unter Studiobedingungen aufgenommen und mit 44,1 kHz digitalisiert: *Das ist ja nun nicht so angenehm*. Dieser Trägersatz würde in einem realen Dialog eine Vielzahl von Fortsetzungsmöglichkeiten bieten. Er ist syntaktisch komplett und durch die semantisch-pragmatische Struktur weder auf eine Turnfortsetzung noch auf einen Turnabschluss festgelegt. Ein anderer Satz könnte die Urteile stark in die eine oder andere Richtung beeinflussen, wenn er z.B. auf der inhaltlichen oder syntaktischen Ebene bestimmte Fortführungen suggeriert. In so einem Fall würde die Interpretation wesentlich stärker von nicht-phonetischen Faktoren bestimmt werden. Trotzdem kann davon ausgegangen werden, dass das phonetische Signalisierungssystem bei einem stärker funktional gebundenen Trägersatz in ähnlicher Weise funktioniert wie bei dem ausgewählten, relativ neutralen Stimulus, nur dass Hörer sich bei Vorhandensein eindeutiger syntaktischer oder inhaltlicher Hinweise hinsichtlich des Äußerungsfortgangs viel stärker auf die Informationen stützen würden, die auf der syntaktischen und inhaltlichen Ebene zugänglich sind. D.h., die funktionale Last bei der Interpretation der interaktionalen Intention des Sprechers würde sich in so einem Fall eher auf syntaktische und/oder wortsemantische Parameter als auf die phonetische Struktur konzentrieren. Der Trägersatz zeigt in der Ausgangsform für die Resynthesen folgende phonetische Charakteristika:

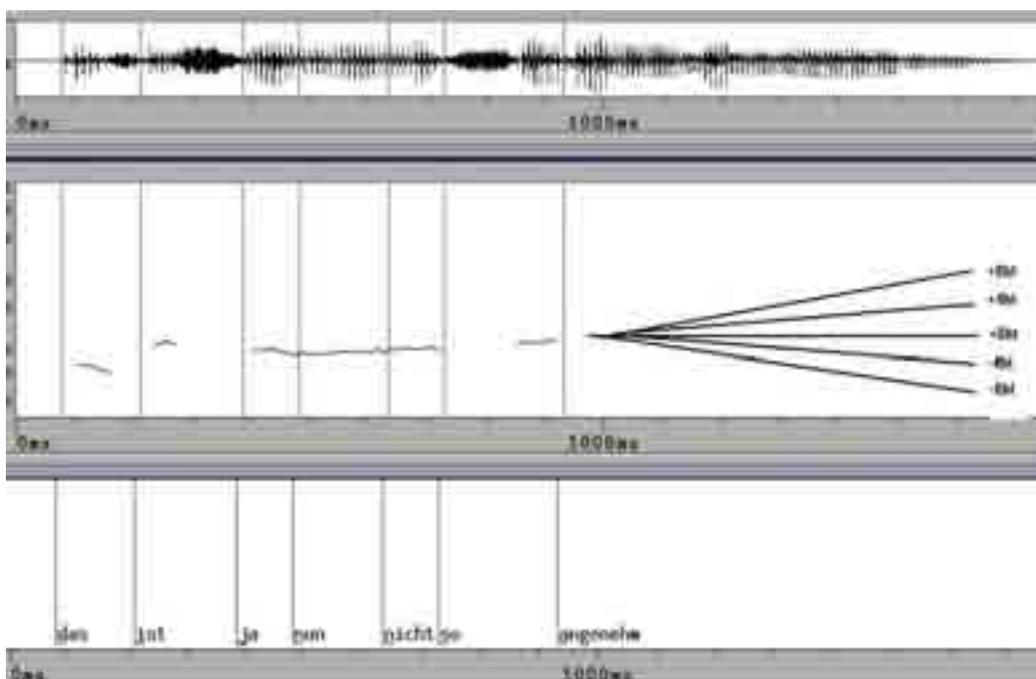
- Tonhöhengipfel auf dem satzakzentuierten *angenehm* bei 111 Hz,
- Dauer des finalen Reims (*angenehm*): 250 ms (wurde mit Faktor 0,7 in seiner Dauer verkürzt, um die in der natürlichen Produktion enthaltene finale Längung zu eliminieren),
- Endpunkt der finalen Kontur bei 70 Hz,
- Durchgehend modale Stimmqualität.

#### (2) Konturverlauf am Phrasenende

Fünf finale Konturverläufe werden in äquidistanten Schritten von vier Halbtönen resynthetisiert. Der Bezugspunkt für die variierte tonale Differenz ist bezogen auf den

tonalen Unterschied zwischen dem 111 Hz Gipfel auf *angenehm* und den F0-Endpunkten am Phrasenende. Es ergeben sich folgende Verläufe (Abbildung 11):

- tief fallend                      -8 Ht (von 111 auf 70 Hz fallend),
- leicht fallend                    -4 Ht (von 111 auf 88 Hz fallend ),
- eben                                -0 Ht (gleichbleibend bei 111 Hz),
- leicht steigend                    +4 Ht (von 111 auf 140 Hz steigend),
- hoch steigend                    +8 Ht (von 111 auf 176 Hz steigend).



**Abbildung 11:** Resynthetisierte melodische Muster

Die gewählten Konturverläufe orientieren sich an den Kategorien des *Kieler Intonationsmodells*. Dort werden insgesamt 12 phrasenfinale Melodiemuster differenziert. Eine experimentelle Überprüfung aller 12 melodischen Muster kann hier jedoch nicht geleistet werden. Deshalb wird nur eine Auswahl der in *KIM* postulierten Muster in die Experimente einbezogen. Die Auswahl beschränkt sich auf 5 Konturen, die einem linearen Verlauf folgen (ausgehend von der letzten satzakzentuierten Silbe der Phrase). Zwei der Konturen sind fallend, zwei steigend und eine verläuft eben auf dem Niveau der vorangehenden satzakzentuierten Silbe. Diese fünf gewählten Muster sind in den etikettierten Produktionsdaten sehr zahlreich vertreten und machen zusammen etwa 90%

aller melodischen Muster am Phrasenende aus. Die restlichen Konturen sind vor allem kombinierte Konturen (fallend-steigend) (Kohler 1991a; Peters 1999) und pseudo-terminale Konturen (Peters 2000).

### **(3) Längung des finalen Reims**

Ausgehend von einem Stimulus ohne finale Längung werden folgende Längungsfaktoren auf den finalen Reim von *angenehm* angewendet.

- Faktor 1,0 (Reimdauer: 250 ms),
- Faktor 1,4 (Reimdauer: 350 ms),
- Faktor 1,8 (Reimdauer: 450 ms),
- Faktor 2,2 (Reimdauer: 550 ms).

Wie im Perzeptionsexperiment 3a und 3b (siehe 5.2) gezeigt wird, werden Reime mit einer Längung ab Faktor 1,8 als klare Indikatoren einer unflüssigen Sprechweise wahrgenommen. Die Längungsstufen 1,8 und 2,2 werden in das Experiment aufgenommen, um eventuelle Wechselwirkungen zwischen der Wahrnehmung einer unflüssigen Sprechweise und der interaktionalen Funktion festzustellen. Ebenfalls kann unter 5.2 belegt werden, dass sich in natürlicher Sprache Längungserscheinungen auch über einen größeren Bereich als den finalen Reim ausdehnen können. Dies steht meist im Zusammenhang mit Planungsschwierigkeiten. Um zu prüfen, ob solche ausgedehnten segmentellen Längungen einen Einfluss auf die Vorhersage von Sprecherwechseln haben, wird eine Serie von Stimuli generiert, in der sich die Längung über das gesamte letzte Wort der Äußerung (*angenehm*) ausdehnt. Folgende Längungsfaktoren werden auf das letzte Wort der Äußerung angewendet:

- Faktor 1,0 (Wortdauer: 515 ms),
- Faktor 1,4 (Wortdauer: 720 ms),
- Faktor 1,8 (Wortdauer: 930 ms),
- Faktor 2,2 (Wortdauer: 1130 ms).

Auch diese verschieden gelängten Varianten werden mit allen fünf Melodiemustern gepaart (siehe 4.3.2).

#### (4) Phonatorische Veränderungen

Wie die Korpusuntersuchungen zeigen, spielen auch phonatorische Veränderungen an prosodischen Grenzen eine Rolle. Sowohl an syntaktischen Abbrüchen als auch an flüssigen turninternen und turnfinalen Grenzen tritt Knarrstimme auf (siehe 3.2.1.1, 4.2.1.4 und 5.1.4). Allerdings ist die durchschnittliche Dauer des geknarrten Signalabschnitts an syntaktischen Abbrüchen kürzer als bei flüssigen Grenzen (83 ms vs. 121 ms). Auch endet die Knarrstimme an syntaktischen Abbrüchen in einigen Fällen mit einem Glottalverschluss, was an flüssigen Grenzen nicht zu beobachten ist. Weiterhin zeigen die Analysen, dass die Knarrstimme an syntaktischen Abbrüchen fast immer aus einem ebenen oder leicht fallenden F<sub>0</sub>-Verlauf heraus einsetzt, während phrasenfinale Knarrstimme in flüssiger Sprache immer mit stark fallender Intonation gekoppelt ist. Um die funktionale Relevanz unterschiedlicher Arten von Knarrstimme zu testen, werden in das Stimulusmaterial auch Varianten mit und ohne finalen Glottalverschluss aufgenommen.

Das technische Vorgehen hierbei ist für beide Arten der Knarrstimme gleich. Bei der Aufnahme des Rohmaterials für die Stimulusgenerierung werden vom Sprecher des Testsatzes zusätzlich mehrere Versionen des Satzes gesprochen, die am Ende der Äußerung entweder eine abrupt einsetzende Glottalisierung mit anschließendem Glottalverschluss haben oder einen tief fallenden Intonationsverlauf, der mit Laryngalisierung endet. Die Stimuli entstehen, indem das Ende der Äußerungen mit Knarrstimme an den Ausgangsstimulus angehängt wird. Die Dauerverhältnisse werden anschließend exakt an die übrigen Stimuli angepasst.

Zur Generierung der Stimuli mit Glottalisierung wird an die Stimuli der Längungsstufe 1,4 (im finalen Reim) für alle Konturen 100 ms Glottalisierung aus einer natürlichen Produktion an den Basisstimulus angehängt, nachdem zuerst 100 ms Modalstimme weggeschnitten wurden. Das entspricht dem im Korpus gemessenen Mittelwert von 95 ms Knarrstimme bei Glottalisierung an syntaktischen Abbrüchen.

Laryngalisierung wird in das Material aufgenommen, indem vom Stimulus mit Längungsstufe 1,4 und tief fallender Kontur die Sequenz *angenehm* abgeschnitten und dann ein anderes *angenehm* angehängt wird, das eine tief fallende Kontur mit finaler Laryngalisierung enthält. Die Dauerverhältnisse werden exakt an die Stimuli der Längungsstufe 1,4 angepasst. Die Laryngalisierung ist 230 ms lang. Das ist etwas länger

als die durchschnittliche gemessene Dauer im Korpus (152 ms). Die Laryngalisierung wird nur an die tief fallende Kontur angehängt, da auch im Korpus eine solche phonatorische Veränderung nur in diesem Kontext auftritt und sie mit allen anderen Konturen unnatürlich klingt. Der laryngalisierte Abschnitt ist von leichter Verhauchung begleitet.

Der Grund, weshalb die phonatorischen Veränderungen nicht systematisch mit allen Längungsstufen gepaart werden, ist, dass dadurch die Anzahl der Stimuli zu groß werden würde und dies die Versuchspersonen überlasten könnte. Der Längungsfaktor 1,4 entspricht einer unauffälligen aber klar vorhandenen Längung des phrasenfinalen Reims. Die so erzeugten Varianten können dann in den Ergebnissen auch nur innerhalb der Serie mit entsprechender Längungsstufe verglichen werden.

#### **(5) Überblick über das gesamte Stimulusmaterial**

Durch die Manipulationen ergeben sich folgende Stimuli:

- Basisstimuli: vier Längungsfaktoren x fünf Konturen = zwanzig Stimuli,
- Ausgedehnte Längung (auf das ganze letzte Wort): drei Längungsfaktoren x fünf Konturen = fünfzehn Stimuli,
- Glottalisierung: ein Längungsfaktor x fünf Konturen = fünf Stimuli,
- Laryngalisierung: ein Längungsfaktor x eine Kontur = ein Stimulus,
- Gesamt:  $20 + 15 + 5 + 1 = 41$  Stimuli.



Audiobeispiele für die Stimuli aus Experiment 2: 4.3

#### **4.3.3 Durchführung**

Die Stimuli werden mit dreifacher Wiederholung randomisiert und zu einer Audiodatei zusammengeschnitten. Zwischen den Stimuli ist immer eine Pause von vier Sekunden. Nach jeweils 10 Stimuli ist eine längere Pause und ein Piepton. In den Pausen zwischen den Stimuli geben die Versuchspersonen ihre Urteile ab. Zu Beginn der Sitzung werden vorher aufgenommene Instruktionen abgespielt, die den Ablauf und die Aufgabenstellung erläutern (siehe Anhang). An dem Experiment nahmen 14 Personen teil.

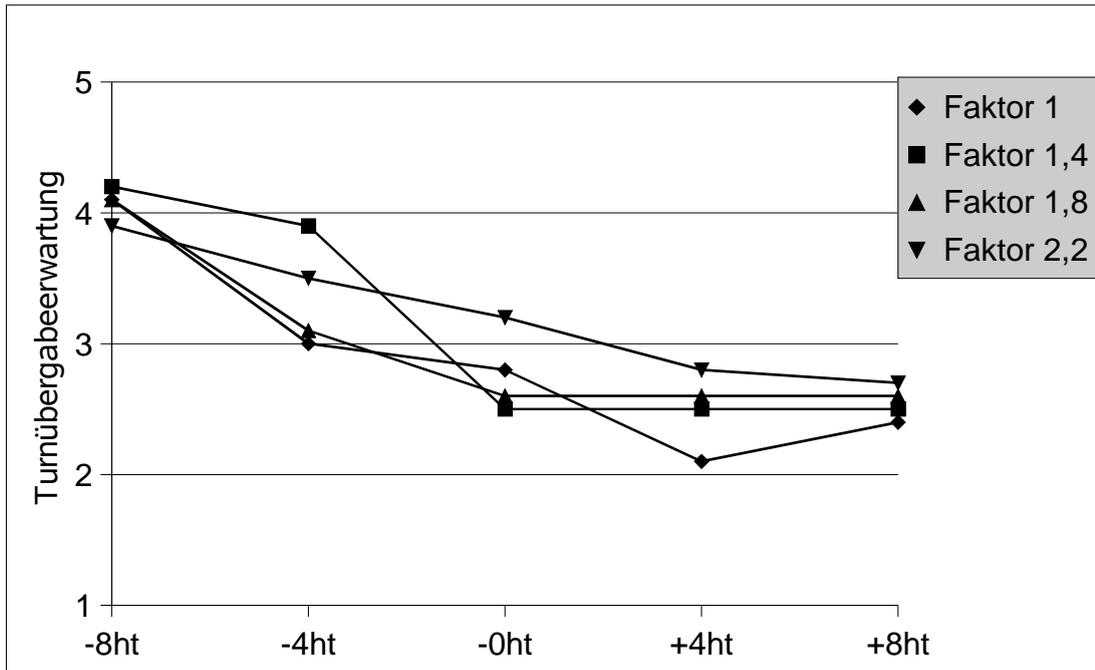
#### 4.3.4 Ergebnisse

Tabelle 24 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

**Tabelle 24:** Ergebnisse der Varianzanalyse von Experiment 2.

<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Längung des finalen Reims	1,17	0,34	0,604
Längung des finalen Wortes	1,32	0,6	0,495
Finale Kontur (über gelängtem Reim)	1,42	4,27	0,041
Finale Kontur (über gelängtem Wort)	1,66	4,81	0,024
Phonation	1	0,021	0,887
Längung (Reim) * Kontur (Reim)	6,18	1,65	0,141
Längung (Wort) * Kontur (Wort)	4,29	1,93	0,114
Phonation * Kontur (Reim)	2,95	1,4	0,258

In diesem Abschnitt werden die oben formulierten Hypothesen anhand der experimentellen Ergebnisse abgearbeitet. Die y-Achse der folgenden Abbildungen ist entsprechend der von 0 bis 6 skalierten Urteile aufgebaut. In den Abbildungen werden die Werte 0 und 6 weggelassen, da die dargestellten Mittelwerte sich nie in diesen Bereichen bewegen. Ein niedriger Wert auf der y-Achse bedeutet, dass die Versuchspersonen eher für eine Fortsetzung des Turns gestimmt haben; ein hoher Wert bedeutet, dass der Eindruck eines abgeschlossenen Turns entstanden ist. Anhand der Abbildung 12 können die ersten fünf Hypothesen überprüft werden:



**Abbildung 12:** Turnübergabeerwartung für die Stimulusserie mit unterschiedlich stark gelängten finalen Reimen und fünf finalen melodischen Mustern.

### Hypothese 1

Melodische Muster an prosodischen Grenzen werden als Indikatoren für den vom Sprecher geplanten weiteren Verlauf des Dialogs interpretiert.

Die Abbildung zeigt einen deutlichen Einfluss des melodischen Verlaufs auf die Urteile der Versuchspersonen, der prüfstatistisch signifikant ist (siehe Tabelle 24). Der finale Konturverlauf ist also relevant für die Erwartungen über den weiteren Dialogverlauf. Die Varianzanalyse zeigt im Rahmen der Berechnung von Einzelkontrasten, dass hierbei die Unterschiede zwischen der stark fallenden Kontur (-8ht) und allen anderen Konturen signifikant werden (schon der Unterschied zwischen der tief fallenden und der leicht fallenden Kontur führt zu einem F-Wert von 5,4 und einem P-Wert von 0,036 alle weiteren Vergleiche haben folglich höhere F-Werte). Ebenfalls signifikant wird der Unterschied zwischen der leicht fallenden Kontur und allen weiteren Mustern (leicht fallend vs. eben:  $F=8,9$ ,  $p=0,011$ ). Die ebenen und steigenden Muster werden untereinander nicht signifikant unterschieden. Hypothese 1 ist bestätigt.

## **Hypothese 2**

Tief fallende tonale Bewegungen (-8ht) am Phrasenende werden als Indikatoren eines nachfolgenden Sprecherwechsels interpretiert.

Auch diese Hypothese wird durch die Ergebnisse belegt. Es zeigt sich, dass die tief fallende Kontur vor allen anderen melodischen Mustern den höchsten Wert erreicht. Allerdings ist zu beachten, dass kein Mittelwert deutlich über den Wert 4 hinausgeht. D.h., dass die Urteile keine eindeutige Einschätzung, sondern nur eine Tendenz zeigen.

## **Hypothese 3**

Eben verlaufende (-0ht) und leicht bis mittelhoch steigende Konturen (+4ht) werden als Indikatoren einer geplanten Äußerungsfortsetzung interpretiert.

Die Urteile zeigen, dass die ebenen und die leicht bis mittelhoch steigenden Muster eher als Signal für eine geplante Äußerungsfortsetzung angesehen werden als das tief fallende Muster. Allerdings bewegen sich die Mittelwerte der Urteile für die ebene und die leicht bis mittelhoch steigende Kontur um den Wert 3, der genau auf der Mitte der Skala liegt. Nur wenige Mittelwerte liegen deutlich darunter. Dieses Ergebnis drückt also eher eine Unentschiedenheit der Versuchspersonen hinsichtlich der funktionalen Zuordnung aus als ein klares Votum für eine erwartete Turnfortsetzung. Auch im Vergleich mit den Ergebnissen von Wichmann & Caspers (2001), die für den hoch-eben endenden Verlauf im Englischen eine Wahrscheinlichkeit von 89% für eine unmittelbare Turnfortsetzung erhalten haben, fallen die Ergebnisse uneindeutig aus. Allerdings ist zu beachten, dass Wichmann und Caspers mit einer 'forced choice'- Befragung gearbeitet haben (siehe 4.1.3). Hypothese 3 kann auf der Basis dieser Ergebnisse weder angenommen noch verworfen werden.

## **Hypothese 4**

Hoch steigende Muster (+8ht) signalisieren eine Frage an den Gesprächspartner und werden somit als Hinweis auf einen nachfolgenden Sprecherwechsel interpretiert.

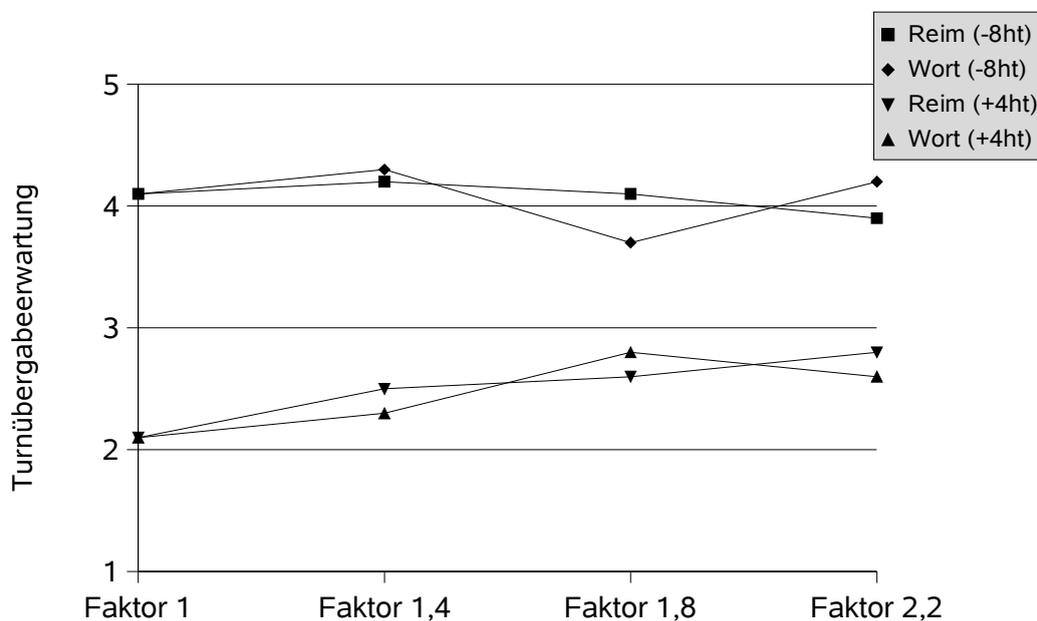
Das hoch steigende Muster erreicht ein ähnliches Antwortprofil wie das leicht bis mittelhoch steigende Muster und wird somit nicht als Indikator für eine Frage an den Gesprächspartner gedeutet, da dies mit der Erwartung eines nachfolgenden Sprecherwechsels einhergehen müsste. Hierfür gibt es zwei mögliche Erklärungen: Der Trägersatz (*Das ist ja nun nicht so angenehm.*) hat eine deklarative Syntax. Deklarative Sätze können zwar durch steigende Intonationsmuster als konversationelle Fragen realisiert werden, allerdings enthält der Satz die Partikeln *ja* und *nun* in direkter Folge. Diese Konstruktion in Verbindung mit einer Frage an den Gesprächspartner wäre sehr ungewöhnlich und hat wahrscheinlich die Wahrnehmung als Frage verhindert. Eine zweite Erklärung, die auch ergänzend zur ersten Erklärung gelten kann, ist, dass für die intonatorische Markierung eines deklarativen Satzes nicht nur der finale melodische Verlauf eine Rolle spielt, obwohl dies das auffälligste Merkmal ist, sondern auch andere Parameter (z.B. eine global angehobene Tonhöhe). Da das Stimulusmaterial nicht günstig gewählt wurde, um Hypothese 4 zu testen, wird diese weder angenommen noch verworfen.

### **Hypothese 5**

Starke finale Längungen, die auf Planungsschwierigkeiten bzw. eine unflüssige Sprechweise hindeuten, senken die Erwartung eines Sprecherwechsels im Vergleich zu schwachen Längungen.

