phonetik AIMS

Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung
Lehrstuhl für experimentelle Phonetik Universität Stuttgart

1997 • VOL. 3 • NO. 4

Intonation und Bedeutung

Jörg Mayer

AIMS - Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, Universität Stuttgart
ISSN 1434-0763

INTONATION UND BEDEUTUNG

Aspekte der Prosodie-Semantik-Schnittstelle im Deutschen

Jörg Mayer

Dissertation zur Erlangung des akademischen Grades Dr. phil.,
vorgelegt an der Fakultät für Philosophie der Universität Stuttgart

August 1997

Erstberichter: Prof. Dr. Grzegorz Dogil
Mitberichter: Prof. Mats Rooth, PhD

Universitä Stuttgart
Institut für Maschinelle Sprachverarbeitung
Azenbergstr. 12, 70174 Stuttgart

Tag der mündlichen Prüfung: 16. Februar 1998

Jörg Mayer
Universitä Stuttgart
Institut für Maschinelle Sprachverarbeitung
Lehrstuhl für Experimentelle Phonetik
Azenbergstr. 12
D-70174 Stuttgart

joemayer@ims.uni-stuttgart.de

Summary

This thesis is an approach to examine structural relations between prosody and discourse semantics by means of laboratory phonology. After a sketch of the basic concepts of suprasegmental phonology and intonational phonetics, the introductory part I continues with a discussion of the tone sequence model, the tone sequence analysis of German intonational patterns and current papers on intonational meaning.

Part II deals with two analysis tools: (1) GToBI(S), a ToBI-based annotation system for prosody and (2) a speech synthesis system, which integrates rule-based F_0 contour generation and PSOLA resynthesis techniques.

In part III, I will present the results of three experiments on (1) disambiguation of sentential adverbs, (2) anaphora resolution, and (3) prosodic correlates of global discourse structure. The leading hypothesis behind the experiments (1) and (2) was that if accent is derivable from the abstract semantic representation, it should be possible to identify semantic structures and resolve semantic ambiguities by looking at the intonational prominence. The results on this hypothesized correlation that I obtained were not always significant and thus sometimes difficult to interpret. However, I can report that in some restricted areas intonational prominence correlates well with formal semantic modelling.

Disambiguation of sentential adverbs: According to Discourse Representation Theory (Kamp & Reyle, 1993) the resolution of ambiguity of focus sensitive adverbs like *only*, *also*, *first*, *again*, etc., is possible without explicit analysis of semantic context inferences (Eberle, 1996). Such an in situ analysis is much faster and more direct than an inference based analysis and is favoured for applications like automatic understanding and machine translation. In a production experiment with the German adverb *erst* ‘*first*’, for

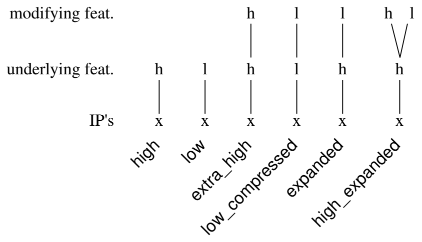
which three different readings can be identified, I tested whether the presence and type of local prominence (pitch-accent) influence the resolution of ambiguities. The first reading presupposes a reference event so that the event expressed in the *erst*-sentence is understood as the first of an elaboration sequence. Elaboration sequences turn out to be preferably realized with L*(H) pitch accents. The other two readings have in common that a set of alternatives to the event expressed in the *erst*-sentence is presupposed. This focal behaviour appears to be realized with the H*(L) pitch accent. The intrinsic order of alternatives which distinguishes the two readings is not intonationally marked.

Anaphora resolution: It has been often claimed that placing an accent on a pronoun in a antecedent-anapher sequence influences the preferred interpretation of the pronoun. An experimental phonetic investigation of this phenomena, however, has (to my knowledge) never been carried out. For purposes of such an experiment, 32 discourse contexts have been constructed. Each discourse consisted of 2 sentences, the second sentence including one or two pronouns. All contexts were resynthesized with systematic variation of the position and the type of pitch accents resulting in 174 items. A perceptual experiment in which 10 subjects had to identify the preferred readings by pointing to pictures was carried out. The results showed no significant changes for semantic reading! Neither the position of pitch accent nor its type changed the preferred interpretation of the antecedent-anapher relation. The sheer presence of (any type of) a pitch accent on an anaphoric pronoun is not sufficient to influence the choice of an antecedent for this pronoun. A production study with comparable data, however, shows that in cases where speakers seek disambiguation they do not simply stress pronouns and unstress everything else like we did in our material for the perceptual experiment. Rather, the relation between the accent in the neutral position (in most cases the direct object) and the additional accent[s] on the pronoun[s] is changed. The F_0-range on pronouns in a neutral realisation (with unaccented pronoun[s]) was in average 10,4 Hz and on stressed lexical heads 82,3 Hz. If the reference switch was attempted at, the F_0-range on the pronouns was increased by the speaker to 36,8 Hz in average, and, at the same time, the F_0-range of the neutral accent

was decreased to 41,3 Hz in average. In order to model this behaviour one would have to devise an experiment in which not the absolute presence or absence of the pitch accent is a variable, but in which relative prominence among all pitch accents is systematically varied. This is, however, not trivial. Firstly, the relative prominence is difficult to model within the standard models of intonation (like the tone sequence model). Secondly, phonetic theory has unresolved problems with the perceptual modelling of relative prominence effects. As long as these phonological and methodological problems are not clarified, the results of the experiments with relative prominence remain uninterpretable.

Prosodic correlates of global discourse structure: The third experiment was designed as an in-depth production study to examine the relation between global discourse structure and prosodic features such as pause duration and register variation. In the SDRS-model (Asher, 1993), the global structure of discourse is defined as a set of formal relations among the individual discourse segments. The relations between discourse segments include elaboration, continuation, contrast, etc. These relations are illustrated by the opening paragraph of a short story which has been very extensively studied in the literature on Discourse Representation Theory. The rendition of the paragraph by one of the semanticists who formally described its underlying discourse structure has been recorded and prosodically described. 18 full intonational phrases have been identified. The intonational phrases were integrated into larger groups by means of pauses. It turned out that the length of a pause correlated well with an abstract discourse relation that Asher formally defined as d-dominance. Informally, it is a relation between a constituent with a more general semantics and a constituent with a more specific semantics. The stretches of discourse which have different d-dominating categories are separated by ‘very long’ (> 1.4 s) pauses. A pause between two constituents for which d-dominance holds is ‘long’ (> 1 s). ‘Short’ pauses (< 0.8 s) are expected between coordinated discourse constituents. Discourse irrelevant intonational phrase boundaries are coded by pauses which are ‘very short’ (< 0.6 s).

Another important global parameter which integrates more than one intonational phrase (IP) and which correlates with the semantic discourse relations is intonational register. According to Clements (1981) the pitch range of a speaker is divided into two categorical register levels represented as {h} and {l}. These underlying register levels (which phonetically may, and do overlap) are further modified by a set of register modification features. In my study I have identified the following set of discourse relevant register categories:



Given this model, the data that we have analysed shows the following tendencies.

- Underlying register features reflect internal structure of the discourse segments: All intonational phrases which are non-final within their discourse segment take the underlying {h} register feature. The segment final IP's associate with the underlying {l} register.

- Register modification features reflect relations among discourse segments: background and elaboration are coded by {l} modification of the underlying {l} register; d-dominating discourse segments are introduced by {hl} register modification of their initial intonational phrase; d-dominating discourse segments which are in the domain of higher d-dominating discourse segments (i.e. sub-topics) are introduced by {h} modification of the underlying register; contrast is coded by {l} modification of the underlying {h} register.

Inhalt

Vorbemerkungen

Ohne die Unterstützung der vielen Freunde und Kollegen, bei denen ich mich hier ganz herzlich bedanken möchte, wäre diese Arbeit in dieser Form nicht möglich gewesen. Alle, die hier genannt sind, haben dazu beigetragen, daß diese Arbeit besser wurde; wenn dennoch Mängel und Unzulänglichkeiten vorhanden sind, so liegt dies allein in meiner Verantwortung.

An erster Stelle möchte ich mich bei meinen Eltern bedanken, für die Unterstützung und das Vertrauen in all den Jahren.

Greg Dogil, der mich ausgebildet und gefördert hat, hat diese Arbeit auf die beste Weise betreut und mit seinen Ideen und Kommentaren bereichert. Gabriele Scharf gab mir Rückhalt, als Freundin und Linguistin; und sie hatte die undankbare Aufgabe, meine Fehler zu korrigieren. Anton Batliner, Ralf Benzmüller, Martine Grice und Matthias Reyelt – der GToBI-Gruppe – danke ich für die unkomplizierte Zusammenarbeit und die vielen wichtigen Diskussionen. Caroline Féry und Kurt Eberle haben Teile dieser Arbeit gelesen und kritisch kommentiert. Mats Rooth und Greg Dogil habe ich zu Danken für die Begutachtung dieser Arbeit, Stefan Rapp für seine geduldige Hilfe. Für die gute Atmosphäre, in der diese Arbeit entstehen konnte, und für alle Hilfen danke ich den Stuttgarter Phonetikern und Sabine Schmid.

Diese Arbeit habe ich geschrieben im Rahmen meiner Mitarbeit im Projekt C4 „Entwicklung eines prosodischen Moduls für die Diskursrepräsentationstheorie" des Sonderforschungsbereichs 340 „Sprachtheoretische Grundlagen für die Computerlinguistik"

(Universität Stuttgart/Universität Tübingen), gefördert von der Deutschen Forschungsgemeinschaft. Eine kurze Zusammenfassung dieser Arbeit wurde als Teil eines SFB-Zwischenberichts veröffentlicht (Dogil, Kuhn, Mayer & Rapp, 1997).

Eine frühere Version von Kapitel 6 habe ich auf dem *Discourse Anaphora and Resolution Colloquium* an der Universität Lancaster (17./18.7.1995) vorgestellt.

Einleitung

Thema dieser Arbeit ist die Interaktion zwischen Prosodie und Semantik. Wie immer bei der Untersuchung linguistischer Schnittstellen bieten sich auch bei diesem Thema mindestens zwei verschiedene Möglichkeiten der Bearbeitung an: Entweder das Thema wird von Spezialisten untersucht und der Schwerpunkt liegt auf einem der beiden Teilbereiche, wobei der jeweils andere Teilbereich meist nur kursorisch bearbeitet wird oder – und dies scheint mir der vielversprechendere Weg zu sein – die Schnittstellenproblematik wird im Rahmen einer interdisziplinären Arbeitgruppe bearbeitet. Wir haben in Stuttgart versucht, mit einem Projekt innerhalb des interdisziplinär ausgerichteten Sonderforschungsbereichs 340 (Stuttgart/Tübingen) diesen zweiten Weg zu verfolgen. An dem Projekt sind Phonologen bzw. Phonetiker, Semantiker und Informatiker beteiligt. Diese Zusammenarbeit dauert nun, zum Zeitpunkt, da ich diese Arbeit schreibe, etwas mehr als 2 Jahre. Insofern dürfte es nicht verwundern, daß das Ziel dieser Zusammenarbeit – ein integriertes phonologisch-semantisches Modell der Prosodie-Semantik-Schnittstelle – noch nicht erreicht ist. Dies ist ein Grund, weshalb diese Arbeit auf den ersten Blick eher an die erste der beiden oben genannten Bearbeitungsmöglichkeiten denken läßt: an die Arbeit eines Spezialisten, in meinem Fall eines Phonologen bzw. Phonetikers. Der zweite Grund dafür, daß der Schwerpunkt der vorliegenden Arbeit auf den phonologischen und phonetischen Aspekten der Prosodie-Semantik-Schnittstelle liegt, ist der, daß bei einer Doktorarbeit die Untersuchungen und Überlegungen des Autors im Vordergrund stehen sollten. Angesichts der zunehmenden Komplexität in allen Teilbereichen der Linguistik führt dies in den allermeisten Fällen zu Arbeiten, die ein bestimmtes Thema von einem spezialisierten Standpunkt aus behandeln. Die vorliegende Arbeit bildet hierin keine Ausnahme. Dennoch habe ich versucht – und ich hoffe, daß dies zum Ausdruck kommt –, das, was ich bei der Projektarbeit und darüber hinaus in Stuttgart über semantische Theorien und Modelle gelernt habe, bei meiner Darstellung und in meinen Analysen zu berücksichtigen. Abgesehen von den Analysen spezieller Phänomene, die in den folgenden Kapiteln behandelt werden und auf die ich in dieser Einleitung nicht näher eingehen möchte, macht sich dieser ‚semantische Einfluß' generell darin bemerkbar, welche Theorien und Ansätze zur Modellierung des semantischen Teils der Prosodie-Semantik-Schnittstelle ausgewählt wurden. Da nicht eine Theorie alle hier behandelten Phänomene abdeckt, werde ich mehrere semantische Ansätze vorstellen, jedoch überwiegend solche, die sich einer gegenseitigen Integration nicht grundsätzlich widersetzen. Im wesentlichen handelt es sich dabei um die Diskursrepräsentationstheorie von Hans Kamp (Kapitel 5 und 7), die Fokussemantik von Mats Rooth (Kapitel 6) und die Theorie der Diskursstruktur von Nick Asher (Kapitel 7). An der Integration dieser drei Modelle arbeiten mehrere internationale Forschergruppen, so daß diese Auswahl eine sinnvolle und zukunftssichere Basis für Untersuchungen der Art, wie sie hier vorgestellt werden, zu sein scheint.

Der Untertitel dieser Arbeit weist bereits darauf hin, daß der erste Satz der Einleitung einer Einschränkung bedarf. Die folgenden Seiten entwerfen kein umfassendes Modell der Interaktion zwischen Prosodie und Semantik. Zu komplex ist das Thema und zuviele Phänomene sind noch nicht einmal erkannt geschweige denn zumindest ansatzweise erforscht. Meine Untersuchungen beschäftigen sich stattdessen mit einzelnen, wie ich meine nicht unwesentlichen Aspekten dieser Interaktion – insbesondere der Interaktion zwischen *Intonation* und Semantik – und sollen Wege aufzeigen, wie man sich dem Ziel, einem umfassenden Modell der Prosodie-Semantik-Schittstelle, annähern kann.

Vor dem experimentellen, dritten Teil der Arbeit, der diesen Untersuchungen gewidmet ist, werden in Teil I zunächst die phonetischen und phonologischen Grundlagen eingeführt. Kapitel 1 stellt grundlegende Konzepte und Paradigmen der suprasegmentalen Phonologie und Phonetik vor. Eine wichtige Funktion dieses Kapitels ist eine – zumindest für diese Arbeit gültige – Vereinheitlichung der terminologischen Vielfalt auf dem Gebiet der Prosodie bzw. der Intonation. Am Ende von Kapitel 1 steht eine Übersicht über das, was in den folgenden Kapiteln unter Intonation verstanden wird. Mit Kapitel 2 folgt dieser allgemeinen Einführung eine vertiefende Darstellung der Phonologie der Intonation. Der Schwerpunkt liegt dabei auf dem Ton-Sequenz-Modell von Janet Pierrehumbert. Nach einem Überblick über das Originalmodell werde ich die Ton-Sequenz-Analyse der Intonation des Deutschen von Caroline Féry darstellen. Den Abschluß von Kapitel 2 bildet eine Diskussion neuerer Ansätze im Bereich Intonation und Bedeutung, auf die ich mich in späteren Kapiteln teilweise beziehen werde.