Insgesamt zeigen die Ergebnisse in Abbildung 12 keine klaren Auswirkung der verschiedenen Längungsstufen des finalen Reims auf die Einschätzungen über den Sprecherwechsel. Auch die Prüfstatistik ergibt keinen signifikanten Effekt der Längung. Im Perzeptionsexperiment 3a kann hingegen eindeutig gezeigt werden, dass die Äußerungen mit final stark gelängten Reimen und Wörtern (Faktor 1,8 und 2,2) als sehr unflüssig wahrgenommen werden (siehe 5.2.1). Abbildung 13 zeigt ergänzend einen Vergleich zwischen dem tief fallenden und dem leicht bis mittelhoch steigenden Muster unter den verschiedenen Längungsbedingungen. Die beiden melodischen Muster wurden gewählt, weil sie hinsichtlich der interaktionalen Bedeutung die deutlichsten funktionalen Tendenzen zeigen. Die Abbildung illustriert, dass sich für die leicht bis mittelhoch steigende Kontur bei zunehmender Längung die Urteile zur Mitte der Skala verschieben, also zunehmende Unsicherheit über den weiteren Gesprächsverlauf ausgedrückt wird.

Hierbei spielt es keine Rolle, ob nur der finale Reim oder das gesamte letzte Wort gelangt ist. Für die tief fallende Kontur gibt es keinen systematischen Einfluss der Dauerrelationen. Hypothese 5 wird folglich verworfen.



**Abbildung 13:** Turnübergabeerwartung für die tief fallende Kontur (-8ht) und die leicht steigende Kontur (+4ht) unter den vier Längungsbedingungen. Getrennt für gelangte finale Reime und gelangte finale Wörter (drei Silben).

### Hypothese 6

Endet der Stimulus mit einer tief fallenden melodischen Bewegung und Laryngalisierung, verstärkt sich die Erwartung eines nachfolgenden Sprecherwechsels gegenüber dem Stimulus mit tief fallender melodischer Bewegung ohne Knarrstimme.

Die Beurteilung für den laryngalisierten Stimulus liegt im Mittel bei 4,2. Der in der Dauerstruktur gleiche Stimulus mit tief fallender Intonation (-8ht) ohne Laryngalisierung wird hinsichtlich der Turnübergabeerwartung mit einem Wert von 4,1 praktisch gleich eingeschätzt. Dieser geringe Unterschied ist statistisch nicht signifikant. Hypothese 6 wird verworfen.

### Hypothese 7

Endet die Variante mit einem Glottalverschluss, führt dies eher zur Erwartung einer Äußerungsfortsetzung als bei vergleichbaren Varianten ohne finalen Glottalverschluss.

Die Urteile zeigen bei Stimuli mit sonst gleichen Merkmalsbündeln keine signifikanten Unterschiede zwischen den Stimuli mit und ohne finalen Glottalverschluss. Hypothese 7 wird somit verworfen.

#### **4.3.5 Zusammenfassung und Diskussion**

Die Versuchspersonen bilden systematische Urteile und können offensichtlich mit der Aufgabenstellung umgehen. Allerdings zeigt deutlich höhere Standardabweichungen als in den Experimenten 1a, 1b und 1c, dass die verschiedenen Hörer die Stimuli weniger einheitlich beurteilen.

Es kann gezeigt werden, dass der melodische Verlauf am Äußerungsende einen Einfluss auf die Einschätzung des weiteren (hypothetischen) Dialogverlaufs hat, während die segmentellen Dauerrelationen hier fast keine Rolle spielen. Vor allem die tief fallende Kontur wird mit Turnabschlüssen assoziiert und die steigenden Konturen mit Turnfortsetzungen. Allerdings bewegen sich die Mittelwerte der Urteile kaum aus dem Bereich zwischen 2 und 4 heraus. Das bedeutet, dass im Mittel keine eindeutigen funktionalen Zuordnungen getroffen wurden, sondern sich lediglich Tendenzen zeigen. Dies hat mehrere Gründe:

- Das individuelle Urteilsverhalten der Versuchspersonen variiert stark. Während die durchschnittliche Standardabweichung in den Experimenten zur Kohäsion (Experimente 1a, 1b, 1c) bei etwa 1 liegt und für alle Stimuli etwa gleich ist, liegt sie im Experiment zur Turnübergabewahrscheinlichkeit durchschnittlich deutlich höher und variiert für die einzelnen Stimuli zwischen 0,8 und 2,3. Die Urteile für die einzelnen Wiederholungen der Stimuli innerhalb einzelner Versuchspersonen hingegen sind relativ konsistent.
- Die wiederholte und kontextfreie Präsentation der Stimuli ist eine unnatürliche Situation und das Urteilsverhalten kann nicht als direktes Abbild der in einer Dialogsituation ablaufenden Hypothesenbildung angesehen werden. Für die Versuchspersonen ist es möglicherweise schwierig, die Einschätzungen über einen Dialogverlauf, die im normalen Gespräch unbewusst ablaufen, bewusst und ohne kontextuelle Hinweise zu treffen.

- Dem phonetischen Signalisierungssystem liegt kein striktes 1:1-Verhältnis zwischen Form und Funktion zugrunde. Die Ergebnisse deuten eher auf eine unterstützende Funktion der phonetischen Komponente hin, die erst in Kombination mit syntaktischen und semantischen Strukturen zu eindeutigen interaktionalen Funktionalisierungen führt.
- Die phonetische Struktur von Äußerungen bietet noch eine Reihe von Variationsmöglichkeiten, die für das Dialogsteuerungssystem eine Rolle spielen können, aber nicht in das Stimulusmaterial integriert wurden. In natürlichen Dialogen ist z.B. in turnfinaler Position oft eine deutlichere Intensitätsabnahme zu beobachten als an turninternen prosodischen Grenzen. Auch unterschiedliche Synchronisierungen von nuklearen Gipfel- und Talpositionen könnten in weitere Experimente einbezogen werden, um zu deutlicheren Urteilen zu gelangen.

Die Urteile zeigen, dass das phonetische Signalisierungssystem zwar funktionale Relevanz besitzt, jedoch kaum eindeutige Zuordnungen hinsichtlich des Gesprächsfortgangs zulässt. Ein klares Ergebnis ist, dass die Veränderungen in der Phonation keine Auswirkungen auf die Urteile haben. Dies ist überraschend, denn die beiden Typen von Knarrstimme zeigen in den Produktionsdaten Verteilungen, die auf eine funktionale Bindung hinweisen. Knarrstimme mit Glottalverschluss ist nur an syntaktischen Abbruchstellen zu beobachten, während weich auslaufende Laryngalisierung gehäuft an prosodischen Grenzen vor Sprecherwechseln vorkommt. Hier scheinen typische Merkmale der Produktion in der Perzeption nicht mit funktionalen Kategorien assoziiert zu werden. Unter diesem Gesichtspunkt muss die von Local & Kelly (1986) postulierte funktionale Differenzierung zwischen *turn-holding-device* (Glottalverschluss) und *turn-yielding-device* (Laryngalisierung) noch einmal kritisch geprüft werden. Aufgrund der experimentellen Ergebnisse liegt nahe, dass die perzeptive Relevanz dieser Parameter von den Autoren überschätzt wird. So ist auch deren generelle Ansicht, dass jedes phonetische Detail für den Kommunikationsprozess von Bedeutung ist, kritisch zu überdenken. Natürlich beziehen sich die Aussagen von Local und Kelly auf das Englische, da aber die von ihnen beschriebenen Phänomene völlig vergleichbar auch in den untersuchten deutschen Daten vorliegen, ist anzunehmen, dass auch die perzeptiven Ergebnisse auf das Englische übertragbar sind.

Ein Einwand, der seitens der Conversation Analysis (CA) gegen diese Kritik erhoben werden könnte, ist, dass es lediglich im durchgeführten Perzeptionsexperiment nicht möglich ist, die perzeptorische Relevanz phonatorischer Veränderungen zu belegen, obwohl diese in natürlichen Gesprächen durchaus vorhanden ist und nur durch die Methoden der sequentiellen Analyse belegt werden kann. Eine solche Argumentation würde allerdings die experimentelle Phonetik insgesamt in Frage stellen und m.E. eine einseitige und dem Forschungsgegenstand nicht angemessene Fixierung auf interpretative Verfahren demonstrieren. Allerdings ist sicher richtig, dass ein Perzeptionsexperiment immer eine Abstraktion einer tatsächlichen Gesprächssituation ist und Versuchspersonen in diesem situativen Kontext u.U. anders auf sprachliches Material reagieren als in einem natürlichen Gespräch. Besonders da in dem Experiment ein bewusstes Urteil über Erwartungen gefordert wurde, die sonst unbewusst gebildet werden, ist nicht sicher, dass das Experiment die kommunikativen Funktionen phonatorischer Veränderungen adäquat erfassen kann. Dass für die melodischen Muster die erwarteten funktionalen Zuordnungen von den Probanden vorgenommen wurden, ist allerdings ein Indiz dafür, dass das Experiment geeignet ist, funktionale Zusammenhänge herzustellen.

#### **4.4 Interpretation von Korpusbeispielen**

Wie bereits bei den turninternen prosodischen Grenzen soll die Subklassifikation turnfinaler prosodischer Grenzen durch die Interpretation von Korpusbeispielen gestützt werden. Hierfür werden jeweils 40 Beispiele der drei unterschiedlich synchronisierten Turnübergänge aus dem LINDENSTRASSE-Korpus herausgegriffen (Übergänge mit Pause, Übergänge mit unmittelbarem Anschluss und Übergänge mit Überlagerung). An jeder der 120 finalen prosodischen Grenzen wird entschieden, ob am Turnübergang das Ende eines möglichen Satzes erreicht wird und ob ein möglicher inhaltlicher Abschluss vorliegt. Weiterhin wird entschieden, ob das Turnende mit einer Frage an den Gesprächspartner einhergeht.

Durch dieses Verfahren soll eine funktionale Subklassifikation in verschiedene Typen von Turnübergängen gestützt werden, denn es hat sich bereits durch die automatische Merkmalsextraktion gezeigt, dass eine Subklassifikation auf rein akustisch/phonetischer Basis wegen der starken Überscheidung der Merkmalsbündel in den drei unterschiedlich

synchronisierten Turnübergängen nicht sinnvoll ist. Auch das durchgeführte Perzeptionsexperiment kann keine eindeutigen funktionalen Bindungen herausarbeiten. Das gewählte interpretative Verfahren bietet die Möglichkeit festzustellen, ob die drei Arten der Übergänge in ihrer semantisch-pragmatischen Struktur unterschiedlich sind. Diese Herangehensweise unterscheidet sich durch den statistischen Ansatz von den Einzelfallbeschreibungen der *Conversation Analysis*. Folgende Hypothesen werden geprüft:

### **Hypothese 1**

Turnübergänge mit Überlagerung zeigen häufiger inhaltliche und/oder syntaktische Unabgeschlossenheit als unmittelbare Übergänge oder Übergänge mit Pausen.

### **Hypothese 2**

Fragen an den Gesprächspartner sind bei überlagernden Turnübergängen seltener zu erwarten als in den anderen beiden Klassen.

Wenn die Interpretation diese Erwartungen bestätigt, kann das Ergebnis herangezogen werden, um Turnübergänge ohne Überlagerung als gezielt vom Sprecher eingeleitet zu klassifizieren und Übergänge mit Überlagerung als vom Hörer eingeleitete Sprecherwechsel.

#### 4.4.1 Ergebnisse

Die Interpretation der 120 Sprecherwechsel liefert folgende Ergebnisse:

**Tabelle 25a:** Inhaltliche Abgeschlossenheit (LINDENSTRASSE).

	<b>Pause (n=40)</b>	<b>Unmittelbarer An- schluss (n=40)</b>	<b>Überlagerung (n=40)</b>
<b>Inhaltlich abgeschlossen</b>	36 (90%)	40 (100%)	39 (97,5%)

**Tabelle 25b:** Syntaktische Abgeschlossenheit (LINDENSTRASSE).

	<b>Pause (n=40)</b>	<b>Unmittelbarer Anschluss (n=40)</b>	<b>Überlagerung (n=40)</b>
<b>Syntaktisch abgeschlossen</b>	39 (97,5%)	38 (95%)	37 (92,5%)

**Tabelle 25c:** Fragefunktion (LINDENSTRASSE).

	<b>Pause (n=40)</b>	<b>Unmittelbarer Anschluss (n=40)</b>	<b>Überlagerung (n=40)</b>
<b>Frage an den Gesprächspartner</b>	11 (27,5%)	8 (20%)	9 (22,5%)

Die Zahlen der Tabellen 25a, b und c zeigen keine verwertbaren Unterschiede zwischen den drei Typen von Turnübergängen. An fast allen Sprecherwechseln wird vom übergebenden Sprecher sowohl die syntaktische als auch die inhaltliche Struktur komplettiert. Das durchgeführte Interpretationsverfahren liefert also keine Hinweise auf funktionale Unterschiede zwischen den drei unterschiedlich synchronisierten Übergängen. Beide gestellten Hypothesen werden verworfen.

Lediglich die finalen melodischen Muster zwischen den Übergängen mit und ohne Frage an den Partner zeigen deutliche Unterschiede in der Verteilung, wie Tabelle 26 zeigt. Unter den Sprecherwechseln mit Frage gibt es wesentlich mehr steigende Muster, und es ist auffällig, dass die melodische Kontur, die am häufigsten in Verbindung mit Fragen steht, die leicht bis mittelhoch steigende Kontur ist. Schwerpunkte bei den Wechseln ohne Frage liegen bei den tief fallenden und den leicht fallenden Verläufen. Es ist aber wichtig zu bemerken, dass auch an Turnabschlüssen, die mit einer Frage an den Dialogpartner verbunden sind, fast alle melodischen Muster vorkommen, obwohl nur 28 Fragen in die Untersuchung eingehen. Auch die Fragefunktion scheint also nicht an bestimmte melodische Muster gebunden zu sein.

**Tabelle 26:** Melodische Verläufe am Turnende (LINDENSTRASSE).

<b>Melodischer Verlauf</b>	<b>Keine Frage (n=92)</b>	<b>Frage (n=28)</b>
Tief fallend	35 (38,0%)	2 (7,1%)
Tief fallend + sehr leicht steigend	6 (6,5%)	0 (0%)
Fallend-steigend	5 (5,4%)	4 (14,3%)
Hoch steigend	0 (0%)	6 (21,4%)
Eben	13 (14,1%)	1 (3,6%)
Leicht fallend	23 (25,0%)	2 (7,1%)
Leicht bis mittelhoch steigend	7 (7,6%)	11 (39,3%)
Eben + sehr leicht steigend	2 (2,2%)	1 (3,6%)
Leicht fallend + sehr leicht steigend	1 (1,1%)	1 (3,6%)

#### **4.4.2 Diskussion**

Es gibt mehrere Erklärungen, warum die Verteilungen in den drei getrennt analysierten Klassen so ähnlich sind und es kaum zu abgebrochenen inhaltlichen und syntaktischen Strukturen kommt. Als erstes ist die generell sehr einvernehmliche und kooperative Einstellung der Dialogpartner zu nennen. Die Gesprächspartner kennen sich gut und achten darauf, dem Gegenüber nicht zu forsch ins Wort zu fallen. Sie sind an der Lösung einer gemeinsamen Aufgabe interessiert, und die Gespräche geben keinen Anlass zu Uneinigkeiten, die ein forsches Unterbrechen des Partners hervorrufen könnten. Möglicherweise wird das kooperative Verhalten der Versuchspersonen zusätzlich durch die Aufnahmesituation verstärkt, da häufiges Unterbrechen des Partners als unhöflich angesehen werden kann und dies in dieser besonderen Gesprächssituation speziell vermieden wird.

Beim Abhören der Dialoge entsteht dennoch an vielen Turnübergängen der Eindruck, dass die Übergabe vom derzeitigen Sprecher nicht geplant war und der Hörer trotzdem an einer geeigneten Stelle zu reden beginnt. Die so unterbrochenen Sprecher machen aber kaum Anstrengungen den Turn zu halten, z.B. indem sie trotz der Unterbrechung weitersprechen. Sie bringen bei überlagernden Einsätzen des Hörers meist nur noch die begonnene syntaktische Konstruktion zu Ende und begeben sich anschließend in die Hörerrolle. Dieses Zuendebringen ist bei Unterbrechungen möglicherweise die übliche Verhaltensweise, im Gegensatz zu einem sofortigem Abbruch innerhalb der begonnenen syntaktischen Struktur, zumal die unterbrechenden Hörer bei

überlagerndem Redebeginn meist einen Punkt in der Äußerung wählen, an dem die inhaltliche und syntaktische Struktur kurz vor einem möglichen Abschluss steht. Auch die gemessene durchschnittliche Dauer der Überlagerungen liegt mit 480 ms niedrig und stützt die obige Argumentation.



Audiobeispiele: Komplettierung der begonnenen syntaktischen Struktur bei unterbrechender Turnübernahme: 2.11, 2.13, 3.3

Auf der anderen Seite müssen Sprecherwechsel mit Pausen nicht unbedingt gezielt eingeleitete Übergaben sein. Die Tabelle 25a zeigt, dass es hier in vier Fällen zu einem Sprecherwechsel kommt, ohne dass ein inhaltlicher Abschluss erreicht wird. Die Erklärung hierfür ist wahrscheinlich, dass die Sprecher in diesen Fällen eine Pause machen, um den weiteren Äußerungsverlauf zu planen, und dass der Hörer in dieser Pause das Wort ergreift und den Turn übernimmt, obwohl der Sprecher dies nicht geplant hat.

#### **4.5 Vergleich zwischen turnfinalen und turninternen prosodischen Grenzen**

In diesem Abschnitt wird eine gegenüberstellende Zusammenfassung der phonetischen Ausprägungen turnfinaler und turninterner prosodischer Grenzen gegeben. Die Tabellen zeigen die Verteilungen melodischer Muster, finaler Längung und phonatorischer Veränderungen im Vergleich zwischen turnfinalen und turninternen prosodischen Grenzen. Diese drei Parameter sind durchgehend vergleichbar (im Gegensatz zu F0-Reset) und der Vergleich soll zeigen, ob sich die funktionalen Unterschiede in prosodischen Strukturen widerspiegeln. Tabelle 27 fasst hierbei die Tabellen 3 und 19 zusammen, Tabelle 28 die Tabellen 9 und 22, und Tabelle 29 stellt die Ergebnisse der Tabellen 6 und 21 gegenüber. Die ausgewerteten Datenbasen sind aufgrund der Zusammenfassung z.T. innerhalb der Spalten der Tabellen nicht identisch (dies ist dann in den Tabellen vermerkt).

**Tabelle 27:** Vergleich melodischer Verläufe an turnfinalen und turninternen prosodischen Grenzen (LINDENSTRASSE und Terminabsprachen).

	<b>Turnfinal mit Pause (n=183)</b> (LINDENSTRASSE)	<b>Turnfinal mit unmittelbarem Anschluss (n=66)</b> (LINDENSTRASSE)	<b>Turnfinal mit Überlagerung (n=128)</b> (LINDENSTRASSE)	<b>Turnintern (n=3399)</b> (LINDENSTRASSE und Termin- absprachen)
<b>Starke melodische Bewegung</b>				
Tief fallend	53 (29,0%)	23 (34,9%)	35 (27,3%)	915 (26,9%)
Tief fallend + sehr leicht steigend	12 (6,6%)	4 (6,1%)	5 (3,9%)	202 (5,9%)
Fallend-steigend	12 (6,6%)	6 (9,1%)	5 (3,9%)	121 (3,6%)
Hoch steigend	13 (7,1%)	4 (6,1%)	3 (2,3%)	65 (1,9%)
<b>Gesamt starke melodische Bewegung</b>	<b>90 (49,2%)</b>	<b>37 (56,1%)</b>	<b>48 (37,5%)</b>	<b>1303 (38,3%)</b>
<b>Schwache melodische Bewegung</b>				
Eben	25 (13,7%)	7 (10,6%)	19 (14,8%)	562 (16,5%)
Leicht fallend	20 (10,9%)	9 (13,6%)	34 (26,6%)	962 (28,3%)
Leicht bis mittelhoch steigend	37 (20,2%)	9 (13,6%)	11 (8,6%)	438 (12,9%)
Eben + sehr leicht steigend	3 (1,6%)	2 (3,0%)	5 (3,9%)	23 (0,7%)
Leicht fallend + sehr leicht steigend	8 (4,4%)	2 (3,0%)	11 (8,6%)	111 (3,3%)
<b>Gesamt schwache mel. Bewegung</b>	<b>93 (50,8%)</b>	<b>29 (43,9%)</b>	<b>80 (62,5%)</b>	<b>2096 (61,7%)</b>
<b>Gesamt</b>	<b>128 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>3399 (100%)</b>

**Tabelle 28:** Vergleich phonatorischer Veränderungen an turnfinalen und turninternen prosodischen Grenzen mit tief fallender Intonation (LINDENSTRASSE).

	<b>Turnfinal mit Pause (n=53)</b>	<b>Turnfinal mit unmittelbarem Anschluss(n=23)</b>	<b>Turnfinal mit Überlagerung (n=35)</b>	<b>Turnintern (n=238)</b>
<b>Modalstimme</b>	30 (56,6%)	14 (60,9%)	27 (77,1%)	159 (66,8%)
<b>Knarrstimme (unregelmäßig)</b>	15 (28,3%)	7 (30,4%)	4 (11,4%)	26 (10,9%)
<b>Knarrstimme (regelmäßig)</b>	2 (3,8%)	0 (0%)	2 (5,7%)	20 (8,4%)
<b>Interferenz oder Unsicherheit</b>	6 (11,3%)	2 (8,7%)	2 (5,7%)	33 (13,8%)
<b>Glottalverschluss</b>	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<b>Gesamt</b>	<b>53 (100%)</b>	<b>23 (100%)</b>	<b>35 (100%)</b>	<b>238 (100%)</b>

**Tabelle 29:** Vergleich segmenteller Längung des turnfinalen Reims an turnfinalen und turninternen prosodischen Grenzen (LINDENSTRASSE und Terminabsprachen).

	<b>Turnfinal mit Pause (n=183) (LINDENSTRASSE)</b>	<b>Turnfinal mit unmittelbarem Anschluss (n=66) (LINDENSTRASSE)</b>	<b>Turnfinal mit Überlagerung (n=128) (LINDENSTRASSE)</b>	<b>Turnintern (n=3399) (LINDENSTRASSE und Terminabsprachen)</b>
<b>Kürzung</b>	1 (0,5%)	0 (0%)	2 (1,5%)	75 (2,2%)
<b>Weder Längung noch Kürzung</b>	47 (25,7%)	18 (27,3%)	40 (31,3%)	1087 (32,0%)
<b>Längung</b>	135 (73,8%)	48 (72,7%)	86 (67,2%)	2237 (65,8%)
<b>Gesamt</b>	<b>183 (100%)</b>	<b>66 (100%)</b>	<b>128 (100%)</b>	<b>3399 (100%)</b>

Alle dargestellten Verteilungen sind sich generell sehr ähnlich. Unterschiede zeigen sich lediglich in einer erhöhten Häufigkeit starker melodischer Bewegungen zwischen turnfinalen Grenzen ohne Überlagerung und turninternen Grenzen sowie in einer erhöhten Häufigkeit von Knarrstimme in der turnfinalen Position. Dies bedeutet, dass über die phonetische Struktur prosodischer Grenzen keine klare Zuordnung zu Turnabschlüssen vs. Turnfortsetzungen erreicht werden kann.

## 4.6 Zusammenfassung und abschließende Diskussion zu den turnfinalen prosodischen Grenzen

Die Ergebnisse aus den drei Analysebereichen (automatische Merkmalsextraktion, Perzeptionsexperimente und interpretative Analyse) ergeben ein komplexes Bild und tragen einiges zur Klärung der untersuchten Fragestellungen bei. Sie werfen aber auch neue Fragen und Probleme auf. Es sei vorangestellt, dass die Ergebnisse eine Aufstellung phonetisch definierter Kategorien turnfinaler prosodischer Grenzen mit eindeutiger funktionaler Zuordnung nicht zulassen, da die Variationsbreite prosodischer Strukturen im Korpus zu groß ist und auch in den Experimenten die Urteile der Versuchspersonen keine klaren Kategorien zeigen.

Als erster Diskussionspunkt stellt sich die Frage, ob es ein striktes, phonetisch kodiertes System gibt, das zwischen turninternen und turnfinalen Grenzen unterscheidet und so die Funktionen *Weiterweisung* vs. *Abschluss* trägt. Die Untersuchungen zeigen, dass die segmentelle Dauerstruktur zu dieser Funktionsopposition kaum beiträgt, sondern vor allem starke melodische Bewegungen Turnabschlüsse markieren. Hierbei macht die tief fallende Kontur den größten Teil der starken melodischen Bewegungen aus. In den Terminabsprachedaten gibt es am Ende von etwa 80% der Turns starke melodische Konturen, während es an prosodischen Grenzen im Turninneren nur knapp 40% sind. Auch das Perzeptionsexperiment zeigt eine Bindung der tief fallenden Kontur mit dem Abschluss von Redebeiträgen. Die Untersuchungen belegen aber auf der anderen Seite auch, dass es im Inneren von Turns zu einer Vielzahl von prosodischen Grenzen kommt, die phonetische Merkmale tragen, die typisch für die turnfinale Position sind. Besonders häufig sind diese Strukturen zu finden, wenn im Inneren von Redebeiträgen semantisch-pragmatische Blöcke abgeschlossen werden. Auch Gilles (2005:21) beobachtet in seinen Daten solche Abschlüsse ohne nachfolgenden Sprecherwechsel:

*Turn-interne Abschlüsse markieren z.B. das Ende von Erzählungen oder Sachverhaltsdarstellungen oder kleinere erzähltechnische, thematische oder argumentative Einschnitte (Ende eines Erzählschritts, eines Themas, eines Subthemas oder eines Einschubs).*

Um also besser mit den Konzepten *Abschluss* vs. *Weiterweisung* arbeiten zu können, ist m.E. eine Untergliederung der beiden Funktionen in *inhaltsbezogen* vs. *dialogbezogen* sinnvoll. So können Abschlüsse im Inneren von Turns auf einer inhaltlichen Ebene als abschließend klassifiziert werden, ohne dass sie auch zu einem interaktionalen Abschluss mit Sprecherwechsel führen. Die Unterteilung in *inhaltsbezogene* und *dialogbezogene* Abschlüsse und die damit verbundene Möglichkeit inhaltlicher Abschlüsse ohne dialogbezogene Abschlussfunktion ist zudem sinnvoll, wenn man von Gesprächssituationen ausgeht, in denen Sprecherwechsel nicht üblich sind. Wenn z.B. ein Vortrag gehalten wird, ist innerhalb der fortlaufenden Rede häufig mit inhaltlichen Abschlüssen zu rechnen, die lediglich die thematische Gliederung (z.B. in Absätze) unterstützen, aber dialogbezogen keineswegs dazu anregen sollen, dass ein Zuhörer das Wort übernimmt.

Auch für das Konzept der Weiterweisung ist zu erwarten, dass die dialogbezogene und die inhaltsbezogene Ebene nicht immer parallel laufen; denn eine inhaltliche Weiterweisung kann zwar mit dem Wunsch nach einer Turnfortsetzung einhergehen, aber auch hier ist ein Auseinandergehen der beiden funktionalen Aspekte vorstellbar, z.B. wenn ein Sprecher ein Thema eröffnet und dem Hörer dann den Turn übergibt, damit dieser Ergänzungen vornehmen kann. Eventuell ist es sinnvoll, dieses Konzept von inhaltlicher Weiterweisung und interaktionalem Abschluss auch auf Fragen zu beziehen, denn eine Frage wirft oft ein Thema auf, ist also inhaltlich weiterweisend, aber in interaktionaler Hinsicht abschließend, da der Turn mit der Frage abgeschlossen wird.

Eine rein phonetisch ausgerichtete Differenzierung zwischen inhaltsbezogenen und dialogbezogenen Abschlüssen und Weiterweisungen ist m.E. allerdings nicht möglich; denn die phonetischen Merkmale der funktionalen Klassen zeigen starke Überschneidungen, und für Hörer ist vermutlich vor allem aus dem inhaltlichen und situativen Gesamtkontext ableitbar, zu welchem Zeitpunkt z.B. ein dialogbezogener Abschluss erreicht wird und eine Übernahme des Rederechts möglich ist oder erwartet wird. Hier liegt auch eines der großen Probleme für experimentelle Untersuchungen der Dialogsteuerung. Es kann nie ganz klar bestimmt werden, inwieweit die Versuchspersonen Urteile über die inhaltliche oder die dialogbezogene Ebene abgeben. Ebenso ist der Einfluss des lexikalischen und situativen Kontextes auf die Urteile schwer abschätzbar.

Die eben diskutierten Funktionen sind Kategorien der Sprecherseite. D.h., es wird angenommen, dass der Sprecher einen (groben) Plan für die Struktur seines Turns hat, zu

einem bestimmten Zeitpunkt einen dialogbezogenen Abschluss erreicht, dies inhaltlich und phonetisch zum Ausdruck bringt und der Hörer ohne Verzögerung in die Sprecherrolle wechselt. Diese Sichtweise ist allerdings idealisiert. In gleichberechtigten Alltagsdialogen sind oft Sprecherwechsel an Positionen zu beobachten, die keinerlei Merkmale eines Abschlusses erkennen lassen. Hier entsteht meistens der Eindruck, dass der Hörer eine unterbrechende Turnübnahme durchführt. Aus diesem Grund muss in die Klassifikation von Sprecherwechseln auch die Aktivität des Hörers und somit eine Kombination der Perspektiven von Sprecher und Hörer einfließen. Um dieses Ziel zu erreichen, sollte eine Unterteilung in *gezielte Turnübergaben*, *erlaubte Turnübernahmen* und *umstrittene Turnübernahmen* auf empirischer Basis belegt werden. Als Ausgangspunkt hierfür stand die Annahme, dass Turnübergänge mit überlagerndem Sprechen eher in die Kategorie der *erlaubten* oder der *umstrittenen* Übergänge fallen, während Übergänge mit unmittelbarem Anschluss oder einer Pause zwischen den Redebeiträgen durch den Sprecher gezielt eingeleitet werden. Diese Annahme müsste sich dadurch bestätigen lassen, dass an den überlagernden Übergängen zumeist Merkmale der interaktionalen Weiterweisung zu finden sind (z.B. inhaltliche und syntaktische Unabgeschlossenheit, schwache melodische Bewegungen), während die nicht-überlagernden Übergänge Merkmale von Abschlüssen tragen (z.B. inhaltliche und syntaktische Abgeschlossenheit, starke melodische Bewegungen).

Die Untersuchungsergebnisse bestätigen eine strikte Trennung dieser Kategorien jedoch nicht. An nahezu allen untersuchten Turnübergängen finden sich sowohl (mögliche) syntaktische Abschlüsse als auch Endpunkte von inhaltlich als komplett interpretierbaren Einheiten. Auch die melodischen Muster, die auf der phonetischen Ebene entscheidend für Abschluss vs. Weiterweisung sind, stützen die angestrebte Einteilung nur zum Teil. Zwar liegt der Anteil der starken melodischen Bewegungen an den überlagernden Übergängen des LINDENSTRASSE-Korpus mit 37,5% am niedrigsten, jedoch an den Übergängen mit unmittelbarem Anschluss auch nur bei 56,1% und an denen mit Pause bei 49,2%. Hieraus lässt sich ableiten, dass unter den überlagernden Übergängen weniger gezielt durch den Sprecher eingeleitete Übergaben sind. Auf der anderen Seite gibt es aber auch mehr als ein Drittel überlagernder Übergänge mit starken melodischen Mustern. Ebenso muss bei den nicht-überlagernden Übergängen der hohe Anteil von finalen prosodischen Grenzen ohne starke melodische Bewegung berück-

sichtigt werden, der eine funktionale Zuordnung als *gezielte Turnübergaben* nicht stützt, da hier der melodische Abschluss fehlt.

Es ist also davon auszugehen, dass sich in jeder Klasse der verschiedenen synchronisierten Übergänge sowohl gezielte als auch erlaubte Übergänge befinden. Das systematische Abhören der Dialoge lässt weiterhin die Aussage zu, dass wirklich umstrittene Übergänge in den Daten kaum vorkommen, da die Dialoge sehr kooperativ verlaufen, und die unterbrochenen Sprecher sich fast immer ohne den Versuch, das Rederecht zu halten, in die Hörerrolle begeben. Die unterbrechenden Hörer ihrerseits wählen die Position ihrer Unterbrechung so, dass der derzeitige Sprecher dort einen (möglichen) inhaltlichen und syntaktischen Abschluss erreicht hat. Dies scheint sogar bei den überlagernden Übernahmen der Fall zu sein. Die Übernahmeposition ist hier in den meisten Fällen so gewählt, dass der unterbrochene Sprecher die begonnene inhaltliche und syntaktische Struktur während der Überlagerung noch komplettieren kann. Dies gilt aber keineswegs für jede Art von Gespräch, sondern ist möglicherweise als spezifische Eigenschaft der LINDENSTRASSE-Dialoge anzusehen. Hier herrscht anscheinend Einvernehmen darüber, dass Unterbrechungen ein normaler Bestandteil der Dialogstrukturierung sind und der Unterbrechende üblicherweise sofort das Rederecht bekommt.



Audiobeispiele für gezielte, erlaubte und umstrittene Turnübergänge: 2.1 – 2.13.

Aus diesen Beobachtungen und Interpretationen der Dialogdaten und aus den experimentellen Ergebnissen ergeben sich einige neue Aspekte hinsichtlich interaktionaler Muster im Dialog und der Rolle der Phonetik in der Dialogsteuerung. Es zeigt sich (vor allem durch das Perzeptionsexperiment), dass die funktionale Bindung intonatorischer Konturen und begleitender phonatorischer Veränderungen hinsichtlich der Turnübergabesignalisierung schwächer ausgeprägt ist als erwartet. Die Ergebnisse deuten eher auf eine unterstützende Funktion der phonetischen Komponente hin, die erst in Kombination mit syntaktischen und semantischen Strukturen zu interaktionalen Funktionalisierungen führt. Dies bedeutet, dass für eine aussagekräftige Klassifikation von einem erweiterten Merkmalsbündel ausgegangen werden muss, das auch die semantische und syntaktische Struktur einbezieht. Hierdurch entsteht ein wesentlich weiterer Rahmen, in dem die Gesprächspartner ihr interaktionales Verhalten steuern und Einschätzungen über adäquate Reaktionen im Gespräch vornehmen.

Aber auch dieses erweiterte Bündel führt m.E. nicht zu einem starren System, das Sprecherwechsel bei bestimmten Bündelungen "vorschreibt" bzw. an anderer Stelle "verbietet". Sprecher und Hörer zeigen im Dialog oft ein großes Maß an Flexibilität und gegenseitiger Anpassungsfähigkeit. So kann das grundlegende Prinzip des Turntakings als *erlaubend* und nicht als *vorschreibend* charakterisiert werden. Dieser erlaubende Charakter lässt sich durch die folgenden zwei Situationen illustrieren, die intuitiv gut nachvollziehbar sind.

- (1) Ein Sprecher schließt eine Äußerung durch melodische, syntaktische und semantische Vervollständigung ab und macht eine kurze Pause. Der Hörer kann an dieser Stelle nun das Wort ergreifen. Oft wird er es aber nicht tun, sondern z.B. durch ein Rezeptionssignal andeuten, dass er weiter in der Hörerrolle bleiben möchte. Üblicherweise wird der Sprecher dann seine Rede fortsetzen (es sei denn er hat eine direkte Frage an den Partner gestellt).



Audiobeispiel: Finalitätsmerkmale und Turnfortsetzung: 1.4

- (2) Ein Sprecher lässt in seiner Rede kein komplexes Merkmalsbündel entstehen, das einen interaktionalen Abschluss erkennen lässt. Dennoch hat der Hörer das Bedürfnis "einzuhaken", weil er z.B. etwas Wichtiges ergänzen möchte. Der Hörer wird die Ergänzung in vielen Fällen als unterbrechende Turnübernahme beginnen, und der vormalige Sprecher wird dies üblicherweise akzeptieren.



Audiobeispiele: Turnübernahme ohne Finalitätsmerkmale: 2.10, 2.12

Anders ist die Situation, wenn eine direkte Frage an den Gesprächspartner gestellt wird. In diesem Fall liegt ein eindeutiger gezielter Turnübergang vor. Der Komplex der Frage wird aber durch die vorgenommenen Untersuchungen nicht explizit analysiert. Hier sind weiterführende Untersuchungen notwendig.



Audiobeispiele: Turnübergaben mit Fragen: 2.3, 2.4, 2.5

Die obigen Erklärungsversuche entfernen sich von der rein phonetischen Fragestellung und machen deutlich, dass der Versuch einer wirklichen Erklärung interaktionaler Strukturen die Grenzen der experimentellen Phonetik zeigt. In diesem Zusammenhang soll noch einmal auf das Forschungsparadigma der Gesprächsanalyse hingewiesen werden, das über kontextuelle Interpretationen versucht, Zugang zu funktionalen Kategorien im Gespräch zu finden, aber kaum zu statistischen Validierungen ihrer Ergebnisse gelangt. M.E. ist eine Kombination experimenteller und interpretativer Verfahren ein geeigneter Weg, sich dem Phänomen "Gespräch" weiter zu nähern, wobei das in diesem Abschnitt gewählte Interpretationsverfahren offensichtlich noch der Verfeinerung bedarf.

#### **4.7 Klassifikationssystem II - Turnfinale prosodische Grenzen**

Das folgende Klassifikationssystem (Tabelle 30) setzt die Ergebnisse der Produktionsdatenanalyse und der Perzeptionsexperimente in Form-Funktions-Beziehungen unterschiedlicher turnfinaler prosodischer Grenzen um. Das aufgestellte System ist ein erster Ansatz zur systematischen Klassifikation interaktional relevanter phonetischer Parameter. Die Validierung des Systems erfordert weitere Forschungsarbeit.

Aus der Datenlage werden sechs zu differenzierende Klassen turnfinaler prosodischer Grenzen abgeleitet. Da die phonetischen Strukturen an prosodischen Grenzen stark variieren, kann keine eindeutige Form-Funktions-Beziehung angesetzt werden. Die aufgestellten Kategorien sind deshalb als prototypische Kernbereiche zu verstehen, von denen es, bedingt durch individuelle und situative Variation, zu deutlichen Abweichungen kommen kann. Da die Ergebnisse verschiedener Untersuchungen zeigen, dass die syntaktische Struktur ein wichtiger Informationsträger für die interaktionale Strukturierung von Gesprächen ist, wird sie mit in die Definition verschiedener prosodischer Grenzen einbezogen. Hinweise aus der Literatur (Local et al. 1985) und eigene informelle Korpusuntersuchungen legen zudem nahe, dass auch Intensitätsverläufe Träger interaktional relevanter Information sein können. Aus diesem Grund wird für einige Kategorien auch die Intensitätsstruktur in das Merkmalsbündel einbezogen.

t in den Klassen PGt1-PGt6 zeigt an, dass es sich um turnfinale prosodische Grenzen handelt.

**Tabelle 30:** Klassifikationssystem II.

<b>Nomenklatur</b>	<b>Merkmalsbündel</b>	<b>Funktionen</b>
PGt1	Tief fallende melodische Bewegung, evtl. Intensitätsabnahme und Veränderung der Stimmqualität, Keine Überlagerung, Syntaktischer Abschluss	Gezielte Turnübergabe: inhaltlicher Abschluss, interaktionaler Abschluss
PGt2	Steigende melodische Bewegung, Keine Überlagerung, Syntaktischer Abschluss	Gezielte Turnübergabe mit Frage an den Gesprächspartner: inhaltliche Weiterweisung, interaktionaler Abschluss
PGt3	Tief fallende melodische Bewegung, Überlagerung, Syntaktischer Abschluss wird während der Überlagerung erreicht	Erlaubte Turnübernahme: inhaltlicher Abschluss, interaktionale Weiterweisung
PGt4	Keine starke melodische Bewegung Überlagerung, Syntaktischer Abschluss / Kein syntaktischer Abschluss	Erlaubte Turnübernahme: inhaltliche Weiterweisung, interaktionale Weiterweisung
PGt5	Keine starke melodische Bewegung, Überlagerung, Hohe Intensität während der Überlagerung, Syntaktischer Abschluss / Kein syntaktischer Abschluss	Erzwungene Turnübernahme: inhaltliche Weiterweisung, interaktionale Weiterweisung

## 5 Prosodische Grenzen in unflüssiger Sprache

In diesem Kapitel wird auf prosodische Grenzen eingegangen, die im Zusammenhang mit Produktionsstörungen entstehen. Die Betrachtung dieses Phänomenbereichs ist für die funktionale Analyse prosodischer Grenzen wichtig, da durch Produktionsstörungen hervorgerufene prosodische Grenzen in spontaner Sprache häufig sind und in ihrer sprachlichen Funktion von inhaltlich gliedernden und gesprächsstrukturierenden prosodischen Einschnitten getrennt werden müssen. Die globale Struktur dieses Kapitels ist vergleichbar mit den Kapiteln zu den flüssigen turninternen und den turnfinalen Grenzen. Es enthält einen Abschnitt zur Produktionsdatenanalyse, wie auch einen experimentellen Teil. Durch diese Untersuchungen sollte ursprünglich auch eine funktional subklassifizierende Einteilung verschiedener unflüssiger Grenzen erreicht werden. Im Laufe der Beschäftigung mit dem Phänomenbereich wurde aber dessen enorme Komplexität immer deutlicher, so dass auf dem derzeitigen Wissenstand kein Klassifikationssystem vorgeschlagen werden kann. Auch auf einen umfassenden Literaturüberblick soll hier verzichtet werden. Einige Aspekte unflüssiger Sprechweise sind bereits in den Literaturüberblicken zu den flüssigen turninternen Grenzen und den turnfinalen Grenzen enthalten. Statt dessen werden vor allem grundlegende Unterschiede zwischen flüssigen und unflüssigen prosodischen Grenzen anhand der Produktionsdatenanalyse herausgearbeitet und in Perzeptionsexperimenten die perzeptorische Relevanz melodischer und temporaler Strukturen bei der Wahrnehmung von Unflüssigkeit überprüft.

In spontaner Sprache kommt es oft zu Planungsproblemen und damit zu Äußerungen, die Fehler enthalten. Diese Fehler (z.B. inhaltlich falsche Aussagen, syntaktische Fehlkonstruktionen oder artikulatorische Fehler) werden häufig von den Sprechern selbst unmittelbar bemerkt. Es ist aus den untersuchten Produktionsdaten bekannt, dass solche Fehler in der Planung und/oder der Ausführung eines Redebeitrags oft prosodische Grenzen verursachen, die dadurch entstehen, dass der Sprecher zögert oder sich selbst unterbricht. Neben eindeutigen Fehlern kommt es durch Probleme bei der simultan zum Sprechen ablaufenden Äußerungsplanung auch immer wieder zu Verzögerungen wie Pausen und segmentellen Längungen, die nicht wie in flüssiger Sprache durch die inhaltliche Gliederungsfunktion erklärt werden können, aber z.T.

deutliche perzeptive Einschnitte verursachen. Daraus erwächst die Frage, wie sich prosodische Grenzen in flüssiger Sprache von prosodischen Grenzen in unflüssiger Sprache unterscheiden und welche phonetischen Merkmale von Hörern als Hinweise auf Planungsprobleme bzw. eine unflüssige Sprechweise wahrgenommen werden. Denn für Hörer ist es wichtig, im Gespräch zwischen Einschnitten, die inhaltliche und syntaktische Strukturen widerspiegeln, und solchen, die aufgrund von Produktionsproblemen auftreten, zu unterscheiden. Geschieht dies nicht, kann das Verstehen eines Redebeitrags und eine angemessene Hörerreaktion im Rahmen eines flüssigen Gesprächsablaufs erheblich erschwert werden.

Der Entstehung von Sprechflussstörungen liegen unterschiedliche Ursachen zugrunde. Folgende Probleme können zu einer unflüssigen Sprechweise führen:

- **Phonetische Versprecher:** Während des Sprechens kann es zu Lautvertauschungen, Fehlartikulationen und anderen artikulatorischen Problemen kommen.
- **Syntaktische Versprecher:** Ein bereits zum Teil ausgesprochener Satz enthält agrammatische Strukturen oder lässt sich nicht sinnvoll ohne eine Verletzung grammatischer Regeln fortsetzen.
- **Inhaltliche Versprecher:** Ein Wort, das nicht der Intention des Sprechers entspricht, wird versehentlich ganz oder zum Teil ausgesprochen und eine Korrektur wird notwendig. In diese Kategorie fallen auch die sog. Freudschen Versprecher (Freud 1901/1971).
- **Wortfindungsprobleme:** Ein Wort, das eigentlich bekannt ist, kann nicht rechtzeitig aus dem Gedächtnis abgerufen werden.
- **Syntaktische Planungsprobleme:** Eine korrekte syntaktische Fortsetzung kann nicht so schnell geplant werden, dass die derzeitige Äußerung flüssig fortgesetzt werden kann.

Die beschriebenen möglichen Ursachen unflüssigen Sprechens haben eines gemeinsam: Der Sprecher weiß im Grunde, was er sagen will, hat aber Probleme bei der Gestaltung der intendierten Äußerung (artikulatorisch, syntaktisch, lexikalisch oder inhaltlich). Hierdurch kann es zu einer unflüssigen Sprechweise kommen, die sich u.U. durch starke prosodische Einschnitte, also prosodische Grenzen, manifestiert.

Unabhängig davon, welche Ursache unflüssiges Sprechen hat, kann m.E. davon ausgegangen werden, dass Hörer Unflüssigkeiten wahrnehmen und dadurch Rückschlüsse auf den mentalen Sprachplanungsprozess des Sprechers möglich werden: Der Planungsprozess wird z.T. durch die phonetische Struktur der Äußerung offenbart, da Planungsprobleme durch bestimmte Exponenten im Signal repräsentiert werden. Das Nachvollziehen des Planungsprozesses hilft dem Hörer bei der inhaltlichen und syntaktischen Interpretation des bereits Gehörten und bei der Einschätzung des weiteren Äußerungs- bzw. Gesprächsverlaufs.

Auch Selting (1995) geht davon aus, dass Produktionsstörungen durch bestimmte prosodische Signale erkennbar werden und eine geplante Fortsetzung oder einen Neustart anzeigen. Sie hebt hierbei vor allem starke segmentelle Längungen in Verbindung mit ebenen melodischen Verläufen hervor. Selting (1995:75)

*Treten bei der Produktion eines Satzes interne Produktionsstörungen oder -verzögerungen auf, dann können diese entweder mit den o.g. prospektiven und retrospektiven lokalen Turnhaltesignalen als Einheiten-interne kontextualisiert werden oder aber es kann ein Neustart erfolgen, der mit den entsprechenden Signalen angezeigt wird. D.h.: Bei global projektierenden Einheiten wie 'Sätzen' können prosodische Signale, z.B. bei Störungen, Fortsetzungen lokal als kohäsive Erweiterungen oder aber als Neuansätze kontextualisieren; die prosodischen Signale kontextualisieren damit lokal den Status und die Beziehung von Elementen zueinander. Eine solche gleichbleibende Halte-Tonhöhenbewegung projiziert dann melodisch eine zu folgende Fortsetzung der Kontur mit lokal fallender oder steigender letzter Tonhöhenbewegung.*

Das grundlegende Problem in der Differenzierung flüssiger und unflüssiger prosodischer Grenzen ist aber, dass in natürlicher Sprache für beide Typen der phonetischen Separierung trotz des funktionalen Unterschieds ähnliche phonetische Parameter zum Einsatz kommen. Vor allem Pausen und Atmungsphasen sowie segmentelle Längungen sind sowohl an flüssigen als auch an unflüssigen Grenzen zu finden. Auch gleich-

bleibende melodische Bewegungen finden sich an prosodischen Grenzen, die beim Abhören nicht als unflüssig erscheinen. Andere Phänomene sind allerdings typisch für prosodische Grenzen, die durch Planungsprobleme entstehen (z.B. Häsitationspartikeln).

Das folgende Kapitel umfasst eine Analyse der im Korpus gefundenen Merkmale unflüssiger prosodischer Grenzen und stellt über Perzeptionsexperimente Beziehungen zwischen akustisch/phonetischen und perzeptiven Größen bei der Wahrnehmung von Sprechflussstörungen her. Unflüssige prosodische Grenzen werden hierbei vor allem in turninterner Position analysiert, denn es soll primär die phonetische Struktur von unflüssigen Grenzen untersucht werden, die Sprecher ohne direktes Zutun des Gesprächspartners produzieren und weniger Unflüssigkeiten, die z.B. durch Unterbrechungen durch den Partner entstehen können.