In Teil II werde ich zwei computerlinguistische Werkzeuge vorstellen, die auf der Basis des Ton-Sequenz-Modells entwickelt wurden, um die intonatorische Analyse und die linguistische Aufbereitung gesprochener Sprache zu unterstützen. Dies ist zum einen ein Werkzeug zur prosodischen Annotation gesprochener Sprache: GTOBI(S) (Kapitel 3). Die Architektur von GTOBI(S) basiert auf dem TOBI-System (TOBI = *Tones and Break Indices*), das zur Annotation prosodischer Dissoziation (*Breaks*) und zur Transkription der Intonation verschiedener Varianten des Englischen, insbesondere des Amerikanischen Englisch entwickelt wurde. Die tonale Komponente von GTOBI(S) adaptiert Férys Ton-Sequenz-Analyse der deutschen Intonation. Das in Kapitel 4 vorgestellte Resynthesesystem stellt dann sozusagen die Implementierung von GTOBI(S) dar: Auf der Basis von GTOBI(S)-Symbolen werden automatisch Grundfrequenzkonturen generiert und einem Sprachsignal, dessen natürliche segmentale Eigenschaften erhalten bleiben, überlagert. Die Anwendungsmöglichkeiten dieses Systems werden kurz diskutiert.

In Teil III werde ich drei experimentelle Untersuchungen vorstellen. Die erste Studie (Kapitel 5) beschäftigt sich mit lokalen intonatorischen Phänomenen und ihrem Beitrag zur Disambiguierung eines Adverbs auf der Satzebene. Behandelt werden die drei Lesarten des Adverbs *erst* und ihre distinktiven prosodischen Korrelate. Die zweite Studie (Kapitel 6) geht über die Satzebene hinaus, sie hat die Interpretation von anaphorischen Einheiten in Zwei-Satz-Diskursen zum Gegenstand und problematisiert den ›lokalen

Charakter‹ lokaler intonatorischer Phänomene. Den ersten Teil bildet ein Perzeptionsexperiment, das die Hypothese, daß allein das Vorhandensein eines Akzents die referentielle Präferenz eines Pronomens beeinflussen kann, überprüft. Zur Erklärung des negativen Ergebnisses dieses Experiments werden dann in einem zweiten Teil Produktionsdaten analysiert. Diese Untersuchung zeigt, daß neben der Lokalisation eines Akzents auch dessen relative Prominenz zu berücksichtigen ist. Die dritte Studie schließlich behandelt die diskurssemantische Strukturierung größerer Diskurseinheiten und deren globalen intonatorischen Korrelate (Kapitel 7). Neben der Pausendauer steht dabei die Analyse der Tonregister im Mittelpunkt. Zur Repräsentation von diskursrelevanten Registermerkmalen wird ein 2-stufiges autosegmentales Modell vorgeschlagen.

Teil I

Prosodie und Intonation

1 Grundlagen

1.1 Suprasegmentale Phonologie

Die suprasegmentale Phonologie beschreibt Phänomene und Prozesse der gesprochenen Sprache, deren Domäne nicht das Segment bzw. das Phonem oder Phon ist. Die Domänen der suprasegmentalen Phonologie sind z.B. die Silbe, das phonologische Wort, die phonologische Phrase etc., in der Regel also Domänen, die größer sind als das Segment. Verschiedene Autoren haben Modelle für eine hierarchische Organisation dieser Einheiten vorgeschlagen (Selkirk, 1984, 1990; Nespor & Vogel, 1986; Hayes, 1989), die sich zwar geringfügig darin unterscheiden, welche Einheiten in die Hierarchie aufgenommen werden, nicht aber darin, an welcher Stelle innerhalb der Hierarchie eine Einheit repräsentiert ist. Ein Beispiel für ein solches Modell phonologischer Konstituenten ist in (1) dargestellt.

(1) Phonologische Äußerung (PU)
↓
Intonationsphrase (IP)
↓
Phonologische/Intermediäre Phrase (ip)
↓
Klitische Gruppe (CG)
↓
Phonologisches Wort (PW)
↓
Fuß (Ft)
↓
Silbe (σ)
↓
Mora (μ)

Die meisten der Modelle, die eine solche hierarchische Organisation annehmen, gehen davon aus, daß jede Ebene vollständig von der nächst höheren Ebene dominiert wird. Selkirk (1990: 180) formuliert diese Hypothese wie folgt:

(2) *Strict Layer Hypothesis*
$P^n \rightarrow P^{n-1}*$ (X* means ‚one or more X's‘)

Dies bedeutet z.B., daß der linke Rand der am weitesten links stehenden intermediären Phrase mit dem linken Rand der dominierenden Intonationsphrase übereinstimmt und entsprechend der rechte Rand der am weitesten rechts stehenden ip mit dem rechten Rand der dominierenden IP:

(3)
$(\qquad\qquad\qquad\qquad)_{IP}$
$(\quad)_{ip} (\quad)_{ip} (\quad)_{ip}$

Außerdem werden durch dieses Prinzip rekursive Strukturen wie in (4) ausgeschlossen.

(4)
$(\qquad\qquad\qquad\qquad)_{IP}$
$(\qquad\qquad)_{IP} (\quad)_{IP}$
$(\quad)_{ip} (\quad)_{ip} (\quad)_{ip}$

Für alle Konstituenten des hierarchischen Modells in (1) wurden von den oben genannten Autoren phonologische Prozesse beschrieben, die die Aufnahme der einzelnen Konstituenten in ein phonologisches Modell rechtfertigen. Einige der Konstituenten sind jedoch nach wie vor umstritten, ebenso wie auch die *Strict Layer Hypothesis* nicht allgemein anerkannt ist (Hyman, Katamba & Walusimbi, 1987; Ladd, 1996). Ich werde im weiteren Verlauf dieser Arbeit nur die Intonationsphrase und die intermediäre Phrase näher behandeln, für eine generelle Diskussion des hierarchischen prosodischen Modells und seines Inventars verweise ich auf Roca (1994: 191ff).

Eine weitere Untergliederung der suprasegmentalen phonologischen Konstituenten und Prozesse, die weithin anerkannt ist, ist die Zuordnung zur lexikalischen oder zur postlexikalischen Ebene der Repräsentation. Vereinfacht gesagt sind lexikalische Prozesse solche, die zur Formierung der phonologischen Repräsentation eines Lexems beitragen, ohne die Einflüsse der phrasalen Komposition, z.B. der syntaktischen Repräsentation zu berücksichtigen. Hierzu zählen beispielsweise die Konstruktion der Fußstruktur, die Silbifizierung und die Zuweisung des Wortakzents. Postlexikalische Prozesse betreffen dagegen die Zuweisung des Satzakzents, die phonologische Phrasierung und die Zuweisung von Phrasengrenzmerkmalen, d.h. auf dieser Ebene sind Informationen aus anderen Komponenten der linguistischen Strukturierung involviert, wie z.B. die syntaktische oder diskursstrukturelle Repräsentation. Diese Unterscheidung in lexikalische und postlexikalische Prozesse ist auch für die Untersuchung der Prosodie-Semantik-Schnittstelle von Bedeutung. Die semantische Relevanz von lexikalischen phonologischen Prozessen ist beschränkt auf die Domäne der lexikalischen Semantik: Eine Segmentkette mit einer bestimmten Silbenstruktur und einer bestimmten metrischen Struktur (welche den Wortakzent repräsentiert) identifiziert ein bestimmtes Lexem aus der Menge der verfügbaren Lexeme – *Panne* und *Tanne* unterscheiden sich aufgrund segmentaler Merkmale, *úmfahren* und *umfáhren* unterscheiden sich aufgrund ihrer metrischen Struktur etc. Postlexikalische phonologische Merkmale beziehen sich dagegen auf Phrasen oder ganze Äußerungen, mithin ist auch ihre Bedeutung (im weitesten Sinne) in der Domäne der Satz- bzw. Diskurssemantik oder in der Pragmatik zu suchen. Postlexikalische Merkmale dienen z.B. dem Ausdruck der Informationsstruktur, der Fokus-Hintergrund-Gliederung oder des Satzmodus, korrelieren also mit höheren Ebenen der linguistischen Organisation von Äußerungen. Aufgrund der Zielsetzung dieser Arbeit, insbesondere Aspekte des Zusammenspiels von Prosodie und Diskursstruktur zu untersuchen, wird im weiteren Verlauf ausschließlich die postlexikalische Komponente der suprasegmentalen Phonologie behandelt.

Zur Definition postlexikalischer phonologischer Domänen werden sowohl phonologische (Nespor & Vogel, 1986; Hayes, 1989), als auch syntaktische und semantische Kriterien herangezogen, wobei neben der phonologischen Evidenz meist die Orientierung an

der syntaktischen Struktur überwiegt (Selkirk, 1984; Cinque, 1993; Hirst, 1993). Semantische Kriterien berücksichtigen in der Regel selten mehr als die Satzsemantik (logische Form) und die Informationsstruktur eines Satzes; diskurssemantische und pragmatische Gesichtspunkte werden ignoriert. Ein Beispiel: Beckman & Pierrehumbert (1986) behandeln ausführlich die phonologische Evidenz für die Annahme von intermediären Phrasen und Intonationsphrasen sowie deren Koinzidenz mit syntaktischen Konstituenten und der Fokus-Hintergrund-Gliederung eines Satzes. Andererseits berichten sie von Untersuchungen (Hirschberg & Pierrehumbert, 1986; Umeda, 1982), die darauf hinweisen, daß Prozesse wie *final lowering*, F_0-Deklination und Anpassung des Stimmumfangs nicht, wie ursprünglich angenommen, an phonologische Domänen von der Größe eines Satzes oder einer phonologischen Äußerung gebunden sind, sondern mit Konstituenten der Diskursstruktur korrelieren. Daraus leiten Beckman & Pierrehumbert die folgenden zwei Schlußfolgerungen ab: 1.) Aufgrund der Korrelation mit der Diskursstruktur ist es nicht plausibel, als Domäne für diese Prozesse eine (neu zu definierende) phrasale Einheit der phonologischen Konstituentenstruktur anzunehmen und 2.) aufgrund des zweifelhaften phonologischen Status ist es sehr wahrscheinlich, daß diese Prozesse den paralinguistischen Phänomenen zuzurechnen sind. Die angesprochenen Prozesse werde ich in den folgenden Absätzen ausführlicher diskutieren, hier geht es mir zunächst um eine Kritik dieses sehr Syntax-orientierten Phonologieansatzes.

Zu Punkt 1.): Die Annahme einer phonologischen Domäne, die mithilfe von Kriterien definiert wird, welche aus der Diskurssemantik abgeleitet sind, ist nicht *per se* ausgeschlossen. Natürlich kann die Notwendigkeit einer solchen phonologischen Kategorie angezweifelt werden, wenn keine unabhängige phonologische Evidenz für ihre Existenz gefunden werden kann. Es ist jedoch nicht einzusehen, weshalb neben Syntax-basierten Definitionen (vgl. Definition der Intonationsphrase in Nespor & Vogel, 1986: 189) nicht auch diskurssemantisch basierte Definitionen erlaubt sein sollen.

Zu Punkt 2.): Der Zweifel am phonologischen Status der angesprochenen Prozesse scheint mir verfrüht angesichts der Tatsache, daß über die Funktion und die phonetische

Ausprägung dieser Prozesse bisher nur sehr wenig bekannt ist. Selbst wenn man die Notwendigkeit einer eigens für diese Prozesse definierten phrasalen phonologischen Konstituente in Frage stellt, so hat doch der völlige Verzicht auf eine phonologische Ebene der Beschreibung Konsequenzen, die nicht empirisch evaluiert sind. Wenn man die kategoriale phonologische Beschreibungskomponente ausschließt, bedeutet dies, daß man direkt von der diskurssemantischen Repräsentation (oder anderen höheren Ebenen der linguistischen Organisation) einer Äußerung auf bestimmte kontinuierliche akustische Parameter zurückschließt. Eine Interaktion der nicht-phonologischen Prozesse mit phonologischen Prozessen und Strukturen wäre damit ausgeschlossen. Dies führt zu der Voraussage, daß die Ausprägung bestimmter akustischer Parameter invariant ist und direkt von einer höheren Ebene der Äußerungsorganisation gesteuert wird, vollkommen unbeeinflußt von der phonologischen Struktur der Äußerung. Zumindest für die oben angesprochenen Parameter ist die Annahme einer solchen direkten Beziehung fragwürdig. Meines Wissens ist es z.B. keineswegs nachgewiesen, daß das Ausmaß von *final lowering* unabhängig ist von der (phonologisch determinierten) tonalen Struktur einer Äußerung, oder daß die Anpassung und Variation des Stimmumfangs unabhängig von der phonologischen Phrasierung beschrieben werden kann. Ich werde in Kapitel 7 auf diesen Punkt zurückkommen und versuchen zu zeigen, daß die Analyse der diskursstrukturell gesteuerten Variation des Stimmumfangs mit dem Instrumentarium der Registerphonologie (Clements, 1981) einen vielversprechenden Ansatz darstellt und durchaus in ein phonologisches Modell der Intonation integriert werden sollte.

Ich möchte diese Diskussion, die in erster Linie dazu dienen soll, die von mir in dieser Arbeit vertretenen Standpunkte zu klären, mit einer weiteren kritischen Anmerkung zur zweiten Schlußfolgerung von Beckman & Pierrehumbert abschließen. Dies betrifft die Unterscheidung von im weitesten Sinne sprachlichen Phänomenen in linguistische, paralinguistische und extralinguistische Phänomene. Die meisten Autoren, die auf diese Einteilung zurückgreifen (es sind dies meist Autoren, die postlexikalische, suprasegmentale Phänomene beschreiben), betonen, daß die Grenzen zwischen diesen drei Klassen

fließend sind (z.B. Cruttenden, 1986: 177ff; Clark & Yallop, 1995: 328ff; Ladd, 1996: 33ff).

> „We can take prosody to be a continuum of functions and effects, ranging from the nonlinguistic or extralinguistic at one end, through the paralinguistic, to the essentially linguistic. At the nonlinguistic end, for example, are features of voice quality that reflect the nature of the speaker's larynx and vocal tract; at the linguistic end are features such as stress and tone, which are functional within specific linguistic systems […]. But note that the term PARALINGUISTIC points to a grey area in between the two reasonably uncontroversial extremes“ (Clark & Yallop, 1995: 329).

Paralinguistisch ist zum Beispiel der lautsprachliche Ausdruck von Emotion (Angst, Überraschung, Ärger etc.) oder von anderen grundlegenden Aspekten der interpersonellen Kommunikation wie Aggression, Beschwichtigung, Spott, Grad des Engagement etc. Typischerweise werden solche kontinuierlich variierenden Zustände und Einstellungen des Sprecher direkt durch die kontinuierliche und globale Variation akustischer Parameter wie Sprechgeschwindigkeit, Lautheit und Stimmqualität ausgedrückt (Ladd, 1996: 8, 33ff). Paralinguistische Merkmale werden also vollkommen unabhängig von der grammatischen Struktur realisiert und infolgedessen auch nicht im Rahmen formal-linguistischer Modelle behandelt.