## **5.1      Symbolbasierte Analyse der Produktionsdaten**

Für die Analyse der phonetischen Merkmale unflüssiger prosodischer Grenzen stellt sich zuerst die Frage, wie im Rahmen einer labelbasierten Korpusabfrage der Zugriff auf unflüssige Grenzen vorgenommen werden kann, denn in der Etikettierung ist keine explizite Trennung zwischen flüssigen und unflüssigen Grenzen enthalten. Da es für die Analyse der phonetischen Merkmale aber notwendig ist, diese Trennung zu erreichen, müssen Kriterien gefunden werden, um gezielt auf prosodische Grenzen zuzugreifen, die im Zusammenhang mit Sprechflussstörungen stehen. Hierfür werden drei Kriterien gewählt:

- In der Etikettierung sind syntaktische Abbrüche gekennzeichnet. Wenn an der Position eines syntaktischen Abbruchs eine prosodische Grenze etikettiert ist, kann davon ausgegangen werden, dass es sich nicht um eine gezielte, inhaltlich gliedernde Grenze handelt, sondern dass es an dieser Position der Äußerung Planungsprobleme gab, die den syntaktischen Abbruch bedingten und ebenfalls zu dem prosodischen Einschnitt geführt haben. Die Klasse der prosodischen Grenzen an syntaktischen Abbrüchen bildet also die erste Gruppe untersuchter prosodischer Grenzen.

- Häsitationspartikeln wie *äh* oder *ähm* sind in den Etikettierdateien markiert. Diese Partikeln werden normalerweise dann von Sprechern produziert, wenn der weitere Äußerungsverlauf geplant wird und dies nicht schnell und flüssig während des fortlaufenden flüssigen Sprechens bewältigt werden kann. Prosodische Grenzen, an denen solche Partikeln vorkommen, stehen daher ebenfalls in Verbindung mit Planungsproblemen und bilden eine weitere Gruppe der Analyse.
- Der dritte Hinweis auf Unflüssigkeit an prosodischen Grenzen liegt in der Etikettierung der segmentellen Dauerstruktur. Die Etikettierer haben im Rahmen der segmentellen Etikettierung an den Positionen, wo sie eine deutliche Zögerungslängung wahrgenommen haben, das Etikett *z:* eingefügt. Diese Kennzeichnung ist häufig im Korpus zu finden und zwar dort, wo einzelne oder mehrere Segmente ungewöhnlich lange Dauern aufweisen. Die Klasse von prosodischen Grenzen, an denen eine solche Markierung gesetzt wurde, bildet den dritten Untersuchungsbereich bei der automatischen Korpusabfrage.



Audiobeispiele: syntaktischer Abbruch: 3.8, Häsitationspartikel: 3.2,

Zögerungslängung: 3.5

In den folgenden Abschnitten werden die extrahierbaren phonetischen Merkmale dieser drei Klassen von unflüssigen prosodischen Grenzen getrennt dargestellt (LINDENSTRASSE-Korpus). Um zu einer klaren Trennung der drei automatisch abfragbaren Analysebereiche zu kommen, wird für jede Klasse das Auftreten der anderen beiden Merkmale ausgeschlossen, wenn also z.B. ein syntaktischer Abbruch zusammen mit einer nachfolgenden Häsitationspartikel vorkommt, wird diese prosodische Grenze in keine der drei Klassen einbezogen. Auf diese Weise können die drei Analysebereiche klar voneinander getrennt werden. In die Auswertung werden nur prosodische Grenzen einbezogen, an denen im Rahmen der Etikettierung eine phrasenfinale Kontur zugeordnet wurde. In vielen Fällen gilt dies nicht, da bei Phrasenfragmenten, die durch eine prosodische Grenze abgebrochen werden, bevor eine satzakzentuierte Silbe in der Phrase auftritt, in der Etikettierung auf eine Kennzeichnung des finalen Melodiemusters verzichtet wird. Folgende Klassen ergeben sich:

- 204 Fälle von prosodischen Grenzen an syntaktischen Abbrüchen (ohne Zögerungslängung und ohne nachfolgende Häsitationspartikel),
- 204 Fälle von prosodischen Grenzen mit nachfolgender Häsitationspartikel (ohne Zögerungslängung und ohne syntaktischen Abbruch),
- 129 Fälle von prosodischen Grenzen mit Zögerungslängung (ohne nachfolgende Häsitationspartikel und ohne syntaktischen Abbruch).

Im LINDENSTRASSE-Korpus kommen an insgesamt 288 prosodischen Grenzen Kombinationen der drei Merkmale vor, die nicht in die Analyse eingehen. Folgende Kombinationen treten auf:

- Syntaktischer Abbruch und Häsitationspartikel: 126 Fälle,
- Syntaktischer Abbruch und Zögerungslängung: 74 Fälle,
- Zögerungslängung und Häsitationspartikel: 55 Fälle,
- Syntaktischer Abbruch, Zögerungslängung und Häsitationspartikel: 33 Fälle.

Sicher gibt es im Korpus auch eine Vielzahl prosodischer Grenzen, die durch Verzögerungen im simultanen Planungsprozess entstehen, ohne eines der oben genannten Etikettier-Merkmale zu zeigen. Auf solche Grenzen könnte lediglich durch eine umfangreiche Interpretation prosodischer Grenzen zugegriffen werden. Deshalb beziehen sich alle nachfolgenden Analysen in diesem Abschnitt auf die oben eingeführten Klassen.

### 5.1.1 Melodische Muster

Tabelle 31 zeigt die etikettierten melodischen Muster an den drei Typen unflüssiger prosodischer Grenzen.

**Tabelle 31:** Melodische Verläufe an turninternen unflüssigen Grenzen (LINDENSTRASSE).

	<b>Syntaktischer Abbruch (n=204)</b>	<b>Häsitations- partikel (n=204)</b>	<b>Zögerungs- längung (n=129)</b>	<b>Gesamt (n=537)</b>
<b>Starke melodische Bewegung</b>				
Tief fallend	19 (9,3%)	26 (12,8%)	13 (10,1%)	58 (10,8%)
Tief fallend + sehr leicht steigend	2 (1,0%)	2 (1,0%)	1 (0,8%)	5 (0,9%)
Fallend-steigend	1 (0,5%)	8 (3,9%)	1 (0,8%)	10 (1,9%)
Hoch steigend	0 (0%)	3 (1,5%)	0 (0%)	3 (0,6%)
<b>Gesamt starke melodische Bewegung</b>	<b>22 (10,8%)</b>	<b>39 (19,1%)</b>	<b>15 (11,6%)</b>	<b>76 (14,2%)</b>
<b>Schwache melodische Bewegung</b>				
Eben	47 (23%)	61 (29,9%)	22 (17,1%)	130 (24,2%)
Leicht fallend	125 (61,3%)	85 (41,7%)	80 (62,0%)	290 (54,0%)
Leicht bis mittelhoch steigend	3 (1,5%)	12 (5,9%)	3 (2,3%)	18 (3,4%)
Eben + sehr leicht steigend	0 (0%)	0 (0%)	3 (2,3%)	3 (0,6%)
Leicht fallend + sehr leicht steigend	7 (3,4%)	7 (3,4%)	6 (4,6%)	20 (3,7%)
<b>Gesamt schwache mel. Bewegung</b>	<b>182 (89,2%)</b>	<b>165 (80,9%)</b>	<b>114 (88,4%)</b>	<b>461 (85,9%)</b>
<b>Gesamt</b>	<b>204 (100%)</b>	<b>204 (100%)</b>	<b>129 (100%)</b>	<b>537 (100%)</b>

Die Verteilungen in den drei Klassen sind ähnlich. In allen Klassen ist der Anteil der starken melodischen Bewegungen an der prosodischen Grenze relativ klein, wobei fast alle Fälle starker melodischer Bewegungen tief fallende Muster sind. Der auffälligste Unterschied ist, dass die leicht fallenden Konturen an syntaktischen Abbrüchen und an Grenzen mit Zögerungslängung über 60% der Fälle ausmachen, während es für die Klasse mit nachfolgender Häsitationspartikel nur knapp 42% sind, sich hier aber ein erhöhter Anteil ebener Konturen findet. Dieser erhöhte Anteil resultiert evtl. daraus, dass

Äußerungsteile vor Häsitationspartikeln häufiger flüssig und inhaltlich komplett abgeschlossen werden und die Planungsschwierigkeiten einen weiteren nachfolgenden inhaltlichen Block betreffen. Bei syntaktischen Abbrüchen hingegen ist das Planungsproblem jedoch eindeutig im aktuellen inhaltlichen Block lokalisiert, da ein inhaltlicher Abschluss ohne syntaktischen Abschluss nicht zu erwarten ist.

Der hohe Anteil leicht fallender Konturen steht im Gegensatz zu den Beobachtungen von Selting (1995), die vor allem ebene F0-Verläufe als melodische Signale für unflüssige prosodische Grenzen ansieht. Weiterhin ist auffällig, dass der leicht fallende melodische Verlauf auch an turninternen flüssigen Grenzen eine hohe Auftretenshäufigkeit hat (28,3%, siehe Tabelle 3). Im Lindenstrasse-Korpus ist dieses Muster auch in der turnfinalen Position mit 16,7% nicht selten (siehe Tabelle 19).



Audiobeispiele: unterschiedliche melodische Muster an unflüssigen prosodischen

Grenzen: 3.1 - 3.4

## 5.1.2 Pausen und Atmungsphasen

An den drei Typen von unflüssigen prosodischen Grenzen finden sich Pausen und Atmungsphasen mit folgenden Anteilen: Grenzen mit syntaktischen Abbrüchen: 36,8% (75 Fälle), Grenzen mit nachfolgenden Häsitationspartikeln: 21,1% (43 Fälle), Grenzen mit Zögerungslängung: 28,7% (37 Fälle). Dieses Ergebnis zeigt, dass Sprechpausen in vielen Fällen als Korrelate unflüssiger Sprechweise vorkommen. Möglicherweise ist die etwas geringere Häufigkeit von Pausen an prosodischen Grenzen mit Häsitationspartikeln und Zögerungslängung damit zu erklären, dass diese phonetischen Phänomene eine Art Ersatz für Pausen darstellen und dadurch deutlich die momentanen Planungsprobleme signalisieren und das Rederecht sichern ("gefüllte Pausen").

### 5.1.3 Längung

Zunächst war auch eine automatische labelbasierte Quantifizierung der segmentellen Längung an unflüssigen prosodischen Grenzen geplant. Nach eingehenden Voruntersuchungen unflüssiger Grenzen wurde diese Vorhaben jedoch eingestellt. Hierfür gibt es folgende Gründe:

- Die segmentelle Dauerstruktur an unflüssigen prosodischen Grenzen unterliegt einer starken Variation in der Position der gelängten Abschnitte. Während sich die segmentelle Längung an flüssigen prosodischen Grenzen relativ zuverlässig auf den finalen Reim beschränkt, dehnt sich Zögerungslängung teilweise über ganze Silbenketten aus oder betrifft einzelne Segmente mitten in Wörtern. Die segmentelle Etikettierung erlaubt keine genaue Positionsbestimmung der etikettierten Zögerungslängungen. Wenn z.B. einem Wort das Label *z:* folgt, das wahrgenommene Zögerungslängung markiert, ist nicht ableitbar, ob die Längung das ganze Wort oder nur das letzte Segment betrifft. Messungen, die auf dieser Basis durchgeführt werden könnten, wären im Nachhinein kaum interpretierbar und geben wenig Aufschluss über tatsächliche Dauerstrukturen. Ein Beispiel für diese Problematik sind die Dauerstrukturen von drei unakzentuierten Realisierungen des Wortes *weil*, die die Sprecherin APE im Dialog 103a innerhalb eines Ausschnitts von fünf Sekunden unflüssiger Sprechweise mit einer starken Variation in der segmentellen Dauerstruktur produziert. Tabelle 32 zeigt die Segmentdauern der drei Realisierungen. Die Realisierung in Zeile 1, im Gegensatz zu den beiden anderen, wirkt nicht gezögert. Die Zahlen zeigen, dass in Realisierung 2, im Vergleich zu Realisierung 1, vor allem der Silbenonset und der Silbennukleus gelängt sind, in Realisierung 3 hingegen der Onset und der Offset der Silbe. Eine Mittelwertbildung über die gesamte Silbe würde also keinen geeigneten Messwert zur Beschreibung der Dauerstruktur ergeben, da die Unterschiede in der Position der gelängten Segmente hierdurch nicht erfasst werden könnten.

**Tabelle 32:** Segmentdauern von drei Realisierungen des Wortes *weil* (in ms).

	[v]	[ai]	[l]
<b>Realisierung 1</b>	30	140	70
<b>Realisierung 2</b>	180	415	110
<b>Realisierung 3</b>	100	130	230



Audiobeispiel: Segmentdauern von drei Realisierungen des Wortes *weil*: 3.7

- Weiterhin ist in den Daten eine große individuelle Fluktuation in Art und Häufigkeit segmenteller Längungen bei Produktionsstörungen beobachtbar. Einzelne Sprecher scheinen deutlich unterschiedliche Strategien zu verfolgen, wenn sie Probleme mit der Äußerungsplanung haben. Einige Sprecher verwenden viele Häsitationspartikeln, andere statt dessen ungefüllte Pausen, dritte wiederum starke segmentelle Dehnungen. Eine wirklich aussagekräftige Untersuchung müsste folglich zuerst das Verhalten unterschiedlicher Sprecher untersuchen, bevor verallgemeinernde Messungen über Korpora mit vielen Sprechern vorgenommen werden.



Audiobeispiele: individuell unterschiedliche Ausprägungen unflüssiger Sprechweise: 3.1, 3.2, 3.3

- Der dritte Punkt, der eine quantitative Untersuchung segmenteller Längung bei Sprechflussstörungen fragwürdig macht, hängt mit der generellen Konzeption der vorliegenden Arbeit zusammen. Ziel der Arbeit ist die Analyse phonetischer Merkmalsbündel an prosodischen Grenzen. Im Korpus tritt aber etikettierte Zögerungslängung auch an vielen Positionen auf, an denen keine prosodische Grenze etikettiert wurde. Insgesamt sind im LINDENSTRASSE-Korpus 403 Fälle von Zögerungslängung etikettiert. Davon stehen 282 Fälle in Verbindung mit einer unmittelbar folgenden prosodischen Grenze. Dass in vielen Fällen keine prosodische Grenze etikettiert wurde, liegt daran, dass die vorhandene Längung nicht zu einem deutlichen perzeptorischen Einschnitt geführt hat. Auch die Perzeptionsexperimente 1a, 1b und 1c (siehe 3.3.2) belegen, dass segmentelle Längung allein nicht zu einer deutlichen Separierung führt. Eine inhaltlich sinnvolle Untersuchung von Zögerungslängung müsste also auch die

Längungen mit einbeziehen, die ohne prosodische Grenzen vorkommen und würde so zu einer erheblichen Erweiterung der ursprünglichen Fragestellung führen.



Audiobeispiele: Längung ohne deutliche perzeptorische Separierung: 3.6, 3.7

### 5.1.4 Phonation

Wie unter 4.2.1.4 beschrieben, ist an syntaktischen Abbruchstellen z.T. Knarrstimme zu beobachten. Tabelle 33 zeigt die Häufigkeiten verschiedener phonatorischer Ausprägungen an prosodischen Grenzen mit syntaktischen Abbrüchen. Für die Analyse werden sämtliche syntaktische Abbruchstellen, die mit einer prosodischen Grenze zusammenfallen einer manuellen akustisch-phonetischen Analyse unterzogen (siehe 3.2.11). Diese Analyse bezieht sowohl turninterne als auch turnfinale syntaktische Abbrüche ein. Die untersuchten Fälle sind nach ihrer Position im Turn getrennt dargestellt. Die Anzahl der syntaktischen Abbrüche liegt hier höher als in den vorigen Abschnitten, da hier auch die prosodischen Grenzen einbezogen werden, für die kein komplettes Merkmalsbündel abgeleitet werden kann.

**Tabelle 33:** Phonationstyp an syntaktischen Abbrüchen (LINDENSTRASSE).

	<b>Turninterne syntaktische Abbrüche (n=544)</b>	<b>Turnfinale syntaktische Abbrüche (n=85)</b>	<b>Gesamt (n=629)</b>
<b>Modalstimme</b>	69,1%	71,8%	69,5%
<b>Knarrstimme (unregelmäßig)</b>	6,3%	4,7%	6,0%
<b>Knarrstimme (regelmäßig)</b>	3,1%	5,9%	3,5%
<b>Glottalverschluss</b>	5,5%	12,9%	6,5%
<b>Rest (Interferenz und unsichere Beurteilung)</b>	16,0%	4,7%	14,5%
<b>Gesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Die durchschnittliche Dauer von Knarrstimme an Abbruchstellen beträgt 96 ms. Die durchschnittliche Dauer von Glottalverschlüssen an Abbruchstellen beträgt 393 ms mit einer hohen Standardabweichung von 743 ms. Der melodische Verlauf an syntaktischen Abbruchstellen ist in der Regel leicht fallend (62,4%) oder eben (23,6%). Tief fallende

Konturen treten nur in 9,5% aller Fälle auf. Glottalisierungen an den Abbruchstellen setzen zum größten Teil aus leicht fallenden oder ebenen Konturen heraus ein (52,2% bzw. 34,8%).



Audiobeispiel: Knarrstimme und Glottalverschluss an syntaktischem Abbruch: 3.8

Die Verteilungen zeigen, dass das Vorkommen von Knarrstimme an syntaktischen Abbrüchen zwar nicht selten ist, dass Knarrstimme (mit oder ohne Glottalverschluss) aber auch nicht der Regelfall ist. Nimmt man alle drei Kategorien zusammen, die mit phonatorischen Veränderungen in Verbindung stehen (Knarrstimme regelmäßig/unregelmäßig und Glottalverschluss) ergibt sich für die turninterne Position eine relative Häufigkeit von 14,9% und für die turnfinale ein Anteil von 23,5%. Der erhöhte Anteil in turnfinaler Position geht vor allem auf den größeren Anteil turnfinaler Glottalverschlüsse zurück. Im Gegensatz zu Knarrstimme an flüssigen prosodischen Grenzen kann Knarrstimme an Abbruchstellen mit Glottalverschlüssen alternieren. Insoweit decken sich die Ergebnisse mit den Befunden von Kohler, Peters & Wesener 2005. Allerdings ist in den Daten aus dem LINDENSTRASSE-Korpus ein hoher Anteil von Glottalverschlüssen in turnfinaler Position festzustellen. Dies ist kaum in Einklang zu bringen mit der Interpretation von Local & Kelly (1986), die gehaltene Glottalverschlüsse als 'turn-holdig-device' funktional einordnen. Hieraus ergibt sich, dass Glottalverschlüsse vor allem in turninterner Position zu finden sein müssten. Die Daten zeigen aber eine gegenläufige Tendenz.

Es scheint vielmehr so zu sein, dass die hier beobachteten Glottalverschlüsse in einem anderen interaktionalen Zusammenhang vorkommen. Neun der zwölf beobachteten Glottalverschlüsse in turnfinaler Position treten während einer Überlagerung durch den Gesprächspartner auf. Der ursprüngliche Sprecher beendet den Turn also mit einem Glottalverschluss, ohne die begonnene syntaktische Struktur abzuschließen. Dies lässt die Interpretation zu, dass Sprecher Glottalverschlüsse z.T. einsetzen, um sich selbst zu unterbrechen, wenn der Partner überlagernd zu sprechen beginnt. Die auditive Überprüfung dieser Fälle zeigt, dass in einigen Fällen der Glottalverschluss so lange gehalten wird, bis der unterbrochene Sprecher das Wort wieder übernehmen kann. Aus diesem Verhalten erklärt sich auch die große Standardabweichung der Verschlussdauern. In einem gewissen Sinn lässt sich dieses Verhalten dann auch wieder mit der Funktion des

Turnhaltens in Verbindung bringen, denn die Selbstunterbrechung mit Glottalverschluss ist hier eine Art vorläufiger Turnübergabe und der Sprecher signalisiert durch die Art des Abschlusses, dass er nicht fertig gesprochen hat und bei nächster Gelegenheit wieder den Turn übernehmen wird. Der Sprecher schneidet sich hierbei gewissermaßen selbst die Luft ab und begibt sich mit durch den Verschluss aufgestauter Luft in "Wartstellung". Gelingt es dem Sprecher nicht, wieder an das Rederecht zu kommen, findet sich in einigen Fällen ein hörbares Lösungsgeräusch des Glottalverschlusses, das von einem deutlichen Ausatmen begleitet wird und anzeigt, dass der Sprecher den Plan der unmittelbaren Wiederübernahme des Rederechts aufgegeben hat.

Da das Vorkommen von Glottalisierung und Glottalverschluss an Abbruchstellen keineswegs die Regel ist, muss in die Interpretation und funktionale Einordnung von Glottalverschlüssen an syntaktischen Abbruchstellen ein weiterer Aspekt einbezogen werden, der sich aus der morphophonologischen Struktur des Deutschen ergibt: Wenn im Deutschen morpheminitial ein Vokal steht, erfordert dies in der phonetischen Realisierung (zumindest in satzakzentuierter Position) den so genannten *harten Vokaleinsatz*. Dieser harte Einsatz kann entweder durch einen dem Vokal voran gestellten Glottalverschluss oder durch Glottalisierung des Vokalonsets realisiert werden. Es ist also möglich, dass die Knarrstimme am Ende syntaktischer Abbrüche bereits einen nachfolgenden harten Vokaleinsatz ankündigt, die eigentliche Realisierung des geplanten Wortes aber ausbleibt. Diesem Problem wird in der Analyse dadurch begegnet, dass alle syntaktischen Abbrüche, denen ein Morphem mit Vokal im Onset folgt, der Restkategorie zugeordnet wird und nicht mit in die eigentliche Auswertung eingeht. Für die verbleibenden Fälle kann aber nicht ausgeschlossen werden, dass ein Sprecher zum Zeitpunkt des Abbruchs bereits eine Äußerungsfortführung geplant hatte, die mit einem Vokal beginnt und die sich bereits am Ende des syntaktischen Fragments durch Glottalisierung ankündigt, der Sprecher dann aber diese Fortführung verwirft und die Äußerung mit einem Wort ohne Vokal im Onset fortsetzt. Fälle dieser Art auszuschließen, ist grundsätzlich nicht möglich, da Vorhersagen über eine geplante und wieder verworfene Struktur nicht aus den Daten ableitbar sind.



Audiobeispiele: Glottalverschluss bei Unterbrechung durch den Gesprächspartner:

2.12, 2.15

### **5.1.5 Zusammenfassung**

Die phonetischen Repräsentationen unflüssiger Sprechweise sind sehr variabel. Zwei typische Ausprägungen können anhand der Daten belegt werden, zum einen prosodische Grenzen mit Zögerungslängung und zum anderen prosodische Grenzen ohne Zögerungslängung, aber mit nachfolgender Häsitationspartikel. Für beide Klassen gilt, dass der melodische Verlauf vor der prosodischen Grenze meist leicht fallend oder eben verläuft. Möglicherweise sind die beiden Ausprägungen sprechertypisch. Allerdings sind auch Kombinationen von Zögerungslängung und Häsitationspartikeln zu beobachten. Beide Strategien führen dazu, dass über die gedehnten Segmente Zeit für den Planungsvorgang gewonnen wird und dem Hörer zugleich signalisiert wird, dass der Sprecher beabsichtigt weiterzusprechen. Auch Knarrstimme (z.T. mit Glottalverschluss) tritt in Verbindung mit syntaktischen Abbrüchen auf. Ein durchgehender systematischer Einsatz dieses phonetischen Parameters zeigt sich jedoch nicht.

## **5.2 Perzeptionsexperimente zur Sprechflüssigkeit (Experiment 3a und 3b)**

Die folgenden Experimente untersuchen den Zusammenhang zwischen der phonetischen Struktur von Äußerungen und der Wahrnehmung von unflüssiger Sprechweise. Das experimentelle Vorgehen ist vergleichbar mit den Experimenten zu den flüssigen turninternen prosodischen Grenzen und den turnfinalen Grenzen. Die im Korpus beobachtete phonetische Variation wird systematisch in Stimulusserien überführt, und Versuchspersonen beurteilen die Stimuli auf Antwortskalen. In den Experimenten zur unflüssigen Sprechweise wird eine Skala zwischen den Polen 'zögerlich' und 'nicht zögerlich' verwendet. Das erste Experiment überprüft die perzeptorischen Einflüsse der segmentellen Dauerstruktur und des melodischen Verlaufs auf die Wahrnehmung von unflüssiger Sprechweise. Im zweiten Experiment wird zusätzlich die Wechselwirkung zwischen syntaktischer und prosodischer Struktur bei der Wahrnehmung von Sprechflussstörungen untersucht.