Der Standpunkt, der in dieser Arbeit vertreten wird und der durch die in Teil III vorgestellten Experimente untermauert wird, ist der, daß lautsprachliche Phänomene, die von der diskursstrukturellen Repräsentation einer Äußerung regelhaft kontrolliert werden, essentiell linguistische Phänomene sind und durchaus mit formal-linguistischen Mitteln analysiert werden können und sollen. Wenn also der Zusammenhang zwischen Diskursstruktur und Prozessen wie *final lowering* tatsächlich besteht, so soll mit der vorliegenden Arbeit dafür argumentiert werden, daß solche Prozesse als linguistische (phonologische) Kategorien in einem integrierten formalen Grammatikmodell behandelt werden und nicht durch die Klassifikation als paralinguistisch von der linguistischen Untersuchung ausgeschlossen werden.

Nach diesem kurzen Überblick über einige Aspekte der suprasegmentalen Phonologie, werde ich im nächsten Abschnitt die Grundlagen der Phonetik suprasegmentaler Phänomene behandeln. Mit der Intonation als einem Teilgebiet der suprasegmentalen Phonologie und Phonetik werde ich dann im Abschnitt 1.3 die eigentliche phonologisch-phonetische Basis dieser Arbeit vorstellen.

1.2 Prosodie, Phonetik und Prominenz

Der klassische Gegenstand der phonetischen Prosodie ist die Untersuchung der folgenden akustischen Parameter: Grundfrequenz (F_0), Intensität und Dauer (Cruttenden, 1986; Neppert & Pétursson, 1986; Clark & Yallop, 1995; Ladd, 1996). Außer diesen drei zentralen Parametern wird in der Regel auch die Stimmqualität, also das laryngale Setting (Phonationstyp, Marasek, 1997) und das supralaryngale Setting zu den prosodischen Parametern gezählt (Laver, 1980; Clark & Yallop, 1995). Grundfrequenz, Intensität und Dauer sind dynamische Parameter, d.h. ihre Ausprägung im Sprachsignal kann sich sehr rasch ändern, die Stimmqualität ist dagegen ein statischer Parameter, der über längere Zeit konstant bleibt.

> „These long-term settings [voice quality and vocal tract state] are the underlying articulatory positions or postures upon which all the dynamics of articulation – both segmental and suprasegmental – are superimposed. The settings have articulatory – and hence acoustic – consequences which pervade the whole stream of speech.“ (Clark & Yallop, 1995: 331).

Aber auch die drei dynamischen Parameter können Ausdruck von eher statischen Merkmalen gesprochener Sprache sein. Die generelle Lautstärke einer Äußerung ist z.B. ein Merkmal der Intensität, die Sprechgeschwindigkeit ist ein Merkmal des Parameters Dauer und der Stimmumfang einer Äußerung korreliert mit der Grundfrequenz. In Analogie zu der Unterscheidung dynamisch vs. statisch verwende ich in dieser Arbeit die Begriffe LOKAL und GLOBAL. Lokale prosodische Merkmale werden innerhalb der Grenzen ‚kleiner' linguistischer Einheiten (Wörter, Silben, Segmente) durch die schnelle Veränderung der Intensität, Dauer und/oder Grundfrequenz realisiert, während globale Merkmale auf Phrasen- oder Äußerungsebene realisiert werden, z.B. durch die Wahl des Tonregisters. Lokale Merkmale finden ihren akustischen Ausdruck also in der raschen Variation der dynamischen Parameter, globale Merkmale in der langsamen, kontinuierlichen Variation der dynamischen Parameter oder der Stimmqualität.

Die Ausprägung der Parameter Grundfrequenz, Intensität und Dauer im Sprachsignal reflektiert jedoch nicht nur suprasegmentale Phänomene, sondern interagiert auch sehr stark mit der segmentalen Struktur einer Äußerung. Die Intensität bzw. ihr psychoakustisches Korrelat Lautheit (*loudness*) wird beeinflußt von der Sonorität eines Segments oder einer Sequenz von Segmenten. Clark & Yallop (1995: 335) geben das folgende Beispiel: In der englischen CVC-Silbe *shack* ist der Vokal, relativ zum konsonantischen Kontext, sonorer als der Vokal in z.B. *wool*. Deshalb wird der Vokal in *shack* als lauter wahrgenommen als der Vokal in *wool*. Diese Wahrnehmung ist zurückzuführen auf die segmentale Struktur und nicht auf eine unabhängig kontrollierte Variation der Intensität.

Die Dauer wird ebenfalls von segmentalen Eigenschaften bzw. von Eigenschaften der Segmentstruktur beeinflußt. So unterscheiden sich z.B. im Deutschen gespannte und ungespannte Vokale in ihrer Dauer (Jessen, Marasek, Schneider & Claßen, 1995). Stimmlose Frikative sind länger als die entsprechenden stimmhaften Frikative (Klatt, 1976). Klatt (1976) berichtet außerdem von zahlreichen Effekten der Segmentdauer und ihrer Abhängigkeit vom segmentalen Kontext, z.B. ist ein Vokal vor einem stimmlosen Konsonant kürzer als vor einem stimmhaften Konsonant und Konsonanten in C-Clustern sind tendenziell kürzer als Konsonanten in einer CV-Silbe. Dauer als eine Eigenschaft linguistischer Einheiten (Segmentdauer, Silbendauer, Dauer einer Phrase) kann nicht getrennt von Einflüssen genereller Timing-Phänomene in der Sprachproduktion betrachtet werden. Neben offensichtlichen Faktoren wie der Sprechgeschwindigkeit beeinflussen z.B. auch emotive und andere psychologische und physiologische Faktoren das artikulatorische Timing und mithin die Dauer linguistischer Einheiten (Williams & Stevens, 1972; Scharf, i.V.).

Die Grundfrequenz unterliegt ebenfalls paradigmatischen und syntagmatischen segmentalen Einflüssen. Ein typischer paradigmatischer Effekt, der in zahlreichen Experimenten nachgewiesen wurde, ist *intrinsic pitch*. Ohala (1978), Ladd & Silverman (1984) und andere haben gezeigt, daß hohe Vokale (z.B. /i/) mit höherer Grundfrequenz produziert werden als tiefe Vokale (z.B. /a/)[1]. Ein kurzer Überblick über die verschiedenen Hypothesen, die diesen Effekt erklären sollen, findet sich in Clark & Yallop (1995: 337). Ein Effekt der Interaktion zwischen Grundfrequenz und segmentaler Struktur ist zum Beispiel der Einfluß von Verschlußlauten auf die F_0-Kontur nachfolgender Vokale. Ein stimmloser Verschlußlaut verursacht eine Anhebung der Grundfrequenz im nachfolgenden Vokal, während ein stimmhafter Verschlußlaut den umgekehrten Effekt bewirkt, die Absenkung der Grundfrequenz (Hombert, 1978). Auch postvokalische Konsonanten können die Ausprägung der Grundfrequenz beeinflussen: vor einem glottalen Verschlußlaut werden höhere F_0-Werte gemessen, vor einem glottalen Frikativ /h/ dagegen tiefere (Clark & Yallop, 1995: 336f).

Mit den Parametern der prosodischen Phonetik werden zahlreiche, linguistisch mehr oder weniger relevante Merkmale der gesprochenen Sprache beschrieben. Tabelle 1-1 gibt einen Überblick über die phonetischen Parameter und einige der mit ihnen verknüpften lautsprachlichen Merkmale.

[1] Der Effekt konnte tatsächlich durch die akustische Analyse des Sprachsignals nachgewiesen werden, insofern ist der psychoakustische Begriff *pitch* mißverständlich. Ich verwende dennoch die Bezeichnung *intrinsic pitch*, da sie in der einschlägigen Literatur üblich ist.

Tabelle 1-1: Phonetische Parameter und lautsprachliche Merkmale

Intensität	*lokal*: rasche Variation der Intensität von Segment zu Segment bzw. von Silbe zu Silbe *global*: genereller Lautstärkeumfang
Dauer	*lokal*: Dehnung bzw. Reduktion von Segmenten/Silben; Pausendauer *global*: Sprechgeschwindigkeit; Rhythmus
Grundfrequenz	*lokal*: rasche Variation der Grundfrequenz (Pitchakzente, Terminaltöne) *global*: Variation des Tonregisters; genereller Stimmumfang, Sprechstimmlage; Deklination

Eine der Hauptaufgaben der linguistischen Prosodie ist der Ausdruck von Prominenz, also die Hervorhebung und Auszeichnung linguistischer Einheiten im lautsprachlichen Kontinuum. Sehr viele der in Tabelle 1-1 aufgeführten Merkmale, insbesondere die lokalen Merkmale, tragen zu dieser Funktion bei. Man unterscheidet im wesentlichen zwei Arten bzw. Level von Prominenz: Prominenz innerhalb von Lexemen und Komposita und Prominenz innerhalb von Phrasen. Der erste Typ, der in der englischsprachigen Literatur (nicht immer ganz einheitlich) als *stress* bezeichnet wird, wird im weiteren Verlauf dieser Arbeit als WORTBETONUNG bezeichnet. Es existieren zahlreiche phonologische Modelle zur Bestimmung der Prominenzverhältnisse innerhalb von Wörtern (Halle & Vergnaud, 1987; Kager, 1995; Halle & Idsardi, 1995). Allen Modellen gemeinsam ist das Konzept der relativen Prominenz, d.h. die Prominenz einer Silbe innerhalb eines Wortes wird immer in Relation zur Prominenz der übrigen Silben des selben Wortes bestimmt; eine Silbe ist entweder prominenter oder weniger prominent als benachbarte Silben. Das Konzept der relativen Prominenz spiegelt sich auch in den phonetischen Korrelaten der Wortbetonung wider. Eine prominente Silbe ist z.B. nicht an ihrer absoluten Dauer zu erkennen, sondern am Verhältnis ihrer Dauer zu den Silbendauern benachbarter Silben.

In der phonetischen Realisierung des Wortakzents unterscheiden sich die Sprachen der Welt (Neppert & Pétursson, 1986: 145). Als stabilstes phonetisches Korrelat der

Wortbetonung im Deutschen gilt nicht, wie häufig angenommen, die Intensität, sondern die Dauer, d.h. betonte Silben bzw. der vokalische Nukleus betonter Silben ist länger als der Nukleus unbetonter Silben (Jessen, 1994; Rapp, 1994; Jessen *et al.*, 1995; Dogil, 1995). Neuere Arbeiten zeigen jedoch durch die differenziertere Untersuchung von Intensitätsmerkmalen, daß auch dem Faktor Intensität eine wichtige Rolle bei der Realisierung der Wortbetonung zukommt (Sluijter & van Heuven, 1994; Sluijter, 1995; Claßen, 1996): Betonte und unbetonte Vokale unterscheiden sich weniger in ihrer Gesamtintensität, als vielmehr darin, wie die Energie in verschiedenen Frequenzbändern verteilt ist. Betonte Vokale haben einen geringeren *spectral tilt*, d.h. größere Intensitäten in den mittleren und hohen Frequenzen des Spektrums als unbetonte Vokale mit geringeren Intensitäten in diesen Frequenzbereichen und damit stärkerem *spectral tilt*.

Prominenz auf der Phrasenebene, englisch *accent*, bezeichne ich im folgenden als AKZENT. Häufig findet man die Bezeichnung SATZAKZENT, da Prominenz auf dieser Ebene Informationen übermittelt, die, basierend auf der syntaktischen Struktur, von höheren Modulen der linguistischen Organisation interpretiert werden (Syntax, Semantik, Pragmatik). Akzente spielen z.B. eine wichtige Rolle bei der Konstitution der Informationsstruktur eines Satzes, also bei der Gliederung in Topik, Fokus und Hintergrund (Uhmann, 1991a,b; Büring, 1995; Kuhn, 1996). Die für die Akzentzuweisung relevante phonologische Domäne ist allerdings die Intonationsphrase, deren Grenzen nicht immer mit den Grenzen syntaktisch definierter Phrasen übereinstimmen (s. Kapitel 2). Die Bezeichnung Phrasenakzent ist jedoch nicht üblich und wäre mißverständlich, da der englische Begriff *phrase accent* kein Prominenzphänomen bezeichnet, sondern ein intonatorisches Merkmal, welches Phrasengrenzen signalisiert (Terminalton). Eine weitere Bezeichnung, die auf die in vielen Sprachen übliche phonetische Realisierung verweist, ist der Begriff PITCHAKZENT (*pitch accent*). In allen Intonationssprachen ist das wichtigste Korrelat von Prominenz auf der Phrasenebene die Variation der Grundfrequenz, d.h. eine linguistische Einheit wird dann als akzentuiert wahrgenommen, wenn sie mit bestimmten, salienten Ereignissen der F_0-Kontur assoziiert ist. Saliente Ereignisse der

F_0-Kontur in diesem Sinne sind entweder Leveltöne (besonders hohe oder besonders tiefe Töne) oder Konturtöne (fallend oder steigend), also lokale Merkmale der Grundfrequenzkontur.

Das Konzept der relativen Prominenz ist auf der Phrasenebene weniger gut ausgearbeitet als auf der Wortebene. Abgesehen von der grundlegenden Relation zwischen akzentuierten und nicht-akzentuierten Einheiten, die selbstverständlich in allen Modellen präsent ist, steht die weitergehende Analyse und Integration dieses Konzepts in die Modelle der phrasalen Prominenz noch aus. Daß es Prominenzunterschiede auch zwischen verschiedenen (pitch-) akzentuierten Einheiten innerhalb der selben Phrase gibt, konnte in verschiedenen Experimenten nachgewiesen werden (Liberman & Pierrehumbert, 1984; Rietveld & Gussenhoven, 1985; Ladd, Verhoeven & Jacobs, 1994; Bartels & Kingston, 1994). Zur Integration solcher Prominenzrelationen zwischen benachbarten Pitchakzenten in ein phonologisches Modell, also zur metrischen Phonologie der phrasalen Prominenz, gibt es jedoch bisher nur sehr wenige Ansätze (siehe z.B. Ladd, 1990; Clements, 1990). Ein Grund hierfür mag sein, daß über den Status solcher gradueller Prominenzunterschiede zwischen Pitchakzenten kein Konsens herrscht. Die offene Frage lautet, ob dieses Phänomen ‚nur' paralinguistischen Status hat und z.B. unterschiedliche Grade der Emphase ausdrückt (Liberman & Pierrehumbert, 1984), oder ob solche Relationen tatsächlich relevant sind für die diskurssemantische Interpretation einer Äußerung (siehe Kapitel 6).

Beide Ebenen der Prominenz, die hier angesprochen wurden, also Wortbetonung und Satzakzent, können in den germanischen Sprachen relativ unabhängig voneinander betrachtet werden (Beckman, 1986; Beckman & Edwards, 1994; Möhler & Dogil, 1995), d.h. eine (wort-) betonte Silbe ist notwendigerweise länger als eine unbetonte Silbe (zumindest im Deutschen), eine (satz-) akzentuierte Silbe trägt dagegen notwendigerweise einen Pitchakzent, muß aber nicht gelängt sein (eine Kritik dieser radikalen Unabhängigkeitshypothese findet sich in Ladd, 1996: 221ff). In den folgenden Kapiteln wird die phrasale Komponente der Prominenz im Mittelpunkt stehen, da insbesondere

Aspekte des Zusammenspiels von Diskurssemantik und Prosodie behandelt werden sollen. Infolge des tonalen Charakters von Pitchakzenten – dem wichtigsten Ausdrucksmittel der phrasalen Prominenz – und des tonalen Charakters anderer diskursrelevanter prosodischer Merkmale, werde ich die *intonatorischen* Kategorien des Deutschen in das Zentrum meiner Analyse stellen.

1.3 Intonation

1.3.1 Begriffsbestimmung

Dieser Abschnitt dient ausschließlich dazu, die von mir in dieser Arbeit bevorzugte Verwendungsweise des Begriffs INTONATION zu verdeutlichen; ich verzichte auf eine Diskussion der (zahlreich vorhandenen) anderen Möglichkeiten, den Begriff zu gebrauchen. Eine kritische Auseinandersetzung mit der Terminologie im Bereich der suprasegmentalen Phänomene ist angesichts der zahlreichen Ambiguitäten und idiosynkratischen Verwendungsweisen zwar wünschenswert, kann aber nicht Teil dieser Arbeit sein. Angesichts des Schwerpunktes der vorliegenden Untersuchung, scheint es mir angebracht und ausreichend, die verwendete Terminologie in meinem Sinne zu definieren, zumal die von mir gewählte Verwendungsweise durchaus auch von anderen Autoren bevorzugt wird.

Intonation beschreibt postlexikalische, suprasegmentale Phänomene der Lautsprache, die primär durch die Variation der Grundfrequenz ausgedrückt werden.

Was unter suprasegmentalen Phänomenen zu verstehen ist, sollte in den vorangegangen Abschnitten deutlich geworden sein. Die Unterscheidung lexikalisch vs. postlexikalisch wurde ebenfalls bereits diskutiert, ich möchte jedoch zwei Punkte nochmals betonen: Postlexikalisch bedeutet erstens, daß intonatorische Merkmale von *höheren* Ebenen der linguistischen Organisation – Syntax, Semantik, Pragmatik – kontrolliert werden, nicht von der lexikalischen Phonologie, und postlexikalisch bedeutet zweitens, daß intonatorische Merkmale von höheren Ebenen der *linguistischen* Organisation kontrolliert

werden, daß also para- und extralinguistische Effekte von der intonatorischen Beschreibung ausgeschlossen sind.

Mit dem dritten Aspekt der Definition – Primat der Grundfrequenzmerkmale – sollen die anderen prosodischen Parameter – Dauer und Intensität – nicht völlig aus der vorliegenden Analyse eliminiert werden. Ihnen wird jedoch ein sekundärer Status zugeschrieben, insofern als die wesentlichen Kategorien, mit denen im folgenden gearbeitet wird, tonaler Natur sind und somit primär mit Grundfrequenzmerkmalen korrelieren, während Dauer- und Intensitätsmerkmale nur unterstützend zur Identifikation dieser Kategorien herangezogen werden.

Intonation im oben beschriebenen Sinne wird von allen Sprachen der Welt verwendet. Hinsichtlich der Komplexität intonatorischer Systeme gibt es jedoch große Unterschiede zwischen den Sprachen. Die Hauptursache für unterschiedlich komplexe Intonationssysteme liegt in der sprachspezifischen linguistischen Funktionalisierung von Tonmerkmalen. Die zwei Hauptklassen, die sich entlang dieser Unterscheidung bilden lassen, sind Tonsprachen auf der einen Seite und sogenannte Intonationssprachen auf der andern Seite. In Tonsprachen ist Ton ein Merkmal des phonologischen Lexikons, Töne sind ebenso bedeutungsunterscheidend wie Phoneme. Intonatorische Variationen der tonalen Struktur einer Äußerung sind in diesen Sprachen sehr stark eingeschränkt, aufgrund der Gefahr, dadurch die lexikalischen Töne zu verdecken. In Sprachen dagegen, die tonale Merkmale nicht lexikalisiert haben, sind komplexere intonatorische Systeme möglich, da kein anderes linguistisches System die tonalen Strukturen von Äußerungen kontrolliert. Zu diesen sogenannten Intonationssprachen zählen fast alle europäischen Sprachen, darunter z.B. das Englische und das Deutsche. Es gibt jedoch auch unter den Intonationssprachen kleinere Unterschiede hinsichtlich der Komplexität des Intonationssystems. Das Intonationssystem des Deutschen gilt z.B. als ärmer im Vergleich zum englischen Intonationssystem (vgl. Féry, 1993 und Abschnitt 2.2). Einer der Gründe hierfür mag sein, daß das Deutsche über ein reicheres Inventar an Gradpartikeln verfügt, so daß differenzierte Bedeutungsunterschiede, die in anderen Sprachen bevorzugt mit intonatorischen Mitteln ausgedrückt werden, im Deutschen auch mit anderen linguistischen Mitteln ausgedrückt werden können.

1.3.2 Intonatorische Phänomene

Intonatorische Merkmale können nach drei funktionalen Kategorien untergliedert werden: Prominenzmerkmale, Phrasierungsmerkmale und Phrasenmerkmale.

PROMINENZ — Die zentralen intonatorischen Merkmale zur Realisierung von Prominenz sind Pitchakzente, also lokale saliente Ereignisse im Tonhöhenverlauf einer Äußerung. Dies können entweder ‚einfache' tonale Ereignisse sein – Gipfel (lokale Maxima) bzw. Täler (lokale Minima) – oder komplexe Ereignisse – fallende bzw. steigende Tonhöhenverläufe. Eine der Grundannahmen, die dieser Analyse von Pitchakzenten zugrundeliegt, ist die, daß die Differenz zwischen einem relativ hohen Pitch-Level und einem relativ tiefen Pitch-Level den elementarsten Aspekt von Grundfrequenzvariationen[2] darstellt und daß die Sprachmelodie im wesentlichen durch die Alternation relativ hoher und relativ tiefer Pitch-Level charakterisiert ist (Collier & 't Hart, 1981, nach der Übersetzung von Ladd, 1996: 15). Alle phonologischen Modelle der Intonation unterscheiden verschiedene Typen von Pitchakzenten[3]. Diese Unterscheidung beruht primär auf der Form der lokalen Pitch-Kontur, mit der ein bestimmter Akzenttyp normalerweise realisiert wird. Voraussetzung hierfür ist eine gewisse Abstraktion von der tatsächlichen Realisierung von Pitchakzenten, also von der tatsächlich beobachtbaren F_0-Kontur. Diese Idealisierung betrifft sowohl mikroprosodische Interferenzen – segmentale Einflüsse,

[2] Ich verwende den akustischen Begriff GRUNDFREQUENZ und den psychoakustischen Begriff PITCH synonym, da die Unterscheidung dieser beiden Beschreibungsebenen in meiner Analyse keine Rolle spielt. In der englischsprachigen Literatur wird bei der Beschreibung intonatorischer Phänomene in der Regel der Begriff *pitch* verwendet, den ich einer mehr oder weniger gut passenden deutschen Übersetzung – ‚wahrgenommene Tonhöhe' o.ä. – vorziehen möchte.

[3] Der Begriff *pitch accent* ist zurückzuführen auf Bolinger (1958) und wird auch in neueren Intonationsphonologien verwendet, z.B. im Tonsequenzmodell (siehe Kapitel 2). Andere Modelle, z.B. das IPO-Modell ('t Hart, Collier & Cohen, 1990), verwenden zwar eine andere Terminologie, teilen jedoch das zugrundeliegende Konzept prominenzsensitiver lokaler Pitch-Ereignisse.

wie sie in Abschnitt 1.2 beschrieben wurden – als auch globale Settings, die die konkrete Ausprägung von Pitchakzenten beeinflussen können. Hier sind insbesondere Merkmale des Stimmumfangs (*pitch range*) zu nennen: bei großem Stimmumfang sind z.B. Tonhöhenbewegungen sehr viel ausgeprägter als bei geringem Stimmumfang. Ein phonologisches Modell intonatorischer Prominenz besteht also aus mindestens zwei Komponenten: 1.) einem Inventar idealisierter und kontrastiver Kategorien – Pitchakzente – und 2.) einem Satz mehr oder weniger expliziter phonetischer Realisierungsregeln.

Pitchakzente sind an metrisch starke Einheiten der prosodischen Struktur einer Äußerung gebunden. Die tontragende Einheit (*tone bearing unit*, TBU) in Sprachen wie Englisch, Holländisch oder Deutsch ist die Silbe, im Japanischen ist es die More. Die Beziehung zwischen, einfach ausgedrückt, Intonation und Text läßt sich sehr anschaulich im Rahmen der autosegmentalen Theorie beschreiben. Die autosegmentale Theorie (Goldsmith, 1976, 1990) ist ein Modell der phonologischen Repräsentation, welches die ungeordnete Matrixrepräsentation phonologischer Merkmale, wie sie im klassischen generativen Modell von Chomsky & Halle (1968) vorgeschlagen wurde, ersetzt durch eine multidimensionale Repräsentation voneinander unabhängiger Autosegmente. Dadurch ist es möglich, verschiedene Klassen von Merkmalen auf unterschiedlichen, relativ unabhängigen ‚Schienen' (*tiers*) zu repräsentieren und die Anwendung phonologischer Regeln auf einzelne Schienen zu beschränken. Geht man nun davon aus, daß z.B. Pitchakzente Autosegmente sind, also eine autonome Klasse von Merkmalen, so läßt sich der Zusammenhang zwischen Intonation und Text einfach durch die Assoziation von Pitchakzenten mit anderen Autosegmenten innerhalb der phonologischen Struktur (Silben, Moren etc.) repräsentieren. Auf diese Weise können die wichtigsten Aspekte der intonatorischen Prominenz – Akzenttyp und Akzentposition – konsistent in einem phonologischen Modell analysiert werden. Die tatsächliche Realisierung, d.h. die konkrete Ausprägung des Pitchakzents und Aspekte der temporalen Koordination zwischen Pitchakzent und phonetischem Korrelat der TBU, wird durch phonetische Regeln abgedeckt.

PHRASIERUNG — Die phonologische Phrasierung von Äußerungen ist sehr variabel und kann nicht immer eindeutig anhand prosodischer Merkmale bestimmt werden (Cruttenden, 1986: 35ff; Féry, 1993: 59ff). In der Regel orientiert sich die Phrasierung an der syntaktischen Struktur, d.h. phonologische Phrasengrenzen fallen mit den Grenzen syntaktischer Konstituenten zusammen. Dieser Zusammenhang ist jedoch weder zwingend noch gibt es feste Regeln, welche phrasalen syntaktischen Konstituenten eine phonologische Phrase bilden (für eine Diskussion dieses Zusammenhangs siehe z.B. Hirst, 1993). Es gibt einige prosodische Merkmale, die einzeln oder in unterschiedlichen Kombinationen eine Phrasengrenze markieren können. Die wichtigsten intonatorischen Hinweise auf Phrasengrenzen sind Terminaltöne und abrupte Veränderungen von Merkmalen des Stimmumfangs. Terminal- oder Grenztöne sind lokale F_0-Bewegungen, die keine Prominenz ausdrücken. Vor allem im Britischen und Amerikanischen Englisch werden Phrasengrenzen häufig durch Pitch-Bewegungen markiert (Cruttenden, 1986; Beckman & Pierrehumbert, 1986), im Deutschen findet sich dieses Merkmal seltener, ist jedoch auch Bestandteil des intonatorischen Systems (Féry, 1993). Bei den Merkmalen des Stimmumfangs möchte ich zwei Phänomene unterscheiden: Registersprünge und Rücksetzung der Deklination.

Das Register legt die obere (*topline*) und untere (*baseline*) Begrenzung des Tonhöhenumfangs einer Phrase fest[4]. Es stellt einen Ausschnitt aus dem generellen Stimmumfang dar, der einem Sprecher oder einer Sprecherin zur Verfügung steht. Das Register bestimmt somit u.a. die vertikale Auslenkung von lokalen Pitch-Bewegungen. In Abbildung 1-1 ist dieser Zusammenhang schematisch dargestellt. Das Register wird stets für eine gesamte Phrase festgelegt (Cruttenden, 1986: 54f), insofern kann ein Registerwechsel bzw. Registersprung eine Phrasengrenze signalisieren.

[4] Cruttenden (1986: 54f) unterscheidet *key effects* und *register effects*. Registereffekte sind Verschiebungen von Top- und Baseline nach oben oder unten, ohne den Abstand zwischen Top- und Baseline zu verändern. *Key*-Effekte sind dagegen Veränderungen des Abstands zwischen Top- und Baseline. Meine Verwendung von des Begriffs REGISTER deckt beide Phänomene ab.

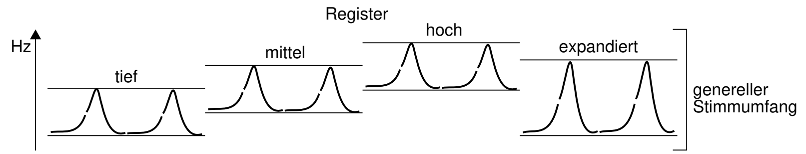


Abbildung 1-1: *Schematische Darstellung von Registermerkmalen.*

Als Deklination (*declination* oder *downtrend*) bezeichnet man das Phänomen, daß die Grundfrequenz im Verlauf einer Äußerung kontinuierlich abfällt, unabhängig von der tonalen Struktur der Äußerung (Ladd, 1984). Andere Modelle gehen davon aus, daß es sich hierbei nicht um einen natürlichen, globalen Abwärtstrend handelt, sondern um einen lokal beschränkten Prozeß – *final lowering* –, der nur die Randstruktur einer Intonationsphrase betrifft und von verschiedenen, mehr oder weniger linguistischen Parametern gesteuert wird (Liberman & Pierrehumbert, 1984; Beckman & Pierrehumbert, 1986). Gussenhoven & Rietveld (1988) schlagen ein Modell vor, welches sowohl globale Tendenzen als auch *final lowering*-Effekte integriert. Unabhängig davon, welche der Analysen bevorzugt wird, kann man jedoch davon ausgehen, daß die Rücksetzung der relevanten F_0-Parameter an eine Phrasengrenze gebunden ist und somit als Hinweis auf die Phrasierung dienen kann.

Neben diesen intonatorischen Merkmalen der Phrasierung sind noch drei wichtige Dauer- bzw. Timingmerkmale zu nennen, die ebenfalls häufig Phrasengrenzen signalisieren: Pausen, Dehnung und *anacrusis*. Vor und nach Intonationsphrasen können Pausen auftreten, je nach Stärke der Phrasengrenze (IP-Grenze oder ip-Grenze; siehe Kapitel 2) von unterschiedlicher Dauer. Dehnung von Silben (*final lengthening*) ist häufig am Ende einer Intonationsphrase zu beobachten (Vaissière, 1983), während *anacrusis* – die Produktion von Silben mit erhöhter Sprechgeschwindigkeit und reduzierter artikulatorischer Genauigkeit – den Anfang einer Intonationsphrase signalisieren kann (Cruttenden, 1986:

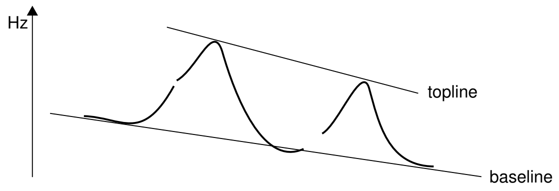


Abbildung 1-2: *Schematische Darstellung einer typischen F_0-Kontur.*

39). Diese drei Merkmale sind jedoch insofern problematisch, als sie häufig auch im Kontext von Hesitationen auftreten können, wobei Hesitationen in der Regel nicht als phrasenbegrenzend sondern als phrasenintern analysiert werden (siehe Kapitel 2).