## **5.2.1 Experiment zur Wahrnehmung unflüssiger Sprechweise (Experiment 3a)**

### **5.2.1.1 Hypothesen**

Folgende Hypothesen werden geprüft:

#### **Hypothese 1**

Längung des phrasenfinalen Reims führt zur Wahrnehmung von unflüssiger Sprechweise, wenn sie einen Schwellenwert überschreitet. Wird die Längung oberhalb des Schwellenwertes noch vergrößert, verstärkt sich die wahrgenommene Unflüssigkeit weiter.

#### **Hypothese 2**

Dehnt sich die phrasenfinale Längung über einen größeren Bereich als den phrasenfinalen Reim aus, führt dies bei einem vergleichbaren Längungsfaktor stärker zur Wahrnehmung von unflüssiger Sprechweise.

#### **Hypothese 3**

Werden in einer flüssig gesprochenen prosodischen Phrase einzelne Segmente im Phraseninneren gelängt, führt dies zur Wahrnehmung von unflüssiger Sprechweise.

#### **Hypothese 4**

Starke melodische Bewegungen dämpfen den Eindruck von Unflüssigkeit ab. D.h., dass bei ansonsten gleicher Merkmalskombination die ebenen, leicht fallenden und leicht steigenden Muster eher mit Unflüssigkeit assoziiert werden als die tief fallenden oder hoch steigenden.

#### **Hypothese 5**

Glottalverschlüsse am Ende der Äußerung verstärken die Wahrnehmung unflüssiger Sprechweise.

### 5.2.1.2 Stimulusmaterial

Die verwendeten Stimuli sind identisch mit den Stimuli aus dem Experiment zur Turnübergabeerwartung (siehe 4.3.2). Lediglich eine zusätzliche Serie wird in den Versuch aufgenommen, in der einzelne Segmente im Inneren der Äußerung gelängt werden. Es folgt eine kurze Zusammenfassung.

Der Trägersatz ist: *Das ist ja nun nicht so angenehm.*

Folgende melodische Verläufe werden äußerungsfinal resynthetisiert:

- tief fallend                      -8 Ht (von 111 auf 70 Hz fallend)
- leicht fallend                    -4 Ht (von 111 auf 88 Hz fallend )
- eben                                -0 Ht (gleichbleibend bei 111 Hz)
- leicht steigend                    +4 Ht (von 111 auf 140 Hz steigend)
- hoch steigend                    +8 Ht (von 111 auf 176 Hz steigend)

Ausgehend von einem Stimulus ohne finale Längung werden folgende Längungsfaktoren auf den finalen Reim von *angenehm* angewendet.

- Faktor 1,0 (Reimdauer: 250 ms)
- Faktor 1,4 (Reimdauer: 350 ms)
- Faktor 1,8 (Reimdauer: 450 ms)
- Faktor 2,2 (Reimdauer: 550 ms)

Folgende Längungsfaktoren werden auf das letzte Wort der Äußerung (*angenehm*) angewendet:

- Faktor 1,0 (Wortdauer: 515 ms)
- Faktor 1,4 (Wortdauer: 720 ms)
- Faktor 1,8 (Wortdauer: 930 ms)
- Faktor 2,2 (Wortdauer: 1130 ms)

Weiterhin wird eine Serie mit finaler Glottalisierung, die in einem Glottalverschluss endet einbezogen (mit Längungsfaktor 1,4 und allen melodischen Mustern) und ein Stimulus mit tief fallender Intonation und Längungsfaktor 1,4, der in Laryngalisierung ausläuft.

Eine weitere Serie kommt hinzu, indem der Nasal in *nicht* mit Faktor 2 bzw. Faktor 3 gelängt wird. Diese Längung wird auf die Stimuli mit leicht gelängten finalen Reimen (Faktor 1,4) angewendet, wobei alle 5 finalen Konturen einbezogen werden. So entstehen weitere 10 Stimuli. Diese Serie wird in das Experiment aufgenommen, da vergleichbare Fälle im Korpus oft beobachtbar sind. In den meisten dieser Fälle ist keine prosodische Grenze etikettiert und auch in den resynthetisierten Stimuli entsteht nicht der Eindruck eines deutlichen prosodischen Einschnitts.



Audiobeispiele: Zögerungslängung wortinitialer Segmente: 3.3, 3.6

Das Stimulusmaterial im Überblick:

- Basisstimuli: vier Längungsfaktoren x fünf Konturen = zwanzig Stimuli
- Ausgedehnte Längung (über das ganze letzte Wort): drei Längungsfaktoren x fünf Konturen = fünfzehn Stimuli.
- Glottalisierung: ein Längungsfaktor x fünf Konturen = fünf Stimuli
- Laryngalisierung: ein Längungsfaktor x eine Kontur = ein Stimulus
- Wortinitiale phraseninterne Längung: ein Längungsfaktor x fünf Konturen x zwei internen Längungsfaktoren = zehn Stimuli
- Gesamt:  $20 + 15 + 5 + 1 + 10 = 51$  Stimuli



Audiobeispiele zu den Stimuli aus Experiment 3a: 4.3, 4.4

### 5.2.1.3 Durchführung

Die Stimuli werden mit dreifacher Wiederholung randomisiert und zu einer Audiodatei zusammengeschnitten. Zwischen den Stimuli ist immer eine Pause von vier Sekunden. Nach jeweils 10 Stimuli ist eine längere Pause und ein Piepton. In den Pausen zwischen den Stimuli geben die Versuchspersonen ihre Urteile ab.

An dem Experiment nahmen 14 Personen teil. Die Durchführung wurde mit der Durchführung von Experiment 2 kombiniert. Zu Beginn der Sitzung wurden vorher aufgenommene Instruktionen abgespielt, die den Ablauf und die Aufgabenstellung erläutern (siehe Anhang).

Hinsichtlich der Sprechflüssigkeit und ihrer Beurteilung ist m.E. von einem Kontinuum von flüssiger zu unflüssiger Sprache auszugehen. Diese Sichtweise beruht auf der Analyse der bereits untersuchten Produktionsdaten, die zeigt, dass prosodische Grenzen in unflüssiger Sprache zum einen nur sehr geringe phonetische Anzeichen von Unflüssigkeit zeigen, zum anderen aber auch Einschnitte vorkommen, die extreme Anzeichen von Unflüssigkeit aufweisen. Auch aus der Sicht des Sprechers können Produktionsprobleme als graduell abgestuft betrachtet werden. Die Suche nach einem Wort oder nach einer syntaktischen Konstruktion kann mehr oder weniger lange dauern und dadurch mehr oder weniger starke Unflüssigkeit in der Rede bedingen. Auch gibt es Sprecher, die generell den Eindruck einer stark unflüssigen Sprechweise hinterlassen, Sprecher, die durchgehend sehr flüssig reden, und Sprecher, deren Sprechweise zwischen diesen Polen liegt. So ist anzunehmen, dass diese Unterschiede von Hörern rezipiert werden und die Interpretation hinsichtlich der Sprechflüssigkeit sich daran orientiert, welche und wieviele Merkmale unflüssiger Sprechweise eine Äußerung enthält und wie stark die einzelnen Merkmale sich ausprägen. Aus diesem Grund wird auch bei der experimentellen Untersuchung wieder mit einer Skala zwischen flüssiger und unflüssiger Sprechweise gearbeitet, die diese graduellen Abstufungen erfassen kann und die folgendermaßen aufgebaut ist:

**Hört sich die Äußerung zögerlich an?**

Ja							Nein
0	1	2	3	4	5	6	

#### 5.2.1.4 Ergebnisse

Tabelle 34 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

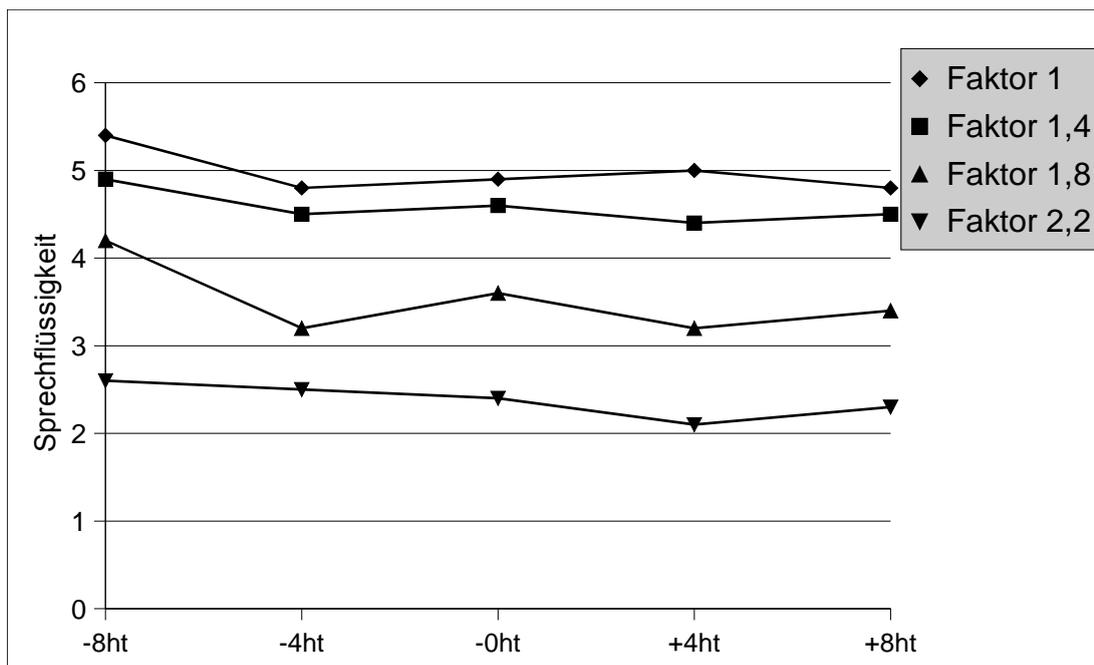
**Tabelle 34:** Ergebnisse der Varianzanalyse von Experiment 3a.

<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Längung des finalen Reims	1,39	100,41	0,000
Längung des finalen Wortes	1,12	112,13	0,000
Phraseninterne Längung	1,67	86,5	0,000
Finale Kontur (über gelängtem Reim)	1,7	4,9	0,022
Finale Kontur (über gelängtem Wort)	2,61	0,655	0,56
Phonation	1	1,5	0,242
Längung (Reim) * Kontur (Reim)	5,89	0,8	0,571
Längung (Wort) * Kontur (Wort)	3,38	1,83	0,149
Phonation * Kontur (Reim)	2,7	1,35	0,276

Die in 5.2.1.1 formulierten fünf Hypothesen werden anhand deskriptiver und prüfstatischer Auswertungen der Hörerurteile im Folgenden überprüft. Wie bereits in den anderen Experimenten dieser Untersuchung stützt sich die Prüfung der Hypothesen vor allem auf direkte Vergleiche der Mittelwerte über die 14 Versuchspersonen für die unterschiedlichen Stimuli und Stimulusserien. Die Werte auf den y-Achsen der Grafiken sind folgendermaßen zu interpretieren: Ein Wert von 6 bedeutet den geringsten Grad von Zögerlichkeit, ein Wert von 0 maximale Zögerlichkeit.

#### **Hypothese 1:**

Längung des phrasenfinalen Reims führt zur Wahrnehmung von unflüssiger Sprechweise, wenn sie einen Schwellenwert überschreitet. Wird die Längung oberhalb des Schwellenwertes noch vergrößert, verstärkt sich die wahrgenommene Unflüssigkeit weiter.



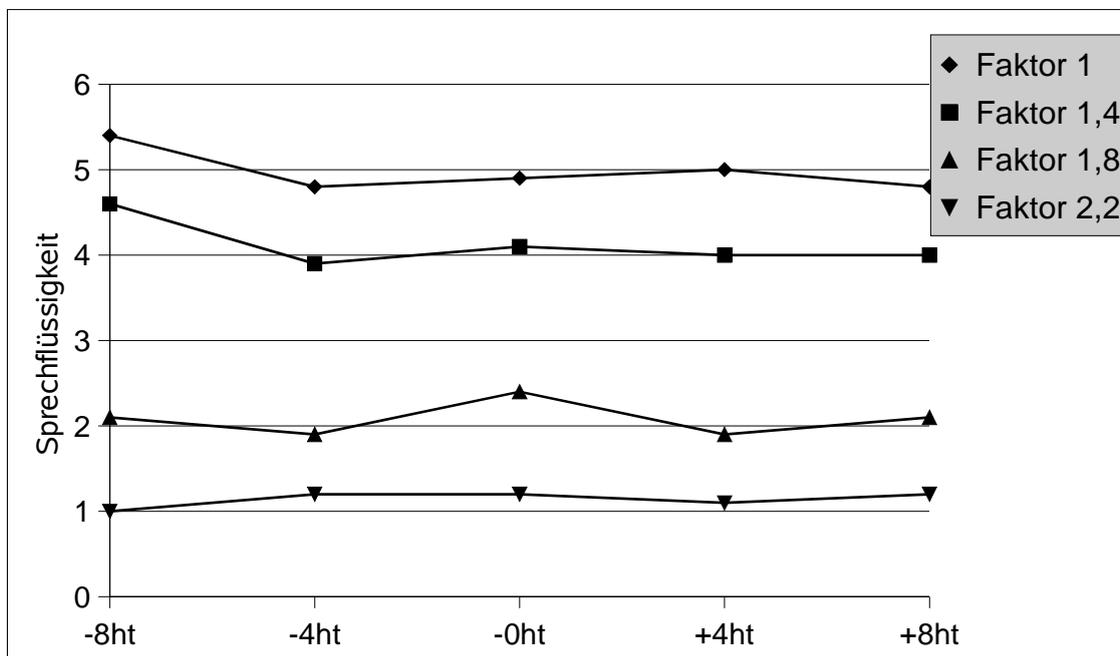
**Abbildung 14:** Wahrgenommene Sprechflüssigkeit für die Stimulusserie mit vier unterschiedlich stark gelängten Reimen und fünf melodischen Mustern am Phrasenende.

Die Dauer des finalen Reims ist eindeutig an der Wahrnehmung unflüssiger prosodischer Grenzen beteiligt (hochsignifikant,  $F=100,41$ ,  $p=0,000$ ; siehe Tabelle 34). Abbildung 14 zeigt, dass die Stimuli mit der Längungsstufe 1 durchgehend als flüssig beurteilt werden. Ein Erreichen des Maximalwerts von 6 ist in Experimenten mit Antwortskalen nicht zu erwarten, da einige Versuchspersonen sich anscheinend scheuen die äußersten Ränder der Skala in ihre Antworten einzubeziehen. So bleiben die Mittelwerte solcher Experimente fast immer unter den Maximalwerten. Die Serie der Längungsstufe 1,4 liegt leicht unterhalb der ungelängten Serie, ist aber noch immer als durchgehend flüssig einzustufen. Der Längungsfaktor von 1,4 entspricht der im Korpus gemessenen durchschnittlichen Längung an turninternen Grenzen ohne syntaktische Abbrüche, Häsitationspartikeln oder Zögerungslängung. Eine weitere Vergrößerung der Reimdauer um 0,4 auf Faktor 1,8 bringt einen deutlichen Sprung der Urteile in Richtung Zögerlichkeit. Die Urteile sind im Mittel etwas oberhalb der Skalenmitte angesiedelt, bewegen sich also klar von der Wahrnehmung als 'nicht-zögerlich' weg. Die noch stärkere Dehnung des finalen Reims (Faktor 2,2) bringt erneut eine deutliche Zunahme der Urteile in Richtung der zögerlichen Sprechweise. Hypothese 1 wird angenommen.

Die Ergebnisse sprechen nicht eindeutig für einen Schwellenwert bei der Wahrnehmung von Unflüssigkeit. Sie belegen eher eine kontinuierliche Zunahme der wahrgenommenen Unflüssigkeit mit zunehmender Reimdauer. Allerdings liegen die Urteile für die Serie der Längungsfaktoren 1 und 1,4 sehr dicht zusammen, während die weiteren hierzu linearen Vergrößerungen der Reimdauer auf die Faktoren 1,8 und 2,2 deutlich stärkere Änderungen der Urteile verursachen. Dieses Urteilsverhalten gibt einen Hinweis auf einen Schwellenwert bei der Wahrnehmung von Unflüssigkeit. Dies wird durch die Berechnung von Einzelkontrasten deutlich: Der Vergleich zwischen Stimuli mit Längungen der Faktoren 1 und 1,4 führt zu einem F-Wert von 12,6 und ist damit hochsignifikant ( $p=0,004$ ). Der Vergleich zwischen Längungsfaktor 1,4 und 1,8 ist ebenfalls hochsignifikant, aber mit einem wesentlich höheren F-Wert von 68 ( $p=0,000$ ). Hypothese 1 wird angenommen.

### Hypothese 2:

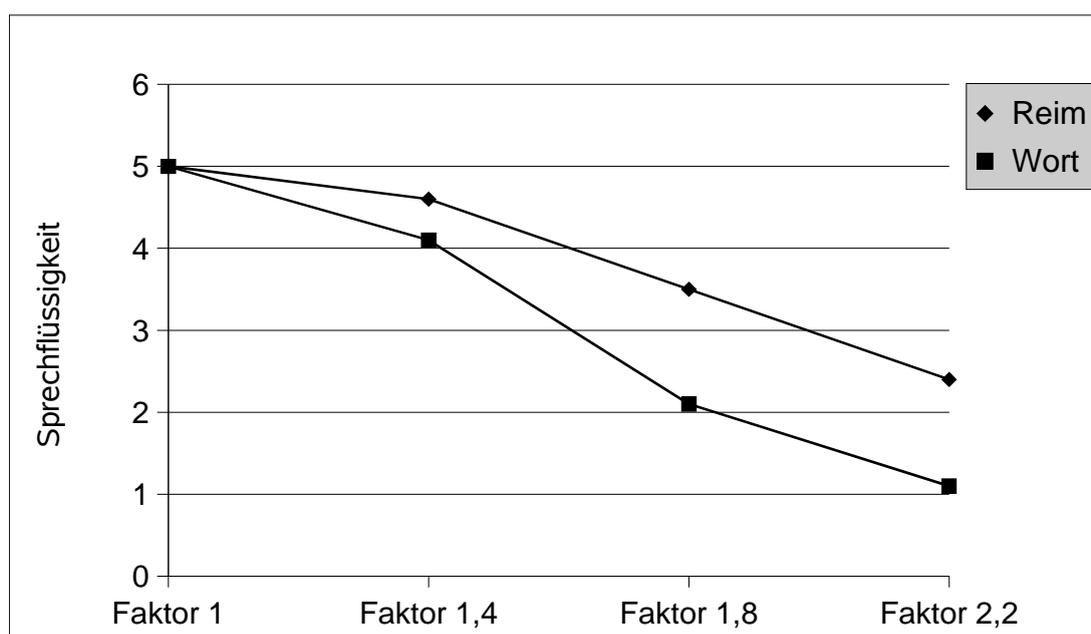
Dehnt sich die phrasenfinale Längung über einen größeren Bereich als den phrasenfinalen Reim aus, führt dies bei einem vergleichbaren Längungsfaktor stärker zur Wahrnehmung von unflüssiger Sprechweise.



**Abbildung 15:** Wahrgenommene Sprechflüssigkeit für die Stimulusserie mit vier unterschiedlich starken Längungen des letzten Wortes des Stimulus (drei Silben) und fünf melodischen Mustern am Phrasenende.

Im Vergleich zwischen Abbildung 14 und Abbildung 15 zeigt sich, dass die Wahrnehmung von Zögerlichkeit ansteigt, wenn sich die Längung am Ende der Äußerung statt über den finalen Reim über das gesamte letzte Wort ausdehnt.

Abbildung 16 stellt diesen Unterschied heraus, indem, gemittelt über alle melodischen Muster, die Serien mit Längung über dem finalen Reim den Serien mit Längung über dem gesamten finalen Wort gegenübergestellt werden. Die Unterschiede sind hochsignifikant ( $F=22,5$ ,  $p=0,000$ ).

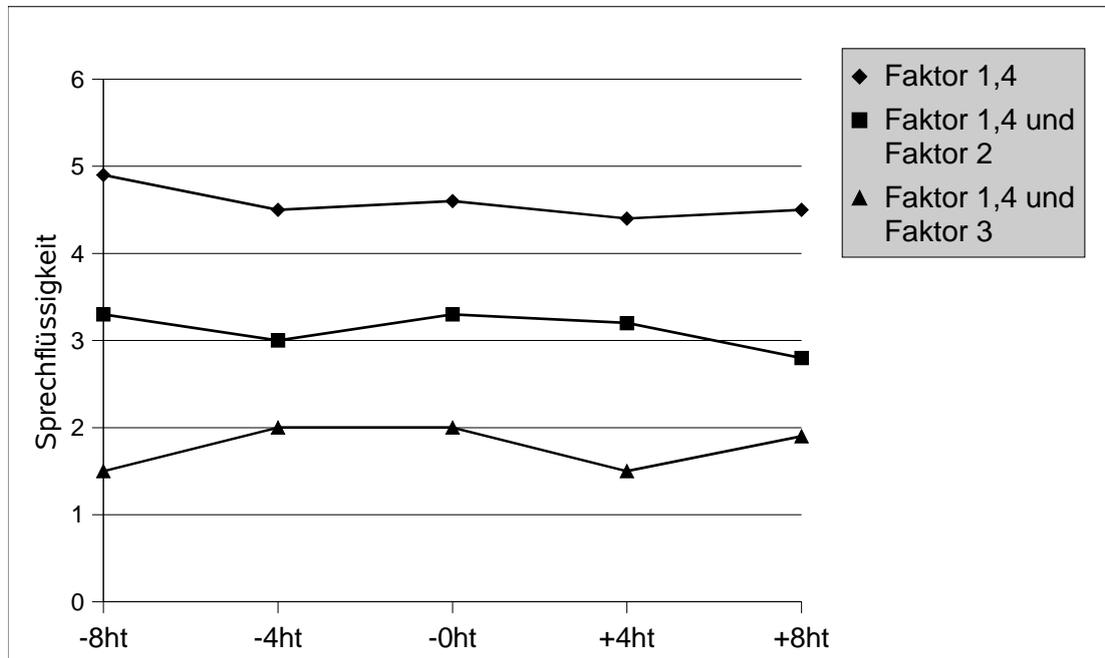


**Abbildung 16:** Wahrgenommene Sprechflüssigkeit im Vergleich zwischen der Stimulusserie mit gelängtem finalen Reim gegenüber der Serie mit Längung des letzten Wortes. Gemittelt über alle melodischen Muster.

Diese Ergebnisse lassen die Schlussfolgerung zu, dass für flüssige prosodische Grenzen die Domäne der finalen Längung auf den letzten Reim vor der prosodischen Grenze beschränkt ist und sich der Grad der finalen Längung nicht über eine etwa eineinhalbfache Dauervergrößerung ausprägt. Wenn von diesen Werten deutlich abgewichen wird, kommt es zur Wahrnehmung einer zögerlichen Sprechweise. Hypothese 2 wird angenommen.

### Hypothese 3:

Werden in einer flüssig gesprochenen prosodischen Phrase einzelne Segmente im Phraseninneren gelängt, führt dies zur Wahrnehmung von unflüssiger Sprechweise.



**Abbildung 17:** Wahrgenommene Sprechflüssigkeit im Vergleich zwischen einer Stimulusserie mit leichter Längung (Faktor 1,4) des finalen Reims und zwei Serien mit zusätzlicher Längung des Nasals in *nicht* um Faktor 2 bzw. Faktor 3.

Abbildung 17 zeigt, dass bei gleicher finaler Längung die Dauervergrößerung einzelner Segmente im Äußerungsinneren zu einer ansteigenden Wahrnehmung als unflüssig führt (hochsignifikant,  $F=86$ ,  $p=0,000$ ; siehe Tabelle 34). Die Signalisierung unflüssiger Sprechweise ist also keineswegs auf die Positionen prosodischer Grenzen beschränkt. Hypothese 3 wird angenommen.

### Hypothese 4:

Starke melodische Bewegungen dämpfen den Eindruck von Unflüssigkeit ab. D.h., dass bei ansonsten gleicher Merkmalskombination die ebenen, leicht fallenden und leicht steigenden Muster eher mit Unflüssigkeit assoziiert werden als die tief fallenden oder hoch steigenden.