PHRASENMERKMALE — Während Prominenz- und Phrasierungsmerkmale als etabliert gelten können, ist der Status von Merkmalen, die die globalen Parameter einer Phrase charakterisieren, in phonologischen Modellen der Intonation noch sehr unklar. An erster Stelle unter den tonalen Phrasenmerkmalen ist das Tonregister zu nennen. Wie bereits erwähnt, determiniert ein gewähltes Register die vertikale Auslenkung lokaler Pitch-Ereignisse bzw., in der Terminologie eines Targetmodells wie des Tonsequenzmodells (Kapitel 2), die vertikale Skalierung der hohen und tiefen Pitch-Targets. Die umstrittene Frage lautet, ob Register ein gradueller, paralinguistisch kontrollierter Parameter ist, oder ob Register in Intonationssprachen, analog zur Analyse afrikanischer Tonsprachen, als phonologische Kategorie behandelt werden kann und sollte (Clements, 1981, 1990; Ladd, 1990, 1994). Ich werde auf diese Problematik in Kapitel 7 dieser Arbeit zurückkommen. Ein weiterer Punkt, der, wie bereits kurz angedeutet, ebenfalls kontrovers diskutiert wird, ist der Bereich Deklination, *final lowering* und *downstep*. Abbildung 1-2 exemplifiziert schematisch die relevanten Phänomene, nämlich Pitch-range-Variationen aufgrund der generell absinkenden Tendenz der Grundfrequenz im Verlauf einer Phrase (*top-* und *baseline*-Deklination). Die konkurrierenden Erklärungs-

modelle unterscheiden sich darin, ob sie diese Phänomene 1.) ausschließlich als globale Eigenschaften einer Phrase analysieren (Fujisaki, 1983), oder 2.) ausschließlich als Produkt lokaler Interaktion (*downstep*) bzw. lokal wirkender Kontrolle (*final lowering*) (Liberman & Pierrehumbert, 1984) oder 3.) als eine Kombination aus lokalen und globalen Faktoren. Je nachdem, welchen Ansatz man wählt, sind diese Phänomene entweder als Phrasenmerkmale zu sehen (globale Interpretation) oder als phonologische Regeln der Interaktion zwischen (lokalen) Kategorien zu interpretieren (lokale Interpretation).

Das im nächsten Kapitel vorgestellte Tonsequenzmodell präferiert die lokale Interpretation dieser Phänomene. *Downstep* ist ein zentrales Konzept dieses Modells, und ich werde an entsprechender Stelle auf die hier behandelte Fragestellung zurückkommen und die Argumentation für diesen Standpunkt dort kurz skizzieren.

2 Die Phonologie der Intonation

Im Rahmen der Entwicklung der autosegmentalen Phonologie etablierte sich Ende der 70er Jahre ein explizit phonologischer Ansatz zur Beschreibung der Intonation. Basierend auf den Arbeiten von Liberman (1975) und Bruce (1977) stellte Pierrehumbert (1980) ein umfassendes, den Prinzipien der autosegmentalen Theorie entsprechendes Modell der Intonation des amerikanischen Englisch vor: das Tonsequenzmodell (TSM). Aufgrund der Kompatibilität mit der weitverbreiteten autosegmentalen Theorie entwikkelte sich das TSM rasch zum vorherrschenden Intonationsmodell in der phonologischen Forschung. Im Laufe der 80er und vorallem der 90er Jahre wurden dann auch zahlreiche TSM-Analysen verschiedener Sprachen der Welt veröffentlicht (z.B. Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988 (Japanisch); Hayes & Lahiri, 1991 (Bengali); Uhmann, 1991a; Féry, 1993 (Deutsch); Mennen & den Os, 1993 (Griechisch); Välimaa-Blum, 1993 (Finnisch); Gósy & Terken, 1994 (Ungarisch); Grice, 1995 (Palermo-Italienisch)). Diese Arbeiten verhalfen dem TSM nicht nur zum internationalen Standard, sondern führten auch zu einer Weiterentwicklung und Modifikation des Tonsequenzmodells selbst (insbesondere die vergleichende Analyse der englischen und japanischen Intonation in Beckman & Pierrehumbert, 1986). Vorallem aufgrund der explizit phonologischen Ausrichtung des TSM aber auch aufgrund seines mittlerweile etablierten Status als Standardmodell der phonologischen Intonationstheorie wurde das TSM für die vorliegende Untersuchung als Ausgangspunkt gewählt und anderen, zum Teil eher phonetisch orientierten Modellen – z.B. dem IPO-Modell ('t Hart, Collier & Cohen, 1990), dem Fujisaki-Modell (Fujisaki, 1983) oder KIM (Kohler, 1991a) – vorgezogen. Im ersten Abschnitt dieses Kapitels werde ich die Grundprinzipien des TSM kurz darstellen (eine ausführliche Diskussion findet sich in Ladd, 1996), um dann in Abschnitt 2.2 die TSM-Analyse des Deutschen von Féry (1993) vorzustellen.

2.1 Das Tonsequenzmodell

Das TSM repräsentiert Pitch-Konturen als eine geordnete Sequenz diskreter, intonatorischer Ereignisse. Die ‚Atome' des TSM sind die *phonologischen* Kategorien H und L. H und L sind abstrakte Tonlevel, denen erst während der Transformation von der phonologischen zur phonetischen Repräsentation konkrete F_0-Werte zugewiesen werden. Die Atome H und L können zwei grundsätzlich verschiedene Funktionen übernehmen, die im letzten Kapitel bereits angesprochen wurden: Ausdruck von Prominenz (Pitchakzente) und Kennzeichnung der Phrasierung (Grenztöne). Es ist üblich, Akzenttöne mit einem Asterisk (*) und Grenztöne von Intonationsphrasen (IP) mit einem Prozentzeichen (%) zu markieren. Grenztöne von intermediären Phrasen[5] (ip) werden mit einem Minuszeichen (-) markiert. Das TSM repräsentiert nur linguistisch relevante, also interpretierbare intonatorische Ereignisse. Der Pitch-Verlauf zwischen solchen relevanten Ereignissen (Transition) wird auf der Ebene der phonologischen Repräsentation ignoriert und bleibt – ebenso wie die konkrete Ausprägung der Tonlevel – der Modellierung mithilfe phonetischer Regeln überlassen. Die Assoziation zwischen den linguistisch relevanten intonatorischen Ereignissen und dem Text bzw. seiner phonologischen Struktur erfolgt gemäß den bereits erwähnten Prinzipien der autosegmentalen Theorie.

Eine Intonationsphrase besteht aus mindestens einem Pitchakzent, mindestens einem sog. Phrasenakzent[6] (ip-Grenzton) und genau einem Terminalton (IP-Grenzton). Wie bereits erwähnt sind sowohl Pitchakzente als auch Grenztöne zusammengesetzt aus H- bzw. L-Tönen, wobei nur Pitchakzente bitonal (oder tritonal; siehe Abschnitt 2.2) sein

[5] Intermediäre Phrasen sind IP-Untereinheiten, also eine zusätzliche Ebene der intonatorischen Phrasierung zwischen Intonationsphrase und phonologischem Wort (Beckman & Pierrehumbert, 1986); ihr Status ist umstritten (Ladd, 1996 und Abschnitt 2.2).

[6] Der Phrasenakzent (*phrase accent*) des TSM darf nicht verwechselt werden mit dem *phrase accent* in Bruces (1977) Modell. Während der *phrase accent* von Bruce seinen Namen zurecht trägt – er ist ein Prominenzmerkmal –, ist die Bezeichnung *phrase accent* oder Phrasenakzent im Sinne des TSM ein wenig irreführend, da er kein Prominenzmerkmal ist, sondern eine Grenztonkategorie (vgl. Ladd, 1983).

können, Grenztöne sind stets monotonal. Somit ergibt sich im Standard-TSM (Pierrehumbert, 1980 und Beckman & Pierrehumbert, 1986) die schematische Struktur in (1) für eine Intonationsphrase[7].

(1) $[\mathrm{T^*(T)\ ((T\text{-})\ T^*(T))\ T\text{-}\ T\%}]_{\mathrm{IP}}$

Sowohl für Pitchakzente als auch für die beiden Grenztonkategorien (Phrasenakzent und Terminalton) wird ein invariantes, sprachspezifisches Inventar angenommen. Nach der Modifikation von Pierrehumberts (1980) Originalmodell durch Beckman & Pierrehumbert (1986) stehen für die Analyse der Intonation des amerikanischen Englisch die folgenden sechs Pitchakzente zur Verfügung:

monotonal: H*, L*

bitonal: H*+L, H+L*, L*+H, L+H*

Diesen sechs Pitchakzenten entsprechen auf der phonetischen Seite charakteristische, wohldefinierte und distinkte Pitch-Verläufe. Die phonetische Entsprechung der monotonalen Akzente sind lokale F_0-Gipfel bzw. -Täler. Komplexe Akzente sind so zu interpretieren, daß immer der Sternton (T*) mit der akzentuierten Silbe assoziiert ist, während *leading tones* (T+) bzw. *trailing tones* (+T) als tonale Targets vor bzw. nach der akzentuierten Silbe realisiert werden. D.h. H*+L erscheint als F_0-Gipfel in der akzentuierten Silbe, gefolgt von einem tiefen Target; H+L* wird hingegen realisiert als hohes Target vor der akzentuierten Silbe, gefolgt von einem F_0-Tal in der akzentuierten Silbe. Das Resultat ist beide Male eine fallende Kontur, da Transitionen zwischen phonologisch repräsentierten Ton-Targets im phonetischen Regelapparat durch einfache Interpolation modelliert werden. Der Unterschied zwischen den beiden Pitchakzenten besteht allein darin, an welcher Position relativ zur betonten Silbe die Pitch-Targets realisiert werden.

[7] T (Ton) ist ein Platzhalter für H oder L; runde Klammern kennzeichnen optionale Konstituenten; der Asterisk kennzeichnet Pitchakzente, das Minuszeichen Phrasenakzente, das Prozentzeichen IP-Grenztöne.

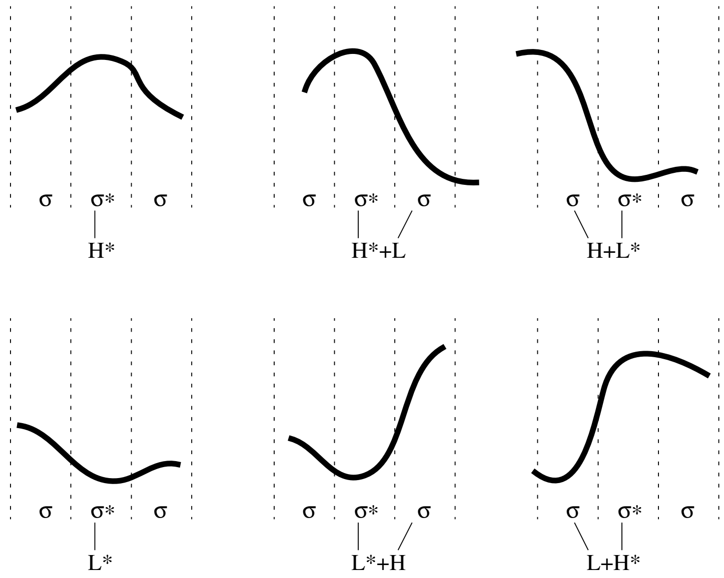


Abbildung 2-1: *Pitchakzente, Ton-Silben-Assoziation und idealisierte Pitch-Verläufe (σ* kennzeichnet jeweils die akzentuierte Silbe).*

Für L*+H vs. L+H* gilt die analoge Interpretation. Abbildung 2-1 zeigt die Prinzipien der Assoziation zwischen Silben und Tönen und die phonetischen Konsequenzen in Form idealisierter Pitch-Konturen. Die genaue temporale Justierung von Pitch-Targets innerhalb von Silben (*temporal alignment*) ist relativ invariant (Ladd, 1996: 67f) und kann durch sprachspezifische phonetische Regeln beschrieben werden (Bruce, 1977 (Schwedisch); Ashby, 1978 (britisches Englisch); Jilka, 1996 (amerikanisches Englisch)).

Für obligatorisch monotonale Phrasenakzente und Terminaltöne bestehen jeweils nur zwei mögliche Ausprägungen: H- und L- bzw. H% und L%. Beide Grenztonkategorien

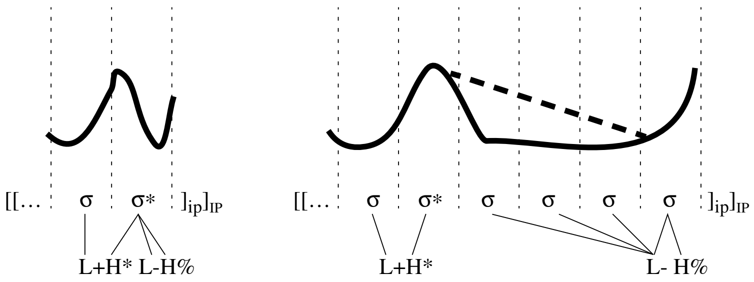


Abbildung 2-2: *Die Prinzipien der autosegmentalen Assoziation am Beispiel der Sequenz L+H* L-H% (in der Terminologie der ›traditionellen britischen Schule‹:* rise-fall-rise*):* leading tones *werden mit der präakzentuierten Silbe assoziiert, Sterntöne mit der akzentuierten Silbe, Phrasenakzente und Terminaltöne mit der letzten Silbe innerhalb der entsprechenden Domäne. Tonal unterspezifizierte Silben werden aufgrund von* feature spreading *mit dem Phrasenakzent assoziiert, dadurch wird die charakteristische Pitch-Kontur korrekt modelliert. Ohne* feature spreading *würde man eine interpolierte Kontur (gestrichelte Linie) erwarten, die jedoch in dieser Form typischerweise nicht realisiert wird.*

sind stets mit der letzten Silbe in der entsprechenden Domäne (ip oder IP) assoziiert. In der Realisierung unterscheiden sich jedoch Phrasenakzent und Terminalton: Letzterer wird durch einen Pitch-Anstieg (H%) bzw. -Abfall (L%) innerhalb der assoziierten Silbe realisiert, während der Phrasenakzent unter Umständen den Pitch-Verlauf innerhalb mehrerer vorangehender Silben determiniert. Die Ursache hierfür ist die Fähigkeit von Phrasenakzentmerkmalen, sich gemäß dem autosegmentalen Prinzip des *feature spreading* auszubreiten: tonal unterspezifizierte Silben zwischen Pitchakzent und Phrasenakzent werden mit dem Phrasenakzentmerkmal assoziiert; dies ist in Abbildung 2-2 illustriert.