Die tief fallenden Konturen werden bei gleichen Dauerverhältnissen als flüssiger beurteilt. Für die stärkste Längungsstufe ist der Effekt nicht mehr zu beobachten. Die hoch steigenden Konturen zeigen keinen Einfluss auf die Urteile. Generell zeigt sich, dass der Einfluss des Parameters 'Kontur' auf die wahrgenommene Sprechflüssigkeit gering ist (siehe Abbildung 13 und 15). Der dämpfenden Einfluss der tief fallenden Kontur ist allerdings hochsignifikant stärker als der der leicht fallenden Kontur (und so auch allen anderen Konturen). Der Einzelkontrast ergibt hier Werte von  $F=21,1$  und  $p=0,000$ . Dieser dämpfende Effekt der tief fallenden Kontur ist somit dafür verantwortlich, dass der Einfluss der Variable 'Kontur' insgesamt signifikant wird ( $F=4,9$ ,  $p=0,022$ ; siehe Tabelle 34). Hypothese 4 wird angenommen.

#### **Hypothese 5:**

Glottalverschlüsse am Ende der Äußerung verstärken die Wahrnehmung unflüssiger Sprechweise.

Die Hypothese wird verworfen, da sich kein signifikanter Einfluss der Variable 'Glottalverschluss' auf die Urteile zeigt (siehe Tabelle 34). Auch dieses Ergebnis belegt, wie schon das Experiment 2 die geringe perzeptorische Relevanz von Glottalverschlüssen an prosodischen Grenzen (siehe 4.3).

#### **5.2.1.5 Zusammenfassung und Diskussion**

Die Versuchspersonen reagieren fein auf Veränderungen in der Dauerstruktur und bringen perzeptorisch auffällige segmentelle Längung mit Störungen der Sprechflüssigkeit in Verbindung. Ab Längungen von Faktor 1,8 des finalen Reims tritt eine deutliche Beurteilung der Stimuli als unflüssig auf. Wird die Längung über einen größeren Bereich als den finalen Reim ausgedehnt (hier die letzten drei Silben) verstärkt sich die wahrgenommene Unflüssigkeit ab Längungsfaktor 1,8 gegenüber den Stimuli, in denen nur der Reim gelängt wurde. Eine leichte Längung im Bereich von Faktor 1,4 führt zu einer schwachen Anhebung der Unflüssigkeitsurteile gegenüber den ungelängten Varianten. Die Urteile belegen für die leichte finale Längung aber insgesamt die Wahrnehmung einer flüssigen Sprechweise. Die Experimente legen nahe, dass als Domäne der flüssigen

finalen Längung der finale Reim vor der prosodischen Grenze anzusehen ist. Eine Verlagerung der gelängten Segmente weg vom finalen Reim deuten auf eine unflüssige Sprechweise hin. Diese Schlussfolgerung bezieht allerdings nicht die Längungserscheinungen ein, die für satzakzentuierte Silben beobachtbar sind. Hierzu müssten separate Versuche durchgeführt werden.

Die melodischen Muster am Phrasenende spielen in der Einordnung als 'zögerlich' vs. 'nicht-zögerlich' kaum eine Rolle. Lediglich das tief fallende Muster führt in einigen Fällen zur Beurteilung als weniger zögerlich als temporal identische Stimuli mit anderen finalen Melodiemustern.

## **5.2.2 Experiment zur Wahrnehmung unflüssiger Sprechweise an unterschiedlichen syntaktischen Strukturstellen (Experiment 3b)**

In den Produktionsdaten kommen unflüssige prosodische Grenzen oft vor, ohne dass sie mit hohen syntaktischen Grenzen zusammenfallen. Die folgende experimentelle Untersuchung soll zeigen, inwieweit die Position einer prosodischen Grenze in Relation zur syntaktischen Struktur die Wahrnehmung als flüssige bzw. unflüssige Grenze beeinflusst. Hierfür werden Stimulusserien generiert, die prosodische Merkmalsbündel verschiedener Art an unterschiedlichen syntaktischen Strukturstellen enthalten. Die Versuchspersonen beurteilen dann, wie im vorangegangenen Experiment, den Grad der wahrgenommenen Unflüssigkeit (siehe 5.2.1.3).

### **5.2.2.1 Hypothesen**

Folgende Hypothese wird geprüft:

Gleiche Bündelungen phonetischer Merkmale erzeugen an niedrigen syntaktischen Grenzen stärker den Eindruck von Unflüssigkeit als an hohen syntaktischen Grenzen.

### 5.2.2.2 Stimulusmaterial und Durchführung

Das verwendete Stimulusmaterial ist identisch mit dem Material aus Experiment 1c (siehe 3.3.2.4.2o). Die Durchführung ist identisch mit Experiment 3a zur wahrgenommenen Zögerlichkeit (siehe 5.2.1.3). An dem Experiment nehmen 17 Versuchspersonen teil.



Audiobeispiele zu den Stimuli aus Experiment 3b: 4.2., 4,5

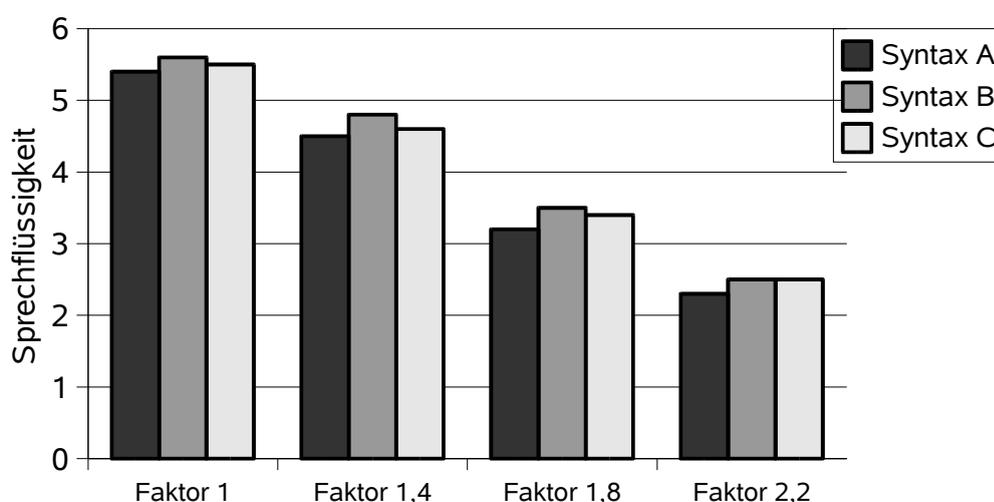
### 5.2.2.3 Ergebnisse

Tabelle 35 fasst die Ergebnisse der Varianzanalyse zusammen. In einigen Fällen sind in Abhängigkeit von der jeweiligen Fragestellung weitere Einzelkontraste zwischen bestimmten Stimuli berechnet worden. Diese Ergebnisse sind in den nachfolgenden Text integriert.

**Tabelle 35:** Ergebnisse der Varianzanalyse von Experiment 3b.

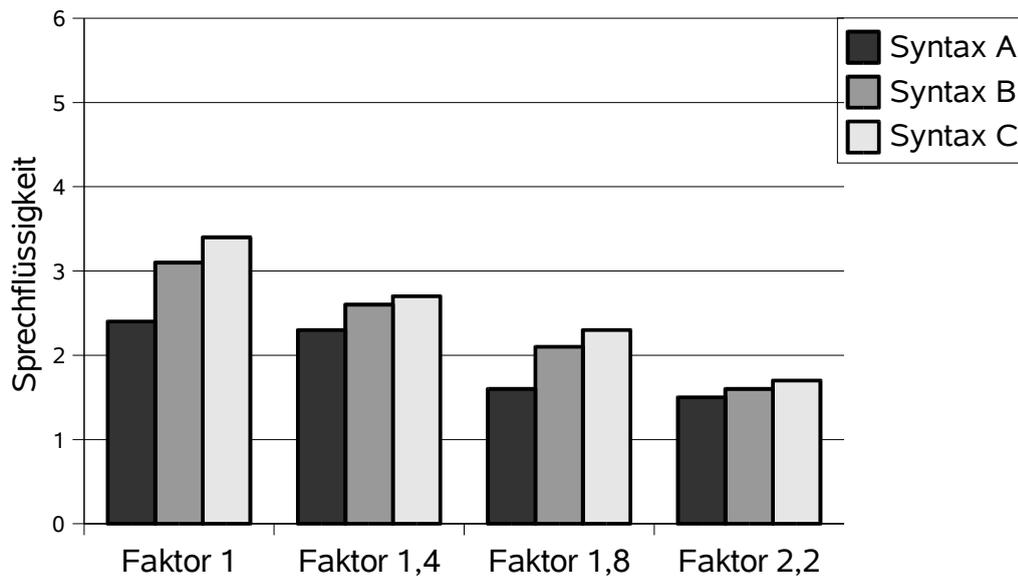
<b>Faktor / Interaktion</b>	<b>df</b>	<b>F</b>	<b>p</b>
Syntax	1,76	5,2	0,015
Längung des finalen Reims	1,59	73,62	0,000
Finale Kontur	1,88	5,02	0,015
Pause	1	26,95	0,000
Syntax * Längung	4,14	0,91	0,468
Syntax * Kontur	3,68	0,69	0,594
Syntax * Pause	1,7	2,61	0,1
Längung * Kontur	3,56	2,74	0,043
Längung * Pause	2,22	24,78	0,000
Kontur * Pause	2	0,2	0,822
Syntax * Längung * Pause	4,46	0,59	0,69
Syntax * Längung * Kontur	7,43	1	0,435
Syntax * Pause * Kontur	2,78	0,42	0,724
Längung * Pause * Kontur	4,81	1,84	0,117
Syntax * Pause * Kontur * Längung	6,25	0,93	0,481

Wie bereits im vorangegangenen Experiment, hat auch in diesem Experiment der melodische Verlauf an den prosodischen Grenzen nur schwache Auswirkungen auf die wahrgenommene Sprechflüssigkeit, obwohl er statistisch auch hier signifikant wird. Aus diesem Grund wird hier nicht weiter auf die melodischen Verläufe eingegangen. In den Abbildungen 18 und 19 sind Mittelwerte über alle melodischen Muster gebildet worden, um zu einer breiteren Datenbasis für die Aussagen über die Auswirkung von Pausen und finaler Längung auf die Urteile der Versuchspersonen zu kommen.



**Abbildung 18:** Wahrgenommene Sprechflüssigkeit für vier Längungsbedingungen unter drei syntaktischen Bedingungen. Stimuli ohne Pausen. Die melodischen Muster sind zusammengefasst.

Abbildung 18 zeigt eine deutliche Auswirkung der Längung des finalen Reims auf die wahrgenommene Sprechflüssigkeit (hochsignifikant,  $F=73,62$ ,  $p=0,000$ ; siehe Tabelle 35). Die Ergebnisse sind denen aus Experiment 3a ähnlich. Die Auswirkung der syntaktischen Strukturstelle auf die Urteile sind relativ gering, aber trotzdem signifikant ( $F=5,2$ ,  $p=0,015$ ; siehe Tabelle 35). Es zeigt sich eine Wahrnehmung stärkerer Unflüssigkeit unter der ersten syntaktischen Bedingung: Wenn also die prosodische Grenze innerhalb einer syntaktischen Einheit vorkommt steigen bei gleichen Merkmalen die Urteile für eine zögerliche Sprechweise leicht an. Die Berechnung von Einzelkontrasten zeigt hochsignifikante Unterschiede zwischen Syntax A und B ( $F=11,7$ ,  $p=0,003$ ) und Syntax A und C ( $F=9,5$ ,  $p=0,009$ ). Syntax B vs. C ergibt kein signifikantes Ergebnis ( $F=0,008$ ,  $p=0,928$ ), (siehe Tabelle 35).



**Abbildung 19:** Wahrgenommene Sprechflüssigkeit für vier Längungsbedingungen unter drei syntaktischen Bedingungen. Stimuli mit Pausen. Die melodischen Muster sind zusammengefasst.

Abbildung 19 zeigt im Vergleich zu Abbildung 18, dass Pausen den Eindruck einer unflüssigen Sprechweise deutlich verstärken. Bereits für die Längungsfaktoren 1 und 1,4 tendieren die Urteile stark in Richtung unflüssig, wenn eine Pause interveniert. Für die Serien mit Pausen lässt sich auch der vorhergesagte Einfluss der syntaktischen Struktur auf die Urteile durchgehend beobachten. Je schwächer der syntaktische Einschnitt an der prosodischen Grenze ist, desto stärker unflüssig erscheint bei gleicher Merkmalsbündelung die Sprechweise. Die geprüfte Hypothese wird folglich angenommen.

#### 5.2.2.4 Zusammenfassung und Diskussion

Es zeigt sich, dass prosodische Grenzen im Inneren syntaktischer Einheiten den Eindruck unflüssiger Sprechweise verstärken. Die Versuchspersonen erwarten also im Rahmen einer flüssigen Sprechweise möglicherweise die Konvergenz syntaktischer und prosodischer Phrasierung. Tritt diese Konvergenz nicht ein, wird die Äußerung als unflüssig eingestuft.

Die starke Auswirkung von Pausen auf die Urteile ist überraschend, vor allem, weil Pausen ein häufiges Korrelat prosodischer Grenzen auch in völlig flüssiger Sprache sind. Eine mögliche Erklärung für den starken Effekt der Pausen ist die kurze Äußerungsdauer, die das Vorkommen einer inhaltlich strukturierenden internen Pause eher unwahrscheinlich macht und die Interpretation als Häsitationspause nahe legt.

### **5.3 Zusammenfassung und Diskussion zu den unflüssigen prosodischen Grenzen**

Die Faktoren, die die Wahrnehmung von unflüssiger Sprechweise verursachen, sind vielfältig und verwoben. Die Perzeptionsuntersuchungen zeigen, dass vor allem starke segmentelle Längung den Eindruck unflüssiger Sprechweise verursacht, besonders wenn diese innerhalb syntaktischer Einheiten auftritt. Die Analyse der Produktionsdaten belegt zudem einen hohen Anteil von prosodischen Grenzen, an denen Häsitationspartikeln vorkommen. Auch hier kann davon ausgegangen werden, dass dies zur Wahrnehmung von Unflüssigkeit führt. Weiterhin sind einige Fälle im Korpus auffindbar, die Knarrstimme und Glottalverschlüsse in Verbindung mit syntaktischen Abbrüchen aufweisen. Eine perzeptorische Relevanz dieses Phänomens kann allerdings nicht nachgewiesen werden.

An den untersuchten unflüssigen Grenzen liegt der Anteil leicht fallender und ebener F<sub>0</sub>-Muster insgesamt bei etwa 80%. Starke melodische Bewegungen an unflüssigen Grenzen sind relativ selten. Sie treten in Verbindung mit finaler Zögerungslängung in etwa 10% der Fälle auf und in Verbindung mit Häsitationspartikeln in etwa 20% der Fälle. Hieraus ergibt sich die Vermutung, dass der erhöhte Anteil tief fallender Konturen vor Häsitationspartikeln damit zusammenhängt, dass vor der prosodischen Grenze mit fallender Kontur ein inhaltlicher Abschluss erreicht wurde und die Häsitationspartikel signalisiert, dass trotz dieses Abschlusses der Turn gehalten werden soll, aber noch Planungsprobleme hinsichtlich der Struktur des Folgenden bestehen. Insofern sollte eine generelle Trennung angestrebt werden, die zwischen Planungsproblemen unterscheidet, die sich auf eine neue inhaltliche Einheit beziehen, und solchen die innerhalb von inhaltlichen und/oder syntaktischen Einheiten auftreten.

Durch die gefundenen phonetischen Merkmale sind die untersuchten unflüssigen prosodischen Grenzen relativ klar von flüssigen Grenzen getrennt. Allerdings gehen in die korpusbasierte Analyse unflüssiger Grenzen von vornherein nur prosodische Grenzen ein, die entweder Zögerungslängung, Häsitationspartikeln oder syntaktische Abbrüche zeigen. Denn es gibt keine anderen Möglichkeiten, labelbasiert auf unflüssige Grenzen zuzugreifen, als über diese Parameter. Es ist aus diesem Grund wahrscheinlich, dass es viele andere prosodische Grenzen im Korpus gibt, die durch Verzögerungen im Sprachplanungsprozess entstanden sind, die aber keine der oben genannten Etikettier-Merkmale tragen. Hierfür spricht auch, dass in den Terminabsprachedialogen etwa ein Drittel aller etikettierten Pausen an Positionen vorkommt, an denen kein Interpunktionszeichen vorliegt. Dies deutet darauf hin, dass viele der Pausen nicht mit hohen syntaktischen Grenzen zusammenfallen. Das letzte Perzeptionsexperiment (Experiment 3b) zeigt, dass Versuchspersonen das Vorkommen prosodischer Grenzen ohne hohe syntaktische Grenzen als Anzeichen von Unflüssigkeit bewerten, also ein Zusammenfallen von prosodischer und syntaktischer Phrasierung erwarten. So liegt nahe, dass es, zumindest bei einigen Sprechern, eine Strategie gibt, Planungsschwierigkeiten nicht durch Zögerungslängung und/oder Häsitationspartikeln zu signalisieren, sondern die entstehenden prosodischen Einschnitte in ihrer phonetischen Form ähnlich den flüssigen Grenzen zu gestalten. Etikettierte Grenzen dieser Art sind aufgrund des Analyseverfahrens zusammen mit den inhaltlich motivierten, gliedernden flüssigen Grenzen in Kapitel 3 untersucht worden. Hier eine Trennung zwischen inhaltlich motivierten und durch Planungsprozesse entstandenen Grenzen herbeizuführen, kann lediglich durch inhaltliche Interpretationen und/oder syntaktische Analysen erreicht werden. Für eine funktionale Zuordnung phonetisch verschiedener Typen unflüssiger Grenzen fehlen bislang ebenfalls klare Hinweise. Auch hierfür können weitere Form-Funktions-Zusammenhänge m.E. am besten durch Interpretationen im Gesamtkontext einer Äußerung herausgearbeitet werden. Es ist also auf dem derzeitigen Wissensstand nicht möglich systematische Form-Funktions-Beziehungen unterschiedlicher unflüssiger prosodischer Grenzen aufzustellen.



Audiobeispiele: Folgende Beispiele illustrieren die schwierige Trennung flüssiger und unflüssiger prosodischer Grenzen durch phonetische Parameter: 3.1, 3.4

## 6 Abschließende Diskussion und Ausblick

Durch die Kombination von automatischen Merkmalsabfragen, Perzeptionsexperimenten und interpretativen Analysen wurden Beziehungen der phonetischen Struktur prosodischer Grenzen zu inhaltlichen Strukturen sowie zu interaktionalen Funktionen hergestellt. Inhaltliche Separierung im Inneren von Redebeiträgen wird vor allem durch tief fallende melodische Bewegungen und Pausen signalisiert, verschiedene Funktionen der Dialogsteuerung werden hauptsächlich durch unterschiedliche melodische Ausprägung an prosodischen Grenzen repräsentiert, und unflüssige Sprechweise ist häufig durch segmentelle Längungserscheinungen in der segmentellen Dauerstruktur kodiert. Allerdings können die prosodischen Grenzen der drei funktionalen Bereiche weder eindeutig mit bestimmten phonetischen Exponenten assoziiert werden, noch lassen sich klar abgrenzbare funktionale Unterklassen innerhalb der flüssigen turninternen, der turnfinalen und der unflüssigen prosodischen Grenzen bilden. Die vorgenommenen Systematisierungen bilden lediglich funktionale Kernbereiche, deren Repräsentation durch typische phonetische Merkmalskombinationen in Perzeptionsexperimenten und durch Korpusanalysen belegt wird. Dass keine 1:1-Zuordnung zwischen kommunikativer Funktion und phonetischer Struktur gefunden wurde, liegt vor allem an der hohen funktionalen und phonetischen Variabilität spontaner Sprache. Aus dieser Situation heraus sollen zum Abschluss der Arbeit zwei Fragen diskutiert werden, die zum einen auf eine kritische Reflexion der geleisteten Untersuchungen zielen und deren Diskussion zum anderen Anregungen für zukünftige Untersuchungen in der funktionsorientierten Phonetik geben soll.

- (1) Was konnten die angewendeten Untersuchungsmethoden leisten?
- (2) Welche grundlegenden Probleme gibt es bei der Analyse von Form-Funktions-Beziehungen in spontaner Sprache?

- (1) Was konnten die angewendeten Untersuchungsmethoden leisten?

Die Kombination von labelbasierten und interpretativen Analysen von Produktionsdaten mit der experimentellen Erhebung von Perzeptionsdaten ist sinnvoll, da sich hierdurch sowohl von der Sprecherseite (Produktion) als auch von der Hörerseite (Perzeption) ein

Zugang zum Untersuchungsbereich bietet. Die labelbasierten Korpusabfragen zeigen, welche phonetischen Strukturen von Sprechern produziert werden, die Perzeptionsexperimente geben Aufschluss über Wahrnehmung und funktionale Relevanz dieser Strukturen beim Hörer, und die interpretativen Analysen und Fallbeschreibungen liefern zusätzliches Detailwissen über phonetische und funktionale Parameter und stellen den Bezug zum Gesprächskontext her. Die drei Untersuchungsmethoden ergänzen sich also gegenseitig und liefern ein komplexes Bild des untersuchten Phänomenbereichs. Folgende Vor- und Nachteile der drei methodischen Stränge zeigten sich im Laufe der Analysen.

Die labelbasierten Korpusabfragen bieten einen statistischen Überblick über die phonetische Ausprägung etikettierter prosodischer Grenzen und somit ein Bild der phonetischen Substanz des Korpus. Hierbei ist es durch Abfragen der Etikettierdateien zwar möglich, eine große Zahl von Phrasengrenzen phonetisch zu beschreiben, diese Beschreibungen sind jedoch eng an die Kategorien der vorhandenen perzeptiv orientierten Etikettierung gebunden, und nicht alle relevanten Parameter können labelbasiert abgefragt werden. So fehlen in der Etikettierung z.B. Informationen über Intensitätsverläufe, präzise Angaben zur Stimmqualität und eine durchgehende Differenzierung von Ein- vs. Ausatmen. Insofern sind die abgeleiteten Merkmalsbündel an den prosodischen Grenzen nicht vollständig. Eine weitere enge Bindung wird im Bereich der prosodischen Etikettierung offensichtlich. Es können nur die Kategorien gefunden und funktional zugeordnet werden, die bereits in die Etikettierung eingegangen sind. Was sich im Detail hinter den melodischen Verläufen verbirgt, die z.B. als 'tief fallend' etikettiert wurden, ist aus der Etikettierung nicht erkennbar. Die vorgenommene Analyse der finalen Längung zeigt aber einen Weg, die rein zählende Untersuchung der Etiketten zu erweitern und durch die Weiterverarbeitung von etikettierten Segmentdauern an neue, nur implizit vorhandene Information über *Dauerhältnisse* zu gelangen. Eine ähnliche Erweiterung der Information könnte durch geeignete Programme auch auf dem Gebiet der melodischen Muster vorgenommen werden. So könnte z.B. automatisch auf sämtliche Fälle der etikettierten Kategorie 'tief fallend' über die Etikettierung zugegriffen werden und der Bereich, der als 'tief fallend' klassifiziert wurde, durch einen F0-Algorithmus analysiert werden. Die erhaltenen Messwerte könnten dann genutzt werden, um Angaben über absolute Endpunkte der Konturen, die Dauern der fallenden Bewegung und somit auch Gradienten von melodischen Bewegung zu berechnen (z.B. in Halbtöne pro Sekunde).

Über diese neue Information könnte dann z.B. der Frage nachgegangen werden, ob inhaltliche Terminalität vor allem durch Umfang oder Endpunkt einer melodischen Bewegung kodiert wird. Vergleichbare Analysen sind auch für Intensitätsverläufe möglich. Insofern besitzt die Etikettierungsebene durch ihre kategorisierende Funktion auch die Eigenschaft, Zugriff zu einer großen Anzahl von (vor-)kategorisierten Fällen zu bieten, so dass durch automatische oder manuelle Weiterverarbeitung neue Information über die Fälle innerhalb der Kategorie gewonnen werden kann. Hier öffnet sich ein weites Untersuchungsfeld, in dem die Etikettierung lediglich das heuristische Instrument darstellt, um zu exakten akustischen Analysen zu gelangen. Hierbei ist allerdings zu bedenken, dass sich in einer so umfangreichen und über die Jahre gewachsenen Datenbasis auch immer Fehler und Inkonsistenzen finden.