Ein zentrales Konzept des TSM ist die Modellierung globaler F_0-Tendenzen mithilfe lokaler Interaktionsregeln. Im Mittelpunkt steht dabei die Analyse der Deklination, also der Tendenz, daß eine imaginäre Linie durch aufeinanderfolgende H-Targets innerhalb einer Phrase in der Regel nicht horizontal, sondern abfallend verläuft. Im Rahmen des

TSM wird dieses Phänomen durch *downstep*-Regeln beschrieben. Im Modell von Pierrehumbert (1980) wird *downstep* durch eine bestimmte Tonabfolge automatisch ausgelöst: jeder L-Ton unmittelbar vor einem H-Ton hat *downstep* des H-Tons zur Folge. Ein *downstepped* H wird um eine Stufe niedriger realisiert als ein vorangegangenes H, wobei die Stufen in einer *downstep*-Sequenz stets von gleicher Größe sind (Liberman & Pierrehumbert, 1984; Pierrehumbert & Beckman, 1988). Bei einer Abfolge H_1 L H_2 L H_3 wird H_1 auf dem höchsten F_0-Niveau n_1 realisiert, H_2 auf dem Niveau $n_2=xn_1$ und H_3 auf dem Niveau $n_3=xn_2$ ($0>x<1$); x bleibt innerhalb einer Phrase konstant. Auch nach der Modifikation durch Beckman & Pierrehumbert bleibt der automatische Mechanismus erhalten, Auslöser von *downstep* sind nun jedoch alle bitonalen Pitchakzente, nicht mehr nur die spezifische Sequenz HL (Beckman & Pierrehumbert, 1986: 276ff). Gegen diese automatische Anwendung von *downstep* wandte sich Ladd (1983) mit der Beobachtung, daß in Intonationssprachen wie dem Englischen (oder dem Deutschen) – im Gegensatz zu Tonsprachen, zu deren Analyse der *downstep*-Mechanismus ursprünglich eingeführt wurde – der Sprecher unabhängig vom tonalen Kontext zwischen der ‚normalen' und der *downstep*-Variante eines Gipfelakzents wählen kann. Ladd begründet dies damit, daß *downstep* etwas bedeutet, „something like 'finality' or 'completeness'" (Ladd, 1996: 90), und daher der Kontrolle des Sprechers und nicht des phonologischen Systems unterliegt. Diese Kritik wurde zwar in späteren Versionen des TSM (z.B. Beckman & Pierrehumbert, 1986) nicht berücksichtigt, in das TSM-basierte Labellingsystem ToBI (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg, 1992), das mit den TSM-Entwicklern zusammen entworfen wurde, ist jedoch die Analyse von Ladd übernommen worden. *Downstep* wird in ToBI nicht durch tonale Interaktion repräsentiert, sondern durch ein kontextunabhängiges Diakritikum '!' (siehe Kapitel 3).

Die Attraktivität des Tonsequenzmodells für die Ausarbeitung der Prosodie-Semantik-Schnittstelle liegt darin begründet, daß im TSM die Melodie einer Phrase (*tune*) kompositional analysiert wird – als linear strukturierte Sequenz phonologischer Merkmale (Töne), die unterschiedlichen funktionalen Kategorien zugeordnet sind (Akzente und

Grenztöne). Dies vereinfacht (bzw. ermöglicht erst) die formale Verknüpfung von Objekten einer kompositionalen Semantik mit Objekten der phonologischen Repräsentation. Als phonologisches Modell der Intonation ist das TSM von Anfang an mit dem Anspruch angetreten, daß alle postulierten Kategorien auf der Ebene der phonologischen Repräsentation bedeutungstragend sind, und daß sich die Bedeutung eines *tunes* aus den Bedeutungen seiner Bestandteile errechnen läßt. Dieser Anspruch wurde mit den Arbeiten von Pierrehumbert & Hirschberg (1990) und Hobbs (1990) eingelöst. Bevor ich jedoch auf diese Arbeiten in Abschnitt 2.3 eingehe, werde ich zunächst die TSM-Analyse der deutschen Intonation von Féry (1993) vorstellen.

2.2 TSM-Analyse des Deutschen

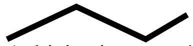
Ein zentrales Konzept der ›traditionellen britischen Schule‹ der Intonationsanalyse – *nuclear tones* (Cruttenden, 1986: Kapitel 3) – spielt im klassischen TSM von Pierrehumbert und Mitarbeitern keine besondere Rolle mehr. Was in der britischen Schule als *nuclear tones* beschrieben wird – komplexe, phrasenfinale Pitch-Konturen (*rise-fall-rise*, *low fall-rise* etc.) – erscheint in Pierrehumberts Modell als Tonsequenz, bestehend aus Pitchakzent, Phrasenakzent und Terminalton. (2) und (3) sind Beispiele für diese Dekomposition und ihre Vorteile (aus Ladd, 1996: 44).

(2) A: I hear Sue‘s taking a course to become a driving instructor.

B: Sue!?

(3) A: I hear Sue‘s taking a course to become a driving instructor.

B: A driving instructor!?



B': A driving instructor!?

Die (schematisch dargestellte) Kontur in (2) ist ein charakteristisches Beispiel für einen *rise-fall-rise nuclear tone*. Die Ton-Sequenz-Analyse beschreibt diese Kontur als L+H* (*rise*) L- (*fall*) H% (*rise*) und postuliert damit, daß die resultierende Kontur zusammengesetzt ist aus mehreren distinkten intonatorischen Ereignissen. Die Schwierigkeiten der holistischen Konfigurationsanalyse[8] zeigen sich bei Beispiel (3). Die charakteristische *rise-fall-rise* Form wird nicht wie in B' angedeutet über die gesamte Äußerung gedehnt, sondern die Kontur wird wie in B realisiert, zusammengesetzt aus distinkten, linguistisch relevanten Ereignissen, die durch linguistisch irrelevante Transitionen verbunden werden: ein *rise* (L+H*-Akzent) in der akzentuierten Silbe <driv> gefolgt von einem *fall* (L-Phrasenakzent) und ein *rise* (H%-Terminalton) in der letzten Silbe <tor>, verbunden durch einen Pitch-Verlauf auf tiefem Level (*spreading* des L-Phrasenakzents) in den dazwischen liegenden, tonal unterspezifizierten Silben <ing instruc>.

Dieser Vorteil der größeren Adäquatheit in der Beschreibung von Pitch-Konturen führte im TSM zur Deklassierung des gesamten Konzepts der *nuclear tones*. Auch ein weiterer Aspekt, der eng mit dem *nuclear-tones*-Konzept verbunden ist, nämlich die grundsätzliche Unterscheidung zwischen *nuclear* und *prenuclear tones*, verlor damit an Bedeutung. Während in der britischen Schule in der pränuklearen Position (also nicht-final) nur einfache Leveltöne vorkommen können, die auch in ihrer Bedeutung sehr stark eingeschränkt sind (Cruttenden, 1986: 62f), ist das Akzentinventar des TSM nicht in dieser Weise restringiert. Alle Akzente sind sowohl nicht-final als auch final (bezogen auf die Intonationsphrase) erlaubt.

[8] Einen Überblick über die *levels-vs.-configuration*-Diskussion bietet Ladd (1996: 59ff).

Der Grund, weshalb die Problematik des *nuclear tone* bzw. der nuklearen vs. pränuklearen Akzente hier angesprochen wurde, ist der, daß Féry (1993) dieser Unterscheidung in ihrer Analyse deutscher Intonationsmuster wieder eine größere Bedeutung zugesteht und auf diese Weise zu einer sehr ökonomischen Repräsentation gelangt. Bevor jedoch erläutert werden kann, an welcher Stelle in Férys Modell diese Unterscheidung Bedeutung erlangt, muß zunächst eine andere Erweiterung, die das klassische TSM durch Férys Analyse erfährt, behandelt werden.

Pierrehumbert und Mitarbeiter gehen nur von *einer* Ebene der tonalen Repräsentation aus (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). Alle Merkmale auf dieser Ebene sind ›basisgeneriert‹, es werden keinerlei Derivationsprozesse angenommen. Féry dagegen unterscheidet zwei Ebenen der Repräsentation: Tiefenstruktur und Oberflächenstruktur (Féry, 1993: 113ff). Die Vermittlung zwischen diesen beiden Ebenen übernehmen sog. *tone linking rules*, wie sie zuerst von Gussenhoven (1984) für das Englische vorgeschlagen wurden. Linking-Prozesse basieren auf der Beobachtung, daß die Phrasierung von Äußerungen sehr stark vom Sprechstil, von der Sprechgeschwindigkeit etc. abhängig ist, daß aber unabhängig davon die zugrundeliegende tonale Struktur als identisch angenommen werden kann. Ein Beispiel:

(4) Detektivromane? Sollen keine Literatur sein?

Die Äußerung in (4) kann sowohl mit zwei Phrasen ([*Detektivromane*]$_{IP/ip}$ [*Sollen keine Literatur sein*]$_{IP/ip}$) als auch innerhalb einer Phrase ([*Detektivromane sollen keine Literatur sein*]$_{IP}$) realisiert werden. In beiden Fällen würde man einen Pitchakzent (*rise*) auf *DetekTIVromane* und auf *LiteraTUR* erwarten. Zunächst die schematische Darstellung einer typischen F_0-Kontur bei der 2-phrasigen Realisierung[9]:

[9] Die bitonalen Pitchakzente konstituieren jeweils eine intermediäre Phrase (ip), ein Phrasenakzent wird im Deutschen nicht benötigt (s.u.).

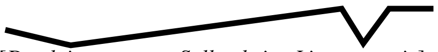
(5)

[[*Detektivromane*]$_{ip}$ [*Sollen keine Literatur sein*]$_{ip}$]$_{IP}$

L*+H L*+H

Beide Phrasen starten auf etwa dem selben F_0-Niveau, dann wird jeweils in der akzentuierten Silbe ein F_0-Tal realisiert (L*), gefolgt von einem Anstieg (+H) und einem hohen Level bis zum Ende der Phrase. Wenn nun, z.B. bei einem schnelleren Sprechtempo, die beiden Phrasen zu einer Phrase verschmolzen werden, werden typischerweise die Konturen in (6) oder (7) realisiert:

(6) partielles Linking



[[*Detektivromane* *Sollen keine Literatur sein*]$_{ip}$]$_{IP}$

L* +HL*+H

(7) komplettes Linking

[[*Detektivromane* *Sollen keine Literatur sein*]$_{ip}$]$_{IP}$

L* L*+H

Der erste (pränukleare) Pitchakzent wird bei der Verschmelzung zweier Phrasen reduziert, indem der Trailing-Ton abgespalten wird und sich entweder mit dem zweiten (nuklearen) Akzent verbindet und mit der präakzentuierten Silbe assoziiert wird – partielles Linking – oder komplett eliminiert wird – komplettes Linking (Féry, 1993: 116ff). Das Resultat ist ein im Vergleich zu (5) flacherer F_0-Anstieg bei partiellem Linking in (6), als Konsequenz aus der Interpolation zwischen L* und gelinktem +H. In (7) folgt dem lokalen Tal des ersten Akzents kein Anstieg, da das H-Target eliminiert wurde, die

Kontur wird zwischen den beiden nunmehr direkt benachbarten tiefen Targets interpoliert.

Mit diesen Linking-Regeln lassen sich aus jeder beliebigen Kombination von zugrundeliegenden bitonalen Pitchakzenten entsprechende Oberflächenrepräsentationen ableiten. Linking-Prozesse lassen sich jedoch nur auf pränukleare, also nicht-finale Akzente anwenden, ein nuklearer Akzent – per Definition immer ein phrasenfinaler Akzent – kann nicht reduziert werden, da im rechten Kontext ein Linking-Partner fehlt; d.h. nukleare Akzente sind – zumindest im Deutschen (s.u.) – stets bitonal, während pränukleare Akzente modifiziert und reduziert werden können. Der Vorteil von Linking-Regeln und einer 2-stufigen Repräsentation ist der, daß scheinbar sehr unterschiedliche Konturen aufgrund einer gemeinsamen zugrundeliegenden Struktur zusammengefaßt werden können. So werden performanzabhängige Varianten durch phonologische Derivationsprozesse abgedeckt, während in der für die Interaktion mit anderen linguistischen Modulen (Syntax, Semantik) relevanten Tiefenstruktur die entscheidenden Gemeinsamkeiten erhalten bleiben. Die Tiefenstruktur repräsentiert nach dieser ‚Entlastung' tatsächlich nur noch bedeutungstragende, distinkte intonatorische Kategorien.

Die Folge ist, daß für die Analyse der Intonationsmuster des Deutschen ein Inventar von fünf Pitchakzenten ausreichend ist, wobei 3 davon nur in sehr restringierten Kontexten verwendet werden können (Féry, 1993: Kapitel 3).

(8) H*+L
L*+H
HH*+L (s.u.)
L*+HL (s.u.)
H*M (s.u.)

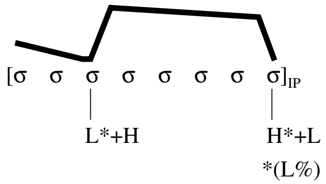
Außerdem kann für die Analyse der deutschen Intonation auf den Phrasenakzent verzichtet werden (Féry, 1993: 74ff). Beide Funktionen, die der Phrasenakzent im Original-TSM erfüllt (nach der Modifikation durch Beckman & Pierrehumbert, 1986), nämlich die

Modellierung tonal unterspezifizierter Silben und die Kennzeichnung intermediärer Phrasengrenzen, werden in Férys Analyse von den Trailing-Tönen der Pitchakzente übernommen:

Nukleare Akzente sind im Deutschen stets bitonal, da der letzte Akzent einer Phrase immer durch eine Grundfrequenzbewegung realisiert wird (Féry, 1993: 81); d.h. ein nuklearer Akzent ist immer zusammengesetzt aus einem Sternton und (mindestens) einem Trailing-Ton (+L oder +H). Dieser Trailing-Ton determiniert das F_0-Niveau bis zur vorletzten Silbe der Phrase, bevor der Terminalton das Niveau in der letzten Silbe bestimmt. Pränukleare Akzente können, wie oben gezeigt, sowohl bitonal als auch monotonal sein, wobei nur ein bitonaler Akzent eine intermediäre Phrase konstituiert (monotonale Akzente sind das Produkt von Linking, also der Verschmelzung von zwei intermediären Phrasen). Das heißt, im Falle eines bitonalen pränuklearen Akzents kann die Kontur zwischen Akzent und intermediärer Phrasengrenze ebenfalls durch *spreading* des Trailing-Tons modelliert werden, während im Falle eines monotonalen pränuklearen Akzents 1.) keine Phrasengrenze gekennzeichnet werden muß und 2.) die Kontur zwischen benachbarten Akzenten durch einfache Interpolation modelliert werden kann (der charakteristische, interpolierte Verlauf zwischen gelinkten Akzenten ist ja die eigentliche Motivation für die Linking-Analyse).

Eine weitere Vereinfachung des Originalmodells stellt Férys Verzicht auf den L%-Terminalton dar (Féry, 1993: 72). Da im Deutschen nach einem nuklearen H*+L Akzent kein weiterer F_0-Abfall in der letzten Silbe einer Intonationsphrase zu beobachten ist, reicht der L-Trailington aus, um den Konturverlauf zu repräsentieren. Das evtl. zu beobachtende tiefere F_0-Niveau eines finalen L-Targets im Vergleich mit einem nicht-finalen L-Target kann durch *final lowering* erklärt werden. Die zweite Möglichkeit, die einen L%-Ton erforderlich machen würde, ist ein tiefer Terminalton nach hohem Pitch-Verlauf nach dem nuklearen Akzent (L*+H L%). Eine solche Kontur, ein F_0-Abfall in der letzten Silbe einer Phrase, dargestellt in (9), wird jedoch nicht als tiefer Terminalton interpretiert, sondern als Prominenzmerkmal, also als H*+L Akzent. Aufgrund dieser beiden Beobachtungen wird der L%-Ton überflüssig.

(9)



Während die L%-Grenze den unmarkierten Fall darstellt, der nicht gekennzeichnet werden muß, kann auf den H%-Ton nicht verzichtet werden. In Férys Modell dient er der Repräsentation von – in traditioneller Terminologie – *fall-rise* Konturen, also einem finalen F_0-Anstieg nach tiefem Trailing-Ton (H*+L H%; Féry, 1993: 91ff)[10].