Trotz der großen Leistungsfähigkeit automatischer Korpusanalysen kann mittels Abfrage der Labeldateien nur in sehr eingeschränkter Weise eine Verbindung zur sprachlichen Funktion der gefundenen phonetischen Muster abgeleitet werden, da in der untersuchten Datenbasis keine Etikettierung semantisch-pragmatischer Strukturen vorhanden ist. Diese Einschränkung führt weiter zu den experimentellen Untersuchungen und deren Vor- und Nachteilen.

Perzeptionsexperimente ergänzen die Korpusabfragen um die funktionale Zuordnung, indem über Hörerurteile die funktionale Relevanz von phonetischen Strukturen überprüft wird. Hierbei werden die phonetischen Beobachtungen der vorangegangenen Korpusabfrage systematisch in Stimulusserien umgesetzt und hypothesengesteuert funktionale Zuordnungen überprüft. Das experimentelle Vorgehen hat somit den Vorteil, dass kontrolliert einzelne formale und funktionale Parameter aufeinander bezogen werden können, indem phonetische und funktionale Merkmalskomplexe im Experiment dekomponiert werden.

Die Ergebnisse der Experimente zeigen, dass die Phonetik der prosodischen Grenzen funktional relevant ist, und zwar sowohl in der Dialogsteuerung als auch in der Signalisierung inhaltlicher Strukturen und kognitiver Prozesse. Da in den Experimenten durchgehend mit Antwortskalen gearbeitet wurde, zeigte sich aber auch, dass es bei keinem der untersuchten Parameter durch dessen Variation zu einem abrupten Kategorienwechsel innerhalb der funktionalen Urteile kam, der eine Postulierung von Schwellenwerten oder scharfen Kategoriengrenzen gerechtfertigt hätte. In den meisten

Fällen haben kontinuierliche phonetische Parameteränderungen auch kontinuierliche Änderungen im Urteilsverhalten nach sich gezogen. In diesem Zusammenhang soll noch einmal kurz auf die Rolle der Prüfstatistik eingegangen werden, die für viele phonetische Kontraste ein hochsignifikant unterschiedliches Urteilsverhalten nachweisen konnte, obwohl die Urteile sich z.T. nur wenig auf der Skala verschoben haben und diese Verschiebung im gesamten Kontext des Experiments nicht als Kategorienwechsel interpretiert werden konnte. Die Arbeit mit Skalen ist also gut geeignet, um kontinuierliche funktionale Übergänge abzubilden, die mit kontinuierlichen Parameteränderungen parallel laufen, und prüfstatische Signifikanzen sollten in Experimenten dieser Art immer kritisch auf ihre erklärende und kategorienbildende Kraft überprüft werden.

Ein wichtiger Punkt in der Konzeption von Experimenten und der Interpretation der experimentellen Ergebnisse ist, dass ein Experiment nie eine natürliche Gesprächssituation abbilden kann. Auch die Experimente dieser Arbeit sind weitgehend kontextfreie Stilisierungen eines wirklichen Kommunikationsprozesses und dürfen nicht übergeneralisiert werden. Besonders wenn es um die experimentelle Überprüfung interaktionaler Verhaltensmuster geht, wird ein hohes Maß an Abstraktion von den Versuchspersonen verlangt, da sie Auskunft über Wahrnehmungsprozesse geben sollen, die in natürlicher Kommunikation unbewusst ablaufen. Dies ist sicher einer der Gründe, warum sich das Verhalten verschiedener Versuchspersonen in den Experimenten z.T. stark unterscheidet und hierdurch z.T. große Standardabweichungen entstehen.

Einen ganz anderen Zugang zu den Daten bieten die vorgenommenen Interpretationen. In der Arbeit wird auf zweierlei Weise interpretierend vorgegangen, zum einen bei der Beurteilung der Stärke inhaltlicher Einschnitte in Kapitel 3 und zum anderen bei der Darstellung der Korpusbeispiele auf der beiliegenden CD-ROM. Die Interpretationen der Korpusbeispiele auf der CD folgen dem Interpretationsschema der Gesprächsanalyse. Die phonetischen Strukturen werden im jeweiligen Gesprächskontext in Verbindung mit sprachlichen Funktionen gebracht. Hierbei werden phonetische Details in die Interpretation einbezogen und die funktionale Struktur, geleitet durch Intuition, aus dem interaktiven, sequentiellen Sprecher- und Hörerverhalten abgeleitet. Diese Art der Interpretation zeigt die feine Abstimmung zwischen den Sprechern ebenso wie den komplexen Einsatz phonetischer Signale und ist vor allem geeignet, einen tieferen

Einblick in die real in einem Gespräch ablaufende Interaktion zwischen den Sprechern zu geben. Das "Teilnehmen" an einem Dialog, das in einer interpretierenden Bearbeitung stattfindet, ist zudem ein wichtiger Blickwinkel, um den Realitätsgehalt experimenteller Untersuchungen ständig zu prüfen und neue Annahmen und Hypothesen zu bilden. Eine statistische Absicherung und weitgehende Verallgemeinerung der gemachten Beobachtungen ist durch dieses Verfahren allerdings nicht zu erreichen.

Die interpretierende Klassifizierung inhaltlicher Einschnitte in Kapitel 3 kann hingegen das interpretative Vorgehen mit dem statistischen Ansatz der Arbeit verbinden, indem anhand definierter Kriterien eine große Zahl von Phrasengrenzen inhaltlich interpretiert wird und die automatisch abgeleiteten Merkmale dieser Phrasengrenzen auf die interpretativen Kategorien bezogen werden. Hierdurch kann belegt werden, dass inhaltliche und phonetische Separierung eng aneinander gekoppelt sind. Ein solches Vorgehen ist geeignet, um interpretative Kategorien klarer zu definieren, sich nicht so stark auf die Intuition des Bearbeiters zu verlassen und die interpretativen, funktionalen Kategorien systematisch mit phonetischen Größen zu assoziieren. Dieser Ansatz ist ausbaufähig und könnte ein Bindeglied zwischen dem Vorgehen der CA und signalphonetischen und experimentellen Forschungsansätzen bilden.

(2) Welche grundlegenden Probleme gibt es bei der Analyse von Form-Funktions-Beziehungen in spontaner Sprache?

Die Untersuchungen zu prosodischen Grenzen haben grundsätzliche Fragen zur Analyse von Form-Funktions-Beziehungen im Bereich der Prosodie aufgeworfen. Obwohl der Kern dieser Arbeit die Frage nach dem funktionalen Stellenwert der phonetischen Struktur prosodischer Grenzen ist, sind die folgenden Überlegungen m.E. relevant für funktionsorientierte phonetische Untersuchung spontaner Sprache im Allgemeinen.

Der Ausgangspunkt der folgenden Argumentation ist, dass in keinem der durchgeführten Arbeitsschritte scharfe Kategoriengrenzen herausarbeitet werden konnten, die Aussagen wie "Merkmalsbündel X hat im Dialog immer Funktion Y" rechtfertigen würden. Es zeigte sich vielmehr, dass es zwar eindeutige Tendenzen in der Beziehung zwischen der phonetischen Substanz und den untersuchten funktionalen Strukturen gibt, dass diese Beziehungen im spontanen Gespräch aber häufig durchbrochen werden. So stellt sich die Frage, warum es kein eindeutiges prosodisches Signalisierungssystem gibt,

das phonetische Form mit sprachlicher Bedeutung verbindet. Die bisher verfolgte Erklärung hierfür ist, dass es verschiedene Arten gibt, eine bestimmte Funktion (z.B. Halten vs. Übergeben des Rederechts) zu signalisieren. So hat jeder Sprecher einen individuellen Sprech- und Kommunikationsstil, einen bestimmten sprachlichen Hintergrund und bestimmte sprachliche Fähigkeiten. Und auch innerhalb der Systeme einzelner Sprecher gibt es für die Signalisierung einer bestimmten Funktion verschiedene Strategien, zwischen denen variiert werden kann. Diese Situation ist eine Quelle großer phonetischer Vielfalt. Um die phonetischen Varianten mit funktionalen Mustern zu assoziieren, ist es notwendig, sie im Gesamtkontext einer Gesprächssituation zu beurteilen und in Abhängigkeit vom Kontext funktionale Strukturen anzusetzen. Dieser Kontext umfasst auf der einen Seite typische linguistische Strukturen wie Morphologie, Syntax und Semantik, auf der anderen Seite aber auch situative, stilistische und emotionale Einflüsse. Dieses Prinzip der kumulativen Evidenz ist folglich ein adäquater Weg, Form und Funktion aufeinander abzubilden, denn alles deutet darauf hin, dass Sprecher und Hörer in Unterhaltungen alle verfügbaren Ressourcen nutzen, um kommunikative Funktionen zu kodieren bzw. zu entschlüsseln. Der Hörer rezipiert und interpretiert z.T. defizitäre Informationen aus völlig unterschiedlichen Ebenen und versucht im Sinne seiner Interpretation des Gehörten adäquat zu reagieren. Das erweiterte, kumulative Merkmalsbündel schafft hierbei Redundanzen in der Signalisierung sprachlicher Funktionen und verringert die Gefahr von Missverständnissen. Hierbei sollte die Rolle der Phonetik stets im Netzwerk der beteiligten Informationsebenen betrachtet und die funktionalen Anbindungen prosodischer Strukturen nicht überschätzt und in starre Systeme gepresst werden. Unter diesem Gesichtspunkt trennen die in dieser Arbeit vorgenommenen Systematisierungen unterschiedlicher prosodischer Grenzen zwar verschiedene phonetische Ausprägungen und ordnen sie funktionalen Strukturen zu. Die Kategorien sind aber als prototypische Kernbereiche zu verstehen, von denen oft abgewichen wird. Diese Abweichungen entstehen durch die starke phonetische Variabilität, die durch das Prinzip der kumulativen Signalisierung funktionaler Strukturen möglich wird, ohne dass hierdurch die inhaltliche und interaktionale Struktur eines Gesprächs regellos oder zufällig wird. Phonetische Parameter können also stark variieren, wenn die "Gesamtrendenz" in einem um syntaktische, semantische und situative Information erweiterten Merkmalsbündel stimmt.

Es kann aber auch eine andere Betrachtungsweise verfolgt werden, in der das Prinzip der Einheitlichkeit sprachlicher Funktionen kritisch betrachtet wird. Es muss in Zukunft verstärkt der Frage nachgegangen werden, was Funktion im Gespräch überhaupt ist und aus welcher Perspektive sie definiert werden kann. Auf der Seite der phonetischen Substanz gibt es messbare Fakten: F0-Verläufe, Intensität, Segmentdauern usw. können systematisch und objektiv erfasst werden. Sprachliche Funktionen und interaktional relevante Bedeutungen jedoch entstehen in der jeweiligen Gesprächssituation zwischen den Teilnehmern des Gesprächs, und es gibt keine Möglichkeit, während eines natürlichen Gesprächs objektiv die Intention des Sprechers und die Wahrnehmung des Hörers zu messen. Deshalb werden Funktionen postuliert und z.B. in stilisierenden Wahrnehmungsexperimenten überprüft, oder es wird im Rahmen der CA davon ausgegangen, dass das Verhalten des Hörers (z.B. wenn er das Rederecht übernimmt) als Indikator für die Intention des vorigen Sprechers angesehen werden kann, das Rederecht abgeben zu wollen. Oftmals scheint es aber so zu sein, dass das Prinzip des *Turnübergabens* gar nicht entscheidend ist, sondern ein erlaubendes *Turnübernehmen* stattfindet, das auf die Initiative des Hörers zurückgeht. Hier wird klar, dass das Konzept 'Funktion' von der Perspektive (Sprecher vs. Hörer) abhängt und eine Turnübernahme ein vielschichtiger kommunikativer Prozess ist, der in der Interaktion ausgehandelt wird. Auch andere funktionale Kategorien, die z.T. eindimensional dargestellt werden, sind bei genauerer Betrachtung komplex. Wenn z.B. ein Äußerungsabschnitt als 'Frage' klassifiziert wird, kann es sich um eine sachliche Informationsfrage handeln, um eine erstaunte Echofrage, um einen ironischen Kommentar und um vieles mehr. Dieses Beispiel zeigt, dass auch hier das Merkmal 'Aktivitätstyp = Frage' nur eine stark generalisierte Eigenschaft der Äußerung erfasst und eine ganze Reihe interaktional relevanter funktionaler Information ausblendet. Funktion entsteht also in einem interaktionalen Prozess und ist, genau wie ein phonetischer Merkmalskomplex, ein Bündel von Eigenschaften, allerdings kein Bündel physikalisch messbarer Parameter, sondern ein Komplex semantischer, pragmatischer und emotionaler Muster. So entsteht die Sichtweise, dass es weder auf der funktionalen, noch auf der phonetischen Seite Strukturen gibt, die durch einzelne Merkmale bzw. eine nicht weiter aufspaltbare Funktion erfasst werden können.

Die vorliegende Arbeit operiert aus dieser Sichtweise, indem die funktionale Relevanz eines phonetischen Merkmalskomplexes in funktionaler Hinsicht auf drei

Ebenen analysiert wird, die sich im Gespräch überlagern: die inhaltlich gesteuerte prosodische Phrasierung, die Ebene der Dialogsteuerung und die prosodischen Grenzen, die durch Planungsschwierigkeiten in die Rede einfließen. Eine Modellierung des Hörerverhaltens kann diese drei Ebenen voneinander trennen, indem angenommen wird, dass Hörer bei ihrer Interpretation prosodischer Grenzen drei separate Beurteilungen treffen, die jeweils skaliert vorgenommen werden. Erstens eine inhaltsbezogene Einordnung zwischen Abschluss und Weiterweisung. Hier laufen die phonetische und die inhaltlich-pragmatische Separierung z.T. parallel. Zweites eine dialogbezogene Einordnung zwischen Abschluss und Weiterweisung, bei der vor allem prosodische Finalitätsmerkmale die Bereitschaft von Turnübergaben signalisieren und drittens eine Einschätzung hinsichtlich des Grades an Unflüssigkeit, gesteuert vor allem durch die temporale Struktur an der prosodischen Grenze. Aus dem Zusammenspiel dieser Beurteilungen ergibt sich der Raum für die Interpretation des Gehörten im aktuellen situativen und semantisch-pragmatischen Rahmen. Aber auch dieser Ansatz zeigt nur eine grobe Richtung und müsste durch eine Vielzahl weiterer funktionaler Skalen ergänzt werden, um zu einer adäquaten Beschreibung zu gelangen.

In einigen Punkten decken sich beschriebene Sichtweisen und Forschungsziele mit denen der *Conversation Analysis*. Allerdings argumentieren einige Vertreter der CA für die Erarbeitung einer *Phonology of conversation* - ein phonologisches Regelwerk für gesprochene Sprache in der sozialen Interaktion. Im Rahmen dieser phonologischen Modellierung werden phonetische Merkmalsbündel in Einzelmerkmale dekomponiert und auf interaktionell relevante Funktionen konversationeller Aktivität bezogen (vgl. Local et al. 1985, 1986, Selting 1995). Hierbei wird die Prosodie als ein weitgehend autonomes Signalisierungssystem betrachtet.

Phonetische Forschung modelliert sprachliches Verhalten häufig durch phonologische Systeme, in denen Verbindungen abstrakter phonologischer Merkmale mit sprachlichen Funktionen definiert werden. Diese Verbindungen werden in einem Regelwerk in Form logischer Beziehungen zusammengefasst. So ein abstrahierendes System erlaubt es, phonetische Varianten unter einer Kategorie zusammenzufassen, wenn funktionale Äquivalenz der Varianten angenommen wird. Dieser Ansatz scheint sinnvoll, solange es z.B. um die Unterscheidung zwischen Wörtern durch Phoneme geht. Sowohl auf der funktionalen als auch auf der phonetischen Seite sind hierbei Kategorien gut definierbar.

Wird aber eine Phonologie der Konversation angestrebt, so ist das oben diskutierte Problem relevant, dass die funktionale Ebene komplex, hoch variabel und von kontextbedingter Interpretation abhängig ist. Somit werden Definitionen funktionaler Kategorien problematisch und postulierte Kategorien schwer generalisierbar. Auch dem Prinzip der kumulativen Signalisierung ist durch die Dekomposition und autonome Betrachtung phonetischer Merkmalsbündel schwer gerecht zu werden. Dies sind vermutlich die Gründe, warum bislang im Rahmen der CA kaum komplexe, abstrahierende phonologische Systeme für konversationelle Aktivitäten aufgestellt wurden.

Es ist also fraglich, ob die Dichotomisierung in Phonetik vs. Phonologie förderlich für die Erforschung von phonetischer Form und sprachlicher Funktion ist und ob auf dem derzeitigen Wissensstand phonologische Abstraktionen und scharfe Kategoriengrenzen sinnvolle Konstrukte sind. M.E. sind Aussagen über den Zusammenhang zwischen Form und Funktion ohne eine phonologische Zwischenebene und ohne zu starke Abstraktionen phonetischer und funktionaler Strukturen besser geeignet, sich dem Forschungsgegenstand weiter zu nähern. Hierbei sollte stets sowohl die Perspektive des Hörers als auch die des Sprechers durch perzeptions- und produktionsbezogene Untersuchungen berücksichtigt und ausgehend vom Prinzip der kumulativen Evidenz vor allem an einem flexiblen Modell funktionaler Strukturen gearbeitet werden. Dies erfordert neue methodische Paradigmen und eine weitere Öffnung der Phonetik gegenüber ihren psychologischen und soziologischen Nachbardisziplinen.

## 7 Literatur

- Altmann, H., A. Batliner und W. Oppenrieder (Hrsg.) (1989). *Zur Intonation von Modus und Fokus im Deutschen*. (Ling. Arbeiten 234), Tübingen: Niemeyer.
- Anderson, A.H., M. Bader, E.G. Bard, E. Boyle, G.M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C.F. Sotillo, H.S. Thompson und R. Weinert (1991). The HCRC Map Task Corpus. *Language and Speech* 34, 351-366.
- Auer, P. (1986). Kontextualisierung. *Studium Linguistik* 19, 22-47.
- Bannert, R. und J. Schwitalla (1999). Äußerungssegmentierung in der deutschen und schwedischen gesprochenen Sprache. *Deutsche Sprache* 4. 314-335.
- Barry, W.J. und A.J. Fourcin (1992). Levels of labelling. *Computer Speech and Language* 6, 1-14.
- Batliner, A. und M. Reyelt (1994). Ein Inventar prosodischer Etiketten für VERBMOBIL. VERBMOBIL Memo 33, Technische Universität Braunschweig, Ludwig-Maximilian-Universität München.
- Batliner, A., R. Kompe, A. Kießling, E. Nöth, H. Niemann und U. Kilian (1995). The prosodic marking of phrase boundaries: Expectations and results. In: A.J. Rubio Ayoso und J.M. López Soler (Hrsg.) *Speech Recognition and Coding: New Advances and Trends*, NATO ASI Series F. 147, 325-328. Berlin/Heidelberg/New York: Springer.
- Beckman, M. E. und J. Pierrehumbert (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255-309.
- Beckman, M. E. und G. M. Ayers (1993). *Guidelines for ToBI labelling*. Technical Report Version 1.5, Ohio State University.
- Bergmann, J.R. (1988). Ethnomethodologie und Konversationsanalyse, Studieneinheit 3: Organisationsprinzipien der sozialen Interaktion: Objekte der Konversationsanalyse. 1-52.
- Boersma, P. und D. Weenink. *Praat: doing phonetics by computer*. URL: [www.fon.hum.uva.nl/praat](http://www.fon.hum.uva.nl/praat).
- Bolinger, D. (1978). Intonation across languages. In: J.H. Greenberg (Hg.) *Universals of Human Language. Vol. 2: Phonology*. 471-524. Stanford, Cal.: Stanford University Press.
- Boomer, D. und A. Dittman (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech* 5, 215-220.
- Boomer, D. (1965). Hesitations and grammatical encoding. *Language and Speech* 8, 148-158.
- Brown, G., A. Anderson, G. Yule und R. Shillcock (1983): *Teaching Talk*. Cambridge: Cambridge University Press.

- Butcher, A. (1981). *Aspects of the Speech Pause: Phonetic Correlates and Communicative Functions. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 15*, Kiel: IPDS.
- Caspers, J. (2000). Experiments on the meaning of four types of single-accent intonation patterns in Dutch. *Language and Speech 43*, 127-161.
- Chen, A.J. (2003). Language dependence in continuation intonation. *Proc. 15th ICPhS, Barcelona*, Volume 1, 1069-1072.
- Close, D., A. Robbins, P. Rubin und R. Stallman (1993). *The GAWK manual*. Free Software Foundation.
- Cohen, A. und J. 't Hart (1967). The anatomy of intonation. *Lingua 19*, 177-192.
- Cook-Gumperz, J. und J. Gumperz (1976). Context in children's speech. *Papers on Language and Context. (Working Paper No. 46)*, Berkeley, CA: Language Behaviour Research Laboratory.
- Core, M. G. und J. F. Allen (1997). Coding dialogs with the DAMSL annotation scheme. *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, 28-35.
- Couper-Kuhlen, E. (1983). Intonatorische Kohäsion. Eine makro-prosodische Untersuchung. *Zeitschrift für Literaturwissenschaft und Linguistik 49*, 74-100.
- Croft, W. (1995). Intonation units and grammatical structure. *Linguistics 33*, 839-882.
- Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics 17*, 77-91.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. (1980). *A First Dictionary of Linguistics and Phonetics*. Boulder, CO: Westview.
- de Pijper, J.R. und A.A. Sanderman (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA 96*, 2037-2047.
- Delattre, P., E. Poenack und C. Olsen (1965). Some characteristics of German intonation for the expression of continuation and finality. *Phonetica 13*, 134-161.
- Essen, O. v. (1964). *Grundzüge der hochdeutschen Satzintonation*. Ratingen, Düsseldorf: Henn.
- Féry, C. (1993). *German Intonational Patterns*. (Linguistische Arbeiten 285).Tübingen: Niemeyer.
- Fletcher, J., R. Wales, L. Stirling und I. Mushin (2002). A dialogue act analysis of rises in Australian English map task dialogues. *Proc. of Speech Prosody 2002*, 299-302.

- Ford, C.E. und S.A. Thompson (1996). Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns. In: E. Ochs, E. A. Schegloff und S. A. Thompson (Hrsg.). *Interaction and Grammar. Studies in Interactional Sociolinguistics 13*, 134-184. Cambridge: Cambridge University Press.
- Ford, C.E., B. Fox und S.A. Thompson (1996): Practices in the construction of turns: The 'TCU' revisited. *Pragmatics 6*, 427-454.
- French, P. und J. Local (1983). Turn-competitive incomings. *Journal of Pragmatics 7*, 701-715.
- Freud, S. (1904/1971). *Zur Psychopathologie des Alltagslebens: Über Vergessen, Versprechen, Vergreifen, Aberglauben und Irrtum*. Frankfurt/M.: Fischer Bücherei.
- Gilles, P. (2000). Intonation der Weiterweisung. *InLiSt 20*, Freiburg.
- Gilles, P. (2005): *Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung*. Linguistik – Impulse und Tendenzen 6. Berlin/New York: de Gruyter.
- Goldman Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London/New York: Academic Press.
- Goldman Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech 15*, 103-113.
- Grabe, E., C. Gussenhoven, J. Haan, E. Marsi, und B. Post (1997). Preaccentual pitch and speaker attitude in Dutch. *Language and Speech 41*, 63-85.
- Grabe, E. (1998). *Comparative Intonational Phonology. English and German*. Diss. Universität Nijmegen (MPI Series in Psycholinguistics 7).
- Greenhouse, S.W. und S. Geisser (1959). On methods in the analysis of profile data. *Psychometrika 32*, 95-112.
- Grice, M. und R. Benz Müller (1995). Transcription of German intonation using ToBI-Tones. The Saarbrücken System. (Phonus 1), 33-51, Saarbrücken
- Grice, M. und S. Baumann (2002). Deutsche Intonation und GToBI. *Linguistische Berichte 191*. 267-298.
- Grosjean F. und A. Deschamps (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica 31*, 144-184.
- Grosjean, F. und J.Y. Dommergues (1983). Les structures de performance en psycholinguistique. *L'Année psychologique 83*, 513-536.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Halliday, M.A.K. (1967): *Intonation and Grammar in British English*. The Hague: Mouton.
- House, D. (1990) *Tonal Perception in Speech*. Lund: Lund University Press.