Neben diesen zahlreichen Vereinfachungen, die mit der ursprünglichen Motivation und Konzeption des Original-TSM weitestgehend kompatibel sind, stellt Féry einen Aspekt zur Diskussion, der eine relativ einschneidende Modifikation des Modells darstellt: die Einführung eines dritten Tonmerkmals M. Das Phänomen, zu dessen Analyse der M-Ton vorgeschlagen wird, ist eine Klasse von Konturen, die sich in vielerlei Hinsicht von ‚normalen' Konturen unterscheiden. Féry (1993: 97ff) bezeichnet diese Konturen in Anlehnung an Ladd (1978), der diesen Begriff zuerst einführte, als *stylized contours*. Diese Konturen werden meist bei Vokativen verwendet, daher auch die Bezeichnung *call*-Kontur von Gibbon (1976). Abbildung 2-3 zeigt eine Realisierung des Vokativs *Angelika!* von einem männlichen Sprecher mit einer charakteristischen stilisierten Kontur. Die erste Silbe startet auf relativ tiefem Niveau, in der nächsten – akzentuierten – Silbe wird ein F_0-Gipfel erreicht (H*) und dieser hohe Level wird auch in der dritten Silbe beibehalten, bevor die letzte Silbe mit einem mittleren F_0-Level realisiert wird. Die

[10] Die Möglichkeit eines zusätzlichen F_0-Anstiegs nach hohem Trailing-Ton (L*+H H%) wird von Féry nicht berücksichtigt, kann jedoch in den von uns analysierten Daten einige Male beobachtet werden und sollte möglicherweise auch phonologisch repräsentiert werden (s. Abschnitt 3.2).

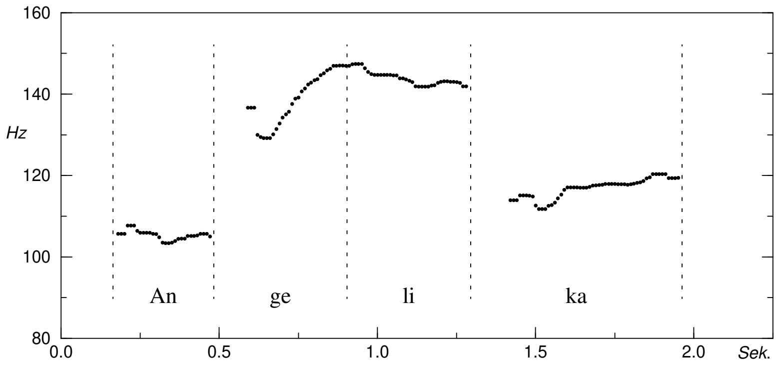


Abbildung 2-3: *Grundfrequenzverlauf einer Realisation von* Angelika! *mit stilisierter Kontur. Die gepunkteten Linien kennzeichnen die Silbengrenzen.*

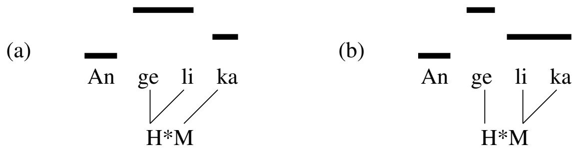
wesentlichen Merkmale der *stylized contour* sind 1.) der hohe Level bis zur vorletzten Silbe und 2.) der Levelton auf mittlerem Niveau in der letzten Silbe. Das zweite Merkmal ist obligatorisch, das erste optional, d.h. beide Konturen in (10) sind möglich und unterscheiden sich nicht in ihrer Interpretation.

(10)

Beckman & Pierrehumbert (1986: 276ff) analysieren diese Art von Konturen als eine Sequenz aus einem beliebigen bitonalen Akzent, gefolgt von H-L%, wobei H-L% den finalen Levelton auf mittlerem Niveau repräsentiert. Die Überlegung ist dabei folgende: Der Phrasenakzent H- ist *downstepped* (darum muß ein bitonaler Pitchakzent als *downstep*-Auslöser angenommen werden (s. Abschnitt 2.1)), während der Terminalton L% *upstepped* ist, ausgelöst durch den H-Phrasenakzent. Das Resultat ist ein abgesenkter H-

Ton und ein angehobener L-Ton – also ein F_0-Level auf mittlerem Niveau. Féry kritisiert, daß „[this] analysis is motivated by theory-internal considerations: it contributes to constrain the theory by restricting the number of tonal values to two, H and L, but in my opinion, it is not intuitive“ (Féry, 1993: 98). Als alternative Analyse schlägt sie die Einführung eines bitonalen Pitchakzents mit M-Trailing-Ton vor: H*M. Dieser Akzent kann nur an nuklearer Position stehen und ist beschränkt auf stilisierte Konturen. Eine weitere Besonderheit ist, daß dies der einzige Akzent ist, dessen Sternton sich ausbreiten kann (optionales *feature spreading*). Die Varianten in (10) werden wie folgt repräsentiert:

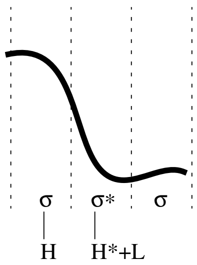
(11)



Der M-Ton ist nur in dieser spezifischen Kombination H*M möglich (Féry, 1993: 102) und nur im spezifischen Kontext stilisierter Konturen. Diese Ausnahmestellung stilisierter Konturen erhält auch dadurch weitere Evidenz, daß die Silben in einer solchen Äußerung häufig stark gelängt sind, insbesondere die letzte Silbe, die obligatorisch mit mittlerem Levelton realisiert wird (Féry, 1993: 101; siehe auch Abb. 2-3).

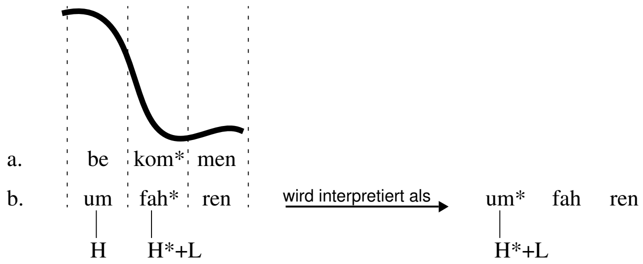
Ich werde die Analyse stilisierter Konturen hier nicht weiter ausführen, da dieser Konturtyp im weiteren Verlauf dieser Arbeit keine besondere Rolle spielt (für eine ausführliche Diskussion verweise ich auf Féry, 1993: 97ff). Zum Abschluß möchte ich noch kurz auf die tritonalen Akzente HH*+L und L*+HL eingehen. HH*+L ist die TSM-Repräsentation des *early peak* (‚früher Gipfel‘; vgl. Kohler, 1991b; Gartenberg & Panzlaff-Reuter, 1991). Der *early peak* hat die Form eines lokalen F_0-Maximums in der präakzentuierten Silbe, gefolgt von einem F_0-Abfall bzw. einem tiefen F_0-Target in der akzentuierten Silbe.

(12) *early peak*



Eine notwendige Bedingung für die Verwendung eines HH*+L Akzents ist die, daß die präakzentuierte Silbe, die mit dem frühen H-Ton assoziiert wird, metrisch schwach ist (vgl. (13)a), da der Pitchverlauf sonst als H*L Akzent interpretiert wird (vgl. (13)b) (Féry, 1993: 103f).

(13)



Der L*+HL Akzent ist ein gewöhnlicher L*+H Akzent, gefolgt von einem F_0-Abfall; er entspricht in etwa dem *late peak* (‚später Gipfel‘) von Kohler und Mitarbeitern (Kohler, 1991b; Gartenberg & Panzlaff-Reuter, 1991). Der Unterschied zwischen den beiden tritonalen Akzenten ist der, daß HH*+L auf mindestens zwei Silben realisiert werden muß, L*+HL kann dagegen gemäß den autosegmentalen Assoziationsprinzipien auch auf einer Silbe realisiert werden. Das bedeutet, daß die *peak*-Verschiebung nach links (*early peak*) durch eine veränderte Akzent-Silben-Assoziation repräsentiert wird, während die

Repräsentation der *peak*-Verschiebung nach rechts (*late peak*) lediglich einen zusätzlichen Trailing-Ton erfordert. Dieser Unterschied in der phonologischen Repräsentation reflektiert die prinzipielle Differenz zwischen *early* und *late peak*: Zwischen *early peak* (HH*+L) und *mid peak* (H*+L) besteht ein kategorialer Unterschied, der auch einen Bedeutungsunterschied induziert, zwischen *mid peak* und *late peak* (L*+HL) besteht dagegen nur ein gradueller Unterschied, der mit dem Grad der Emphase korreliert (vgl. die Arbeiten von Kohler und Mitarbeitern in Kohler, 1991c).

Zusammenfassung: Zur Beschreibung der wichtigsten Intonationsmuster des Deutschen im Rahmen des Tonsequenzmodells wird das folgende Toninventar benötigt:

- zwei Basis-Pitchakzente, H*+L und L*+H, die sowohl in nuklearer als auch in pränuklearer Position stehen können;
- drei ergänzende Pitchakzente, H*M, HH*+L und L*+HL, die nur in nuklearer Position stehen können[11] und z.T. auf einen sehr spezifischen Kontext beschränkt sind;
- ein Terminalton, H%, zur Beschreibung steigender IP-Grenzen; für die Modellierung fallender bzw. tiefer IP-Grenzen ist der nukleare Pitchakzent ausreichend;

Auf Phrasenakzente kann verzichtet werden, ihre Funktion übernehmen die Trailing-Töne der bi- oder tritonalen Pitchakzente. Durch die Anwendung optionaler Linking-Regeln kann eine Vielzahl scheinbar sehr unterschiedlicher Oberflächenkonturen

[11] Féry (1993: 103) bemerkt, daß auch Sequenzen von zwei aufeinanderfolgenden *early peak*-Akzenten beobachtet werden können. Im Gegensatz zu den zwei Basisakzenten ist jedoch bei einem ‚pränuklearen' HH*+L-Akzent kein Tonlinking zu erwarten, so daß dem *early peak* die charakteristische Eigenschaft eines pränuklearen Akzents fehlt. Außerdem ist zu erwarten, daß ein ‚pränuklearer' HH*+L-Akzent nur vor einem (nuklearen) HH*+L-Akzent stehen kann. Insofern sollte eine HH*+L-Sequenz adäquater als nuklearer Akzentkomplex analysiert werden, denn als eine konventionelle Abfolge aus pränuklearem und nuklearem Akzent.

adäquat modelliert und aus einer relativ geringen Menge phonologisch distinkter Grundstrukturen abgeleitet werden.

2.3 Intonation und Bedeutung: Neuere TSM-basierte Ansätze

Wie die Überschrift zu diesem Abschnitt bereits andeutet, werde ich im folgenden nicht die sehr umfangreiche Literatur zur Bedeutung verschiedenster intonatorischer Merkmale aufarbeiten, sondern mich stattdessen auf zwei neuere Ansätze konzentrieren – Pierrehumbert & Hirschberg (1990) bzw. Hobbs (1990) einerseits und Büring (1995) bzw. Steedman (1997) andererseits –, die auf der Ton-Sequenz-Analyse beruhen und die bereits angedeutete Konzeption einer kompositionalen Semantik von intonatorischen Kategorien verfolgen. Die Arbeit von Pierrehumbert & Hirschberg und der Kommentar von Hobbs stehen dabei im Mittelpunkt, da in diesen Arbeiten tatsächlich versucht wird, in der Tradition von z.B. Bolinger (1982), Gussenhoven (1984) oder Ward & Hirschberg (1985) ein umfassendes System der *intonational meaning* auf der Basis des Tonsequenzmodells zu entwickeln. Die Arbeiten von Büring und Steedman dienen dagegen eher der Illustration neuerer Entwicklungen auf dem klassischen Gebiet der Syntax-Phonologie-Schnittstelle, also der Interaktion zwischen Intonation und Informationsstruktur. Auf die für die Analyse der deutschen Intonation wichtigen Beiträge des Münchner Intonationsprojekts[12] zur satzmodalen Funktion der Intonation und des Kieler Intonationsprojekts[13], das ebenfalls Aspekte der *intonational meaning* anschneidet, werde ich an dieser Stelle nicht gesondert eingehen, da diese Arbeiten nicht auf dem Tonsequenzmodell basieren und die (nicht unproblematische) Transformation der Ergebnisse in eine TSM-basierte Beschreibung der deutschen Intonation im Rahmen der vorliegenden Arbeit nicht gelei-

[12] siehe die Beiträge von Altmann, Batliner, Oppenrieder und anderen in Altmann (1988) und Altmann, Batliner & Oppenrieder (1989).

[13] siehe die Beiträge von Kohler, Gartenberg, Hertrich und Panzlaff-Reuter in Kohler (1991c).

stet werden kann. An einzelnen relevanten Stellen meiner Untersuchung werden diese Arbeiten jedoch genannt und berücksichtigt.

2.3.1 Ein kompositionales Modell für *intonational meaning*

Das kompositionale Modell von Pierrehumbert & Hirschberg (1990; im weiteren Verlauf: P&H) beruht auf dem semantischen Konzept der *mutual beliefs* (Clark & Marshall, 1981; Joshi, 1982): „we understand the *mutual beliefs* of a discourse to be those beliefs that conversational participants come to belief to be shared among them as a direct result of the conversational interaction“ (P&H: 285; Hervorhebung im Original). Es handelt sich also um Annahmen des Sprechers bzw. des Hörers/der Hörer bezüglich dessen, welche Propositionen aus der Menge der möglichen Propositionen bei den Diskursteilnehmern als bekannt und akzeptiert vorausgesetzt werden können. Auf der Basis dieser Konzeption hat die Intonation die Funktion, eine bestimmte Relation zwischen dem propositionalen Gehalt der aktuell intonierten Äußerung und den *mutual beliefs* der Teilnehmer am aktuellen Diskurs zu spezifizieren. Diese Spezifizierung entspricht der *tune*-Bedeutung. Die zentrale Aussage der Arbeit von Pierrehumbert & Hirschberg ist nun, daß diese Relation (*tune*-Bedeutung) kompositional interpretiert werden kann: zusammengesetzt aus dem Beitrag bzw. der Bedeutung der Pitchakzente, der Phrasenakzente und des Grenztons, wobei die Art des Beitrages der drei Grundkategorien grundsätzlich unterschieden wird. Die Basis für diese Unterscheidung bildet der unterschiedliche phonologische ‚Skopus‘, über den die drei Kategorien verfügen: Pitchakzente beziehen sich auf phonologische Wörter, Phrasenakzente auf intermediäre Phrasen und Grenztöne auf ganze Intonationsphrasen. Diese phonologische Hierarchie wird von den Autorinnen folgendermaßen auf die semantische Funktion abgebildet:

1. Pitchakzente
   „Pitch accents convey information about the status of the individual discourse referents, modifiers, predicates, and relationships specified by the lexical items with which the accents are associated.“ (P&H: 286)

2. Phrasenakzente
   „Phrase accents convey information at the level of the intermediate phrase. […] We propose the S [Speaker] chooses phrase accent type to convey the degree of relatedness of one such phrase to preceding and succeeding intermediate phrases.“ (P&H:287)

3. Grenztöne
   „The boundary tones contribute information about the intonational phrase as a whole. […] We believe that boundary tones convey information about relationships among intonational phrases – in particular, about whether the current phrase is to be interpreted with particular respect to a succeeding phrase or not.“ (P&H: 287)

> „Together, these intonational features can convey how S intends that H [Hearer] interpret an intonational phrase with respect to (1) what H already believes to be mutually believed and (2) what S intends to make mutually believed as a result of subsequent utterances.“ (P&H: 288)

Ausgehend von dieser Grundkonzeption entwickeln P&H ein Modell für *intonational meaning*, welches jeder Kategorie des Pierrehumbert‘schen Tonsequenzmodells eine eigene, spezifische Bedeutung zuweist. Ich werde dieses Modell im folgenden kurz skizzieren, für eine ausführliche Diskussion der einzelnen Kategorien verweise ich auf die entsprechenden Abschnitte in P&H (Seiten 288ff).