- IPDS (1995). *The Kiel Corpus of Spontaneous Speech*, Volume 1, CD-ROM#2. Kiel: IPDS.
- IPDS (1996). *The Kiel Corpus of Spontaneous Speech*, Volume 2, CD-ROM#3. Kiel: IPDS.
- IPDS (1997). *xassp* user's manual (advanced speech signal processor under the X window system). In: A.P. Simpson, K.J. Kohler, T. Rettstadt (Hrsg.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 33, 31-115, Kiel: IPDS.
- Isacenko, A. und H.-J. Schädlich (1970). *A Model of Standard German Intonation*. Volume CXIII of *Janua Linguarum, Series practica*. The Hague: Mouton.
- Jefferson, G. (1986): Notes on 'latency' in overlap onset. *Human Studies* 9, 153-183.
- Jones, D. (1960). *An Outline of English Phonetics*. Cambridge: Heffner.
- Jonson, B. (1640). *The English Grammar Made by Ben Jonson for the Benefit of all Strangers out of the Observation of the English Language now Spoken and in Use*. Reprint 1928, London: Lanston.
- Kingdon, R. (1958) *The Groundwork of English Stress*. London: Longman.
- Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3, 129-140.
- Klatt, D. und W.E. Cooper (1975). Perception of segment duration in sentence contexts. In: A. Cohen, S. Nooteboom (Hrsg.), *Structure and Process in Speech Perception*. 69-89, Heidelberg: Springer.
- Klatt, D. (1976): Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *JASA* 59, 1208-1221.
- Klinghardt, H. und G. Klemm (1920). *Übungen im Englischen Tonfall für Lehrer und Studierende*. Cöthen: Otto Schulze Verlag.
- Kohler, K.J. (1977). *Einführung in die Phonetik des Deutschen*, Berlin: Erich Schmidt.
- Kohler, K.J. (1983). Phonetic boundary signals in German. *Phonetica* 40, 89-134.
- Kohler, K.J. (1991a). A model of German intonation. In: K.J. Kohler (Hg.), *Studies in German Intonation*, AIPUK 25, 295-360, Kiel: IPDS.
- Kohler, K.J. (1991b). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. In: K.J. Kohler (Hg.), *Studies in German Intonation*, AIPUK 25, 115-186.
- Kohler, K.J. (1992). Automatische Generierung der kanonischen Transkription und des Aussprachlexikons. In: K.J. Kohler (Hg.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 26, 175-196, Kiel: IPDS.

- Kohler, K.J. (1995a). Einführung in die Phonetik des Deutschen, (2. Aufl.). Berlin: Erich Schmidt.
- Kohler, K.J. (1995b). PROLAB – the Kiel system of prosodic labelling. *Proc. 13<sup>th</sup> ICPPhS*, Stockholm, Volume 3, 162-165.
- Kohler, K.J., M. Pätzold und A.P. Simpson (1995). From Scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech. In: K.J. Kohler, M. Pätzold und A.P. Simpson (Hrsg.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 29*, Kiel: IPDS.
- Kohler, K.J. (1997). Parametric control of prosodic variables by symbolic input in TTS synthesis. In: J. van Santen, R. Sproat, J. Olive und J. Hirschberg (Hrsg.), *Progress in Speech Synthesis*, 459-477. New York: Springer.
- Kohler, K.J., M. Pätzold und A.P. Simpson (1997). From the acoustic data collection to a labelled speech data bank of spoken Standard German. In: A.P. Simpson, K.J. Kohler, T. Rettstadt (Hrsg.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 32*, 1-30, Kiel: IPDS.
- Kohler, K.J. (2001). Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? In: N. Grønnum und J. Rischel (Hrsg.), *To Honour Eli Fischer-Jørgensen: Festschrift on the Occasion of her 90<sup>th</sup> Birthday, February 11<sup>th</sup>, 2001*, Travaux du Cercle Linguistique de Copenhague 31, 174-211, Copenhagen: Reizel.
- Kohler, K.J. (2005a). Form and function of non-pitch accents. In: K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 97-123, Kiel: IPDS.
- Kohler, K.J. (2005b). Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. In: K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 125-142, Kiel: IPDS.
- Kohler, K.J., B. Peters und T. Wesener (2005). Phonetic exponents of disfluency in German spontaneous speech. In K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 185-201, Kiel: IPDS.
- Koiso, H., Y. Horiuchi, S. Tutiya, A. Ichikawa und Y. Den (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech* 41, 295-321.
- Ladd, R.D. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lehiste, I (1963). Review of K. Hadding-Koch. *Language* 39, 352-360.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In: A. Cohen und S. Nooteboom (Hrsg.), *Structure and Process in Speech Perception*, 195-206. Heidelberg: Springer.

- Lehiste, I. (1979): Perception of sentence and paragraph boundaries. In: B. Lindblom und S. Öhman (Hrsg.) *Frontiers of Speech Communication Research*, 191-201, London: Academic Press.
- Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge/London: MIT Press.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Liebermann, P. (1967). *Intonation, Perception, and Language*. Research Monograph 38. Cambridge/Mass.: MIT Press.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* 140, 44-53, New York: McGraw-Hill.
- Lindbloom, B. (1978): Final lengthening in speech and music. *Travaux de l'Institut de Phonétique de Lund*, 85-101.
- Local, J., W.H.G. Wells und M. Sebba (1985). Phonology for Conversation. Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9, 309-330.
- Local, J. und J. Kelly (1986). Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies* 9, 185-204.
- Local, J., J. Kelly und W. Wells (1986) Towards a phonology of conversation: turn-taking in urban Tyneside speech. *Journal of Linguistics* 22, 411-437.
- Local, J. (2000-2002). Phonetics of talk-in-interaction: a study of the interactional functioning of phonetic detail in everyday talk. *British Academy Readership*. <http://www-users.york.ac.uk/~lang4/ba/Phonetics-of-talk.pdf>
- Martin, J.G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior* 9, 75-78.
- Megyesi, B. und S. Gustafson-Capkova (2002). Production and perception of pauses and their linguistic context in read spontaneous speech in Swedish. *Proc. ICSLP, Denver*, 2153-2156.
- Moulines, E. und F. Carpentier (1990): Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9, 453-467.
- Nakatani, C.H. und J. Hirschberg (1994). A corpus-based study of repair cues in spontaneous speech. *JASA* 95, 1603-1616.
- Nespor, M. und I. Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris.
- O'Connor, J.D. und G.F. Arnold (1973). *Intonation of Colloquial English: a Practical Handbook*. 2. ed., London: Longman.
- Ohala, J. und M. Hirano (1967). Studies of pitch change in speech. *UCLA Working papers in phonetics*, 80-84.
- Ohala, J.J. (1983). Cross-language use of pitch: an ethological view. *Phonetica* 40, 1-18.

- Ohala, J.J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica* 41, 1-16.
- Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *JASA* 54, 1235-1247.
- Ostendorf, M., P. Price, J. Bear und C. Wightman (1990). The use of relative duration in syntactic disambiguation. *Proc. 3<sup>rd</sup> DARPA Workshop on Speech and Natural Language*, 26-31.
- Pätzold, M. und A.P. Simpson (1994). Das Kieler Szenario zur Terminabsprache. *Verbmobil Memo* Nr. 53.
- Pätzold, M. (1997). KielDat – data bank utilities for the Kiel Corpus. In: A.P. Simpson, K.J. Kohler, T. Rettstadt (Hrsg.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 32, 117-126, Kiel: IPDS.
- Peters, B. (1999). Prototypische Intonationsmuster in deutscher Lese- und Spontansprache. In: K.J. Kohler (Hg.), *Phrase-level Phonetics and Phonology of German*, AIPUK 34, 1-175. Kiel: IPDS.
- Peters, B. (2000). Individuelle und geschlechtsspezifische Unterschiede in der prosodischen Gestaltung deutscher Lese- und Spontansprache. In: H.D. Schlosser (Hg.) *Sprache und Kultur*, Forum Angewandte Linguistik 38, 153-162, Frankfurt/M.: Lang.
- Peters, B. (2001). 'Video Task' oder 'Daily Soap Szenario': Ein neues Verfahren zur kontrollierten Elizitation von Spontansprache. URL: [http://www.ipds.uni-kiel.de/pub\\_exx/bp2001\\_1/Linda21.html](http://www.ipds.uni-kiel.de/pub_exx/bp2001_1/Linda21.html)
- Peters, B. und K.J. Kohler (2004). *Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM*. URL: [www.ipds.uni-kiel.de/kjk/pub\\_exx/bpkk2004\\_1/TrainerA4.pdf](http://www.ipds.uni-kiel.de/kjk/pub_exx/bpkk2004_1/TrainerA4.pdf)
- Peters, B. (2005a). The database – The Kiel Corpus of Spontaneous Speech. In: K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 1-6, Kiel: IPDS.
- Peters, B., K.J. Kohler und T. Wesener (2005a). Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache – Statistische Verteilung und sprachliche Funktion. In: K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 7-54, Kiel: IPDS.
- Peters, B., K.J. Kohler und T. Wesener (2005b). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In: K.J. Kohler, F. Kleber, B. Peters (Hrsg.), *Prosodic Structures in German Spontaneous Speech*, AIPUK 35a, 143-184, Kiel: IPDS.
- Pheby, J. (1984). Phonologie: Intonation. In: K. E. Heidolph, W. Flämig, W. Motsch (Hrsg.), *Grundzüge einer deutschen Grammatik*. 2., unveränderte Auflage, 839-897, Berlin: Akademie.

- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT. (Distributed 1987 by Indiana University Linguistics Club).
- Pierrehumbert, J. und J. Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In: J. Cohen, P. Morgan und M. Pollack (Hrsg.), *Intentions in Communication*, 271-311. Cambridge, Mass.: MIT Press.
- Pompino-Marschall, B. (1995). Einführung in die Phonetik. Berlin/New York: de Gruyter.
- Price, P.J., M. Ostendorf, S. Shattuck-Hufnagel und C. Fong (1991). The use of prosody in syntactic disambiguation. *JASA* 90, 2956-2970.
- Rietveld, T., J. Haan, L. Heijmans und C. Gussenhoven (2002). Explaining attitudinal ratings of Dutch rising contours: morphological structure vs. the frequency code. *Phonetica* 59, 180-194.
- Sacks, H. (1963) Sociological description. *Berkeley Journal of Sociology* 8, 1-16.
- Sacks, H., E.A. Schegloff und G. Jefferson (1974). A simplest systematics for the organisation of turn-taking for conversation, *Language* 50, 696-735.
- Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of *uh huh* and other things that come between sentences. In: D. Tannen (Hg.): *Analyzing discourse: Text and talk*. Georgetown University Round Table on Languages and Linguistics 1981, 71-91, Washington, D.C.: Georgetown University Press.
- Schegloff, E. (1996): Turn-Organization: one intersection of grammar and interaction. In: E. Ochs, E. Schegloff, S. Thompson (Hrsgs.), *Interaction and Grammar*, 52-133, Cambridge: Cambridge University Press.
- Schwitalla, J. (1997). *Gesprochenes Deutsch. Eine Einführung*. Berlin: Erich Schmidt.
- Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. In: T. Fretheim (Hg.), *Nordic Prosody Vol. 2*, 111-140, Trondheim: TAPIR.
- Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology* 3, 371-405.
- Selting, M. (1993): Phonologie der Intonation. Probleme bisheriger Modellen und Konsequenzen einer neuen interpretativ-phonologischen Analyse. *Zeitschrift für Sprachwissenschaft* 11, 99-138.
- Selting, M. (1995). *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Selting (1998). TCUs and TRPs: the construction of 'units' in conversational talk. *InLiSt* 4.
- Selting, M., P. Auer, B. Barden, J. Bergmann, E. Couper-Kuhlen, S. Günthner, U. Quasthoff, C. Meier, P. Schlobinski und S. Uhmman (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173, 91-122.

- Shriberg, E., A. Stolcke, D. Hakkani-Tür und G. Tür (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication* 32, 127-154.
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert und J. Hirschberg (1992). ToBI: A standard for labelling English prosody. *Proc. of the Second International Conference on Spoken Language Processing*, 867-870.
- Simpson, A. P. (1991). *Writing Phonological Statements from Naturally Occurring Talk: An Experiment in Method*. Ph. D. thesis, University of York.
- Sweet, H. (1877). *A Handbook of Phonetics*. London.
- Sweet, H. (1890). *Primer of Phonetics*. Oxford.
- Swerts, M. und R. Collier (1992): On the controlled elicitation of spontaneous speech. *Speech Communication* 11, 463-468.
- Szczepek, B. (2000). Formal Aspects of collaborative productions in English conversation. *InLiSt* 17.
- t'Hart, J.R., R. Collier, und A. Cohen (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Thurston, L.L. und E.J. Chave (1929) *The Measurement of Attitude*. Chicago: University of Chicago Press.
- Trager, G. L. und H. L. Smith (1951). *An Outline of English Structure*. Norman, OK: Battenburg Press.
- Trubetzkoy, N.S. (1939). Grundzüge der Phonologie. (Travaux du Cercle Linguistique de Prague VII), Prag.
- Uldall, E. (1964). Dimensions of meaning in intonation. In: D. Abercrombie, D.B. Fry, P.A.C. MacCarthy, N.C. Scott, J.L. Trim (Hrsg.). *In Honour of Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday*, 271-279, London: Longman.
- URL(1): <http://verbmobil.dfki.de>
- URL(2): <http://www.lindenstrasse.de>
- Watzlawick, P. (1983): *Anleitung zum Unglücklichsein*. München: Piper
- Webster, N. (1789): *Dissertations on the English Language*. Boston.
- Wells, J., W. Barry und A. Fourcin (1989). Transcription, labelling and reference. In: A. Fourcin, G. Harland, W. Barry und V. Hazan (Hrsg.). *Speech Technology Assessment. Towards Standards and Methods for the European Community*, 141-159, Chicester: Ellis Harwood.
- Wichmann, A. (2000). *Intonation in Text and Discourse. Beginnings, Middles and Ends*. Studies in Language and Linguistics. Harlow: Pearson Education.

- Wichmann, A. und J. Caspers (2001). Melodic cues to turn-taking in English: evidence from perception. *Proc. 2<sup>nd</sup> SIGdial Workshop on Discourse and Dialogue, Aalborg*. 180-185.
- Wightman, C.W., S. Shattuck-Hufnagel, M. Ostendorf. und P.J. Price, (1992). Segmental duration in the vicinity of prosodic phrase boundaries. *JASA* 91, 1707 - 1717.
- Willems, N.J. (1983). Towards an objective course in English intonation: standardized precepts. In: M. van den Broecke, V. van Heuven, W. Zonneveld (Hrsg.), *Sound structures. Studies for Antonie Cohen*, 281-296, Dordrecht/Cinnaminson: Foris Publications.
- Wilson, T.P. (1978). Theorien der Interaktion und Modelle soziologischer Erklärung. In: Arbeitsgruppe Bielefelder Soziologen (Hrsg.), *Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit. Band 1: Symbolischer Interaktionismus und Ethnomethodologie*, 54-79. Reinbek bei Hamburg.
- Zwirner, E. und K. Zwirner (1937). Phonometrischer Beitrag zur Frage der Lesepausen. *Arch. Néerl. Phon. Exp.* 13, 111-128.

## **Anhang**

Der Anhang enthält die schriftlichen Vorlagen für die Instruktionen der Perzeptionsexperimente. Die Präsentation der Instruktionen vor den Experimenten erfolgte als vorgelesene Version von CD.

### **Einleitung Experiment 1a, 1b – Kohäsion an turninternen prosodischen Grenzen**

In dem folgenden Hörexperiment geht es um die phonetische Gliederung von Äußerungen. Äußerungen können phonetisch als eine globale Einheit gesprochen werden, wie in dem folgenden Fall:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." ohne interne Pause, ohne finale Längung und mit ebener Kontur.*

Hier werden sie wahrscheinlich alle sagen, dass diese Äußerung phonetisch nicht in verschiedene Teile untergliedert ist. Äußerungen können aber auch so gesprochen werden, dass eine ebensolche Untergliederung phonetisch angezeigt wird, wie in dem folgenden Beispiel:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit kurzer interner Pause, leichter finaler Längung und stark fallender Kontur.*

Wenn sie die beiden Beispiele jetzt noch mal hintereinander hören, werden sie im zweiten gegenüber dem ersten eine deutliche Trennung wahrnehmen.

*Es folgen die beiden obigen Audiobeispiele.*

Diese Trennung kann von unterschiedlicher Stärke sein. Bei dem folgenden Beispiel urteilen sie sicher, dass es sich um einen starken Grad der Trennung handelt:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit langer interner Pause, starker finaler Längung und stark fallender Kontur.*

Im folgenden Beispiel aber eher um einen schwachen Grad der Trennung:

*Es folgt ein Audiobeispiel des Satzes "Das ist mein jüngster Sohn – mit seiner Freundin." mit ohne interne Pause, leichter finaler Längung und leicht fallender Kontur.*

In dem nun folgenden Hörexperiment möchten wir von Ihnen Auskunft darüber haben, wie sie die Stärke der phonetischen Trennung in Äußerungen wahrnehmen. Die Skala dieser Trennung rangiert zwischen dem Fehlen einer Trennung an einem Ende und einer extremen Trennung am anderen Ende. In der Mitte dieser Skala ist ein mittlerer Grad der Trennung. Die gesamte Skala wird eingeteilt in sieben Stufen von 0 (keine Trennung) bis 6 (extremer Trennungsgrad); der Wert drei entspricht einem mittleren Trennungsgrad. Sie werden im Experiment in vielfacher Wiederholung einen Satz hören, der immer die gleichen Wörter enthält. Er variiert aber in der Aussprache und zwar im Grad der Trennung in zwei Äußerungsteile.

Zum Ablauf des Experiments: Auf dem Ihnen vorliegenden Fragebogen sehen Sie Blöcke von jeweils zehn Zeilen, die in sieben Antwortspalten unterteilt sind. Bitte geben Sie für jede der aufeinander folgenden Äußerungen in der entsprechenden Zeile ein Urteil darüber ab, wie stark Sie die Trennung innerhalb der Äußerung wahrnehmen. Nach jeder vorgespielten Äußerung ist eine Pause, in der Sie Ihr Urteil abgeben können. Bitte kreuzen Sie für jede Äußerung auf der entsprechenden Zeile den Grad der wahrgenommenen Trennung an. Nach jeweils zehn Äußerungen kommt eine längere Pause und ein Piepton. Um Ihnen den Ablauf besser zu verdeutlichen werden jetzt 10 Probebeispiele vorgespielt, die Sie bitte in Block 1 beurteilen. Danach können noch Fragen zum Ablauf gestellt werden.

Vielen Dank im Voraus für Ihre Mitarbeit.

## **Einleitung Experiment 1c – Kohäsion an turninternen prosodischen Grenzen**

Die Einleitung für Experiment 1c ist identisch mit dem obigen Text. Es ist allerdings eine kurze Passage geändert worden, die sich darauf bezieht, dass hier drei unterschiedliche Sätze beurteilt werden sollen.

## **Einleitung Experiment 2 - Turnübergabeerwartung**

Das nun folgende Experiment ist völlig parallel zu dem eben durchgeführten Experiment aufgebaut. Sie bekommen wieder den selben Satz in vielfacher Wiederholung zu hören und sollen wieder Beurteilungen zu dem Gehörten auf Skalen abgeben.

Die Fragestellung hierbei ist aber eine ganz andere als eben. Sie sollen nun eine Einschätzung darüber abgeben, wie sich wohl der Sprecher nach dem geäußerten Satz in einer Gesprächssituation verhalten würde. Ob er noch weitersprechen wird oder ob er seine Äußerung mit diesem Satz beendet und das Wort dem Gesprächspartner überlässt.

Hierfür können Sie sich vorstellen, dass der Satz in einem zwanglosen Gespräch zwischen Freunden geäußert wird. Z.B. als Kommentar zu einer Schilderung eines Missgeschicks oder als Reaktion auf die Beschreibung eines Zahnarztbesuchs. Nach diesem Kommentar - *Das ist ja nun nicht so angenehm* - gibt es für den Sprecher zwei Möglichkeiten: Weitersprechen oder nicht weitersprechen. Das Experiment hat zum Ziel herauszufinden, welche phonetischen Eigenschaften von Äußerungen dazu dienen, dem Hörer zu signalisieren, ob noch weitergesprochen werden soll oder ob eine Äußerung zum Ende gekommen ist.

Auch in diesem Experiment wird mit Skalen gearbeitet, damit auch Urteile gegeben werden können, die eine gewisse Unsicherheit in der Einschätzung des Gesprächsfortgangs ausdrücken. Wenn Sie also klar der Meinung sind, dass der Sprecher noch weitersprechen wird, kreuzen Sie bitte die 0 an. Wenn Sie denken, dass der Sprecher am Ende seiner Äußerung angekommen ist, kreuzen Sie die 6 an. Die 3 in der Mitte der Skala würde eine völlige Unentschiedenheit bedeuten. Bitte scheuen Sie sich auch in diesem

Experiment nicht davor, die Extreme der Skala mit in ihre Beurteilungen einzubeziehen. Das Experiment dauert etwa 15 Minuten.

Nun folgt wieder ein Block mit 10 Wiederholungen, damit Sie sich an die Aufgabenstellung gewöhnen können. Danach können noch Fragen gestellt werden.

### **Einleitung Telexperiment 3a: Sprechflüssigkeit**

Im folgenden Hörexperiment geht es um den Unterschied zwischen einer zögerlichen und einer flüssigen Sprechweise. Jeder kennt es aus dem Alltag, dass man manchmal schnell und sicher die eigenen Gedanken äußern kann und zusammenhängend und flüssig spricht. Manchmal aber fallen einem die richtigen Worte nicht gleich ein, und man kommt ins Stocken - so entsteht eine zögerliche Sprechweise.

In dem nun folgenden Hörexperiment geht es darum, dass Sie Äußerungen hinsichtlich der Sprechweise beurteilen. Also eine Einschätzung darüber abgeben, ob es sich um eine flüssige oder eher um eine unflüssige Sprechweise handelt. Sie werden im Experiment in vielfacher Wiederholung einen Satz hören, der immer die gleichen Wörter enthält, sich aber in phonetischen Details unterscheidet.

Auf den Antwortbögen vor Ihnen sehen Sie eine Antwortskala mit sieben Stufen. Die Skala rangiert von 0 bis 6. Nehmen Sie bitte für jede Wiederholung des Satzes eine Beurteilung hinsichtlich der Flüssigkeit der Äußerung vor. Wenn Sie den Eindruck einer völlig flüssigen Sprechweise haben, kreuzen Sie bitte die 6 an. Wenn der gehörte Satz einen sehr zögerlichen Eindruck macht, kreuzen Sie die 0 an. Viele der Äußerungen werden aber vermutlich zwischen diesen Extremen liegen. Machen sie Ihr Kreuz dann an einer Stelle der Skala, die für Ihr Empfinden diesem mittleren Grad der Flüssigkeit entspricht. Bitte scheuen Sie sich nicht, auch die äußeren Bereiche der Skala mit auszunutzen. D.h. wenn Sie den Eindruck einer deutlich zögerlichen Sprechweise haben, kreuzen sie ruhig die 0 am linken Ende der Skala an. Wenn Ihnen die Äußerung völlig flüssig erscheint wählen Sie die 6.

Nach jeder vorgespielten Äußerung ist eine Pause, in der Sie Ihr Urteil abgeben können. Nach jeweils zehn Äußerungen kommt eine längere Pause und ein Piepton. Das ganze Experiment dauert etwa 15 Minuten. Um Ihnen den Ablauf besser zu verdeutlichen werden jetzt 10 Probebeispiele vorgespielt, die Sie bitte im ersten Antwortblock - dem Testblock - beurteilen. Danach können noch Fragen zum Ablauf gestellt werden.

Vielen Dank im Voraus für Ihre Mitarbeit.

### **Einleitung Telexperiment 3a: Sprechflüssigkeit**

Die Einleitung für Experiment 3b ist identisch mit dem obigen Text. Es ist allerdings eine kurze Passage geändert worden, die sich darauf bezieht, dass hier drei unterschiedliche Sätze beurteilt werden sollen.

## Lebenslauf Benno Peters

29.10.1970	geboren in Kiel
1977-1981	Grundschule Raisdorf
1981-1990	Gymnasium Elmschenhagen
1991-1992	Zivildienst an der Rudolf-Steiner-Schule Kiel
1993-1998	Studium an der Christian-Albrechts-Universität zu Kiel, HF: Phonetik und digitale Sprachverarbeitung, NF: Allg. Sprachwissenschaft, Psychologie
1999-2006	Wiss. Angestellter am Institut für Phonetik und digitale Sprachverarbeitung der Christian-Albrechts-Universität zu Kiel



---

Benno Peters