2.3.1.1 *Pitchakzente*

Hier kann zunächst unterschieden werden zwischen den beiden monotonalen Akzenten H* und L*. H* zeichnet ein lexikalisches Item als ‚neu im Diskurs' aus. Der Sprecher signalisiert mit H*, daß die mitgeteilte Information – das akzentuierte Item und die Prädikation – zu den bereits existierenden *mutual beliefs* hinzu addiert werden soll. Die Bedeutung von L* kann dagegen charakterisiert werden als „*salience-without-predication*" (P&H: 292). Ein L*-akzentuiertes Item wird zwar aufgrund der Akzentuierung hervorgehoben, d.h. es ist salient in der Diskursrepräsentation, es wird aber ausgeschlossen aus der Prädikation, die gemäß der Intention des Sprechers zu den *mutual beliefs* hinzu addiert werden soll. P&H führen verschiedene Motivationen für einen L*-Akzent an: z.B. kann dieser Akzent verwendet werden, wenn der Sprecher gar nicht in der Lage ist, etwas über das akzentuierte Item auszusagen, also beispielsweise in Entscheidungsfragen:

(14) *Do prunes have feet*[14]
L* L* H-H%

Die Prädikation wird in diesem Fall dem Diskurspartner überlassen. Ein weiterer Kontext für einen L*-Akzent ist dann gegeben, wenn die mitgeteilte Information bereits Teil der *mutual beliefs* ist – Beispiel (15) – bzw. wenn der Sprecher darauf hinweisen möchte, daß die mitgeteilte Information eigentlich Teil der *mutual beliefs* sein sollte – Beispiel (16).

[14] Alle englischsprachigen Beispiele in diesem Abschnitt sind, sofern nichts anderes angegeben ist, aus P&H entnommen.

(15) Kontext: Der Sprecher wird um eine Wunschliste zum Geburtstag gebeten und es ist bekannt, daß sein sehnlichster Wunsch eine Pavoni Espressomaschine ist:
Well, I'd like a Pavoni …
L* L* L* L-H%

(16) Kontext: Dem Sprecher, bekannt als Vegetarier, wird vorgeschlagen, gemeinsam ein Chateaubriand für zwei Personen zu bestellen:
I don't eat beef
L* L* L-H%

P&H führen noch einige weitere Beispiele an, die für ihre Interpretation des L*-Akzents sprechen, die ich jedoch hier nicht behandeln möchte (siehe P&H: 291ff).

Die vier bitonalen Akzente L*+H, L+H*, H*+L und H+L* werden von P&H in zwei Untergruppen zusammengefaßt: L+H-Akzente und H+L-Akzente. L+H-Akzente, so P&H, „are employed by S to convey the salience of some *scale* […] linking the accented item to other items salient in H's mutual beliefs" (P&H: 294; Hervorhebung im Original). Die weitere Unterscheidung zwischen L*+H und L+H* ist dann zurückzuführen auf die Grundbedeutung von L* bzw. H*. L*+H evoziert eine partielle Ordnung (*scale*) und signalisiert gleichzeitig den Mangel einer Prädikation. Diese Kombination führt zu einem Eindruck der ‚Unsicherheit' (P&H: 295f).

(17) A: *Alan's such a klutz.*
B: *He's a good badminton player*
L*+H L-H%

In (17) drückt B's Erwiderung Unsicherheit darüber aus, ob die Tatsache, daß Alan ein guter Badmintonspieler ist, eine relevante Information darstellt bezüglich des Grades von Alans Ungeschicklichkeit. Es wird also ein *scale* evoziert – Grad der Ungeschicklichkeit –, während die Relevanz der Information bezüglich des evozierten *scale* vom

Sprecher bezweifelt wird. Daher wird ein Akzent gewählt, der die Information nicht zu den *mutual beliefs* hinzu addiert, sondern sie stattdessen den Diskursteilnehmern zur Evaluation überläßt.

L+H* evoziert ebenfalls eine partielle Ordnung und signalisiert gleichzeitig, daß die mitgeteilte Information den *mutual beliefs* hinzugefügt werden soll. Dieser Akzent wird sehr häufig in Korrekturen verwendet (P&H: 296).

(18) Kontext: A nimmt irrtümlich an, es sei Januar, dabei ist es noch Dezember:
A: *It's awfully warm for January.*
B: *It's even warm for December*
L+H* L-H%

In solchen Korrekturen substituiert der Sprecher (im Beispiel: B) einen neuen skalaren Wert für einen anderen im Kontext verfügbaren Wert mit der Intention, den neuen Wert den *mutual beliefs* hinzuzufügen.

H+L-Akzente haben ebenfalls einen gemeinsamen Bedeutungsaspekt, nämlich daß der Hörer aus seiner Repräsentation der *mutual beliefs* die mit dem akzentuierten Item mitgeteilte Information erschließen können sollte. Für H*+L gilt erstens, daß – vergleichbar mit H* – eine Prädikation gemacht wird, und daß zweitens der Hörer einen Inferenzpfad – ableitbar aus der Repräsentation der *mutual beliefs* – identifizieren soll, der diese Prädikation unterstützt (P&H: 297). Dieser Akzent hat häufig einen ‚pädagogischen Beigeschmack', was nicht verwunderlich ist, „since teaching involves pointing the student to inference relationships between old and new information" (P&H: 298). Die Bedeutung des zweiten H+L-Akzents – H+L* – ist sehr ähnlich: Er signalisiert, daß die mitgeteilte Information bereits Bestandteil der *mutual beliefs* ist, und daß der Hörer dies eigentlich wissen sollte.

2.3.1.2 Phrasenakzente

Phrasenakzente haben, wie oben bereits erwähnt, Skopus über intermediäre Phrasen. Ihre Funktion ist es, die Beziehung zwischen benachbarten Phrasen dieses Typs zu signalisieren (P&H: 302ff). Ein hoher Phrasenakzent H- bedeutet, daß die aktuelle Phrase gemeinsam mit der folgenden Phrase eine interpretative Einheit bildet, während ein tiefer Phrasenakzent L- die Separation zwischen zwei benachbarten Phrasen betont. P&H illustrieren diesen Unterschied anhand von disjunktiven Listen und Exhaustivität:

(19) *Do you want apple juice or orange juice*
H* H- H* L-L%

(20) *Do you want apple juice or orange juice*
H* L- H* L-L%

In (19) werden *apple juice* und *orange juice* aufgrund des H-Phrasenakzents als Einheit interpretiert, nämlich als die Menge der verfügbaren Säfte. Diese Aufzählung wird daher mit großer Wahrscheinlichkeit als exhaustive Liste aufgefaßt. In (20) liegt dagegen die Emphase auf dem separaten Status der beiden Safttypen – aufgrund des L-Phrasenakzents. Die beiden Items bilden daher keine Einheit und die Aufzählung erscheint eher als nicht-exhaustiv.

2.3.1.3 Grenztöne

Die Bedeutung der Grenztöne H% und L% ist vergleichbar mit der Bedeutung der Phrasenakzente H- und L-, der wesentliche Unterschied ist lediglich der Skopus: Grenztöne beziehen sich auf ganze Intonationsphrasen. H% signalisiert ebenso wie H- „*forward reference*" (P&H: 304ff), d.h. für die Interpretation einer H%-gekennzeichneten Phrase soll die folgende Phrase mitberücksichtigt werden. L% signalisiert dagegen keine *forward reference*. Diese Unterscheidung hat z.B. Konsequenzen für die Anaphernresolution im Diskurs:

(21) a. *My new car manual is almost unreadable*
L-L%

b. *It's quite annoying*
L-H%

c. *I spent two hours figuring out how to use the jack*
L-L%

In (21), mit einer H%-Grenze nach b, bilden b und c eine Einheit. Dies führt dazu, daß als Referent für das Pronomen *it* die Proposition in c – *‚my spending two hours figuring out how to use the jack'* – bevorzugt wird. Im Gegensatz hierzu Beispiel (22):

(22) a. *My new car manual is almost unreadable*
L-H%

b. *It's quite annoying*
L-L%

c. *I spent two hours figuring out how to use the jack*
L-L%

Die H%-Grenze nach a führt nun zu einer Gruppierung von a und b, und somit zu einer Präferenz für die NP in a – *‚my new car manual'* – als Referenten von *it*.

2.3.1.4 Modifikationen des P&H-Modells

In einem Kommentar zu der Arbeit von Pierrehumbert und Hirschberg entwickelt Hobbs (1990) eine Reihe von Vereinfachungen und Generalisierungen. Auch diese Modifikationen sollen hier kurz skizziert werden, da sie dem Modell insgesamt zu größerer Konsistenz verhelfen und den kompositionalen Aspekt noch mehr in den Vordergrund stellen.

Zunächst schlägt Hobbs eine – im wesentlichen terminologische – Vereinfachung der zugrundeliegenden semantischen Konzeption vor. Die wichtigste semantische Funktion der Intonation ist, so Hobbs, die Differenzierung der mitgeteilten Information in *new*,

given und *false* (Hobbs, 1990: 314). Diese drei Kategorien lassen sich anhand von Teilmengen der Menge aller möglichen Propositionen definieren. Eine Teilmenge der Menge aller möglichen Propositionen entspricht den *private beliefs* des Sprechers, eine andere Teilmenge entspricht den *private beliefs* des Hörers (weitere Teilmengen können für weitere Diskursteilnehmer angenommen werden). Die Menge der *mutual beliefs* entsteht aufgrund der Überlappung zwischen den individuellen Teilmengen (*private beliefs*) der Diskursteilnehmer. Eine weitere Teilmenge der Menge aller möglichen Propositionen bilden „all those propositions that are believed by neither the speaker nor the hearer, many of them because they are false“ (Hobbs, 1990: 314). Aufgrund dieser Differenzierung können die folgenden Definitionen formuliert werden (*loc.cit.*): Eine mitgeteilte Information ist *new*, wenn es sich dabei um ein Element der Teilmenge der *private beliefs* des Sprechers handelt; sie ist *given*, wenn es sich um ein im Äußerungskontext relevantes Element der Teilmenge der *mutual beliefs* handelt; und sie ist *false*, wenn es sich um ein Element der Menge der *false propositions* des Sprechers handelt. Aufgrund dieser Überlegungen charakterisiert Hobbs die semantische Funktion der Intonation folgendermaßen:

> „The prototypical utterance is a bid to turn some of the speaker's private beliefs into mutual beliefs. The utterance is anchored referentially in mutual belief and reaches out into the speaker's private beliefs. An utterance thus contains the given propositions that anchor it referentially and the new propositions that the speaker is informing the hearer of. Much of intonational meaning is concerned with keeping these different partitions of the space of propositions distinguished.“ (*loc.cit.*)

Eine weitere von Hobbs vorgeschlagene Vereinfachung betrifft die Zuordnung von Tönen (H und L) zu funktionalen Einheiten (Pitchakzente, Phrasenakzente, Grenztöne) und die daraus resultierende unterschiedliche Interpretation der Töne. Hobbs schlägt vor, die interpretative Differenzierung von Tönen auf zwei Klassen zu beschränken (*loc.cit.*): 1.) Sterntöne und *leading tones* (‚Präfixe') und 2.) *trailing tones* (‚Suffixe'), Phrasenakzente und Grenztöne. Die erste Klasse umfaßt die monotonalen Pitchakzente H* und L* sowie die beiden bitonalen Akzente L+H* und H+L*. Die Töne H bzw. L haben in dieser Klasse eine jeweils identische Bedeutung: H bedeutet *new*, L bedeutet *not new*. *Not new* kann gemäß der obigen Unterteilung sowohl *given* als auch *false* heißen. Die monotonalen Akzente haben also die einfache semantische Funktion, eine Information als salient auszuzeichnen („important for the correct interpretation of the utterance" *loc.cit.*) und zusätzlich die Information als *new* (H*) oder *not new* (L*) zu kennzeichnen. Die präfigierten Akzente erhalten ihre – komplexe – Bedeutung, neben der Salienzfunktion, durch die Kombination der einfachen H- bzw. L-Bedeutungen: „L+H* says something like, "you might think this information is not new, but it really is new," and H+L* says, "you might think this information is new, but it really is not new."" (*loc.cit.*). Die Modifikation durch ein Präfix drückt also eine Art Korrektur aus, eine Korrektur, „of what the speaker believes the hearer might incorrectly believe the status of the information to be" (*loc.cit.*).

Die Bedeutung von H und L in der zweiten Klasse ist ebenfalls jeweils identisch, d.h. unabhängig von der Funktion des Tons: H signalisiert *incompleteness* oder *open-endedness*, L signalisiert nicht *incompleteness* oder *open-endedness*, d.h. „[it] is not the case that a L suffix signals completeness; it merely fails to signal incompleteness" (*loc.cit.*). Open-endedness wird von Hobbs folgendermaßen charakterisiert: „What I've just conveyed by that morpheme or phrase requires further discussion before it is entered into mutual belief, or before its status with respect to mutual belief is agreed upon" (Hobbs, 1990: 315). Diese Bedeutung von H bzw. L gilt für alle Töne der 2. Klasse, also sowohl für Trailing-Töne (Suffixe), als auch für Phrasenakzente und Grenztöne.

Damit sind alle relevanten Komponenten des Hobbs'schen Modells beschrieben. Mithilfe dieser Basiseinheiten – zusammengefaßt in Tabelle 2-1 – kann, so Hobbs, die

Bedeutung von komplexeren *tunes* auf sehr einfache Weise kompositional analysiert werden.

Tabelle 2-1: Die Bedeutung von H- bzw. L-Tönen nach Hobbs (1990)

	Bedeutung von	
	H	L
Klasse I (Sterntöne, Präfixe)	salient + new	salient + not new
Klasse II (Suffixe, Phrasenakzente, Grenztöne)	incomplete, open-ended	Unterspezifikation bezüglich completeness

Anhand einiger der oben angeführten Beispiele soll die Hobbs'sche Reanalyse des P&H-Ansatzes im folgenden kurz illustriert werden. In der Analyse der monotonalen Akzente H* und L* unterscheiden sich die beiden Ansätze nicht wesentlich. Sowohl bei Hobbs als auch bei P&H kennzeichnet der H*-Akzent neue Information. Die L*-Analyse von P&H – *salience without prediction* – wird von Hobbs verfeinert: „In my account, there are just two ways something can fail to predicate, that is, fail to be proposed as new. It can be proposed as given or as false" (Hobbs, 1990: 315). Ein größerer Unterschied besteht zwischen den beiden Ansätzen in der Analyse bitonaler Akzente. In ihrer Diskussion der bitonalen Akzente und der Kompositionalität von Akzentbedeutungen schreiben P&H: „we note that items differing only in the location of the star have closely related meanings" und „[…] the meaning of each particular pitch accent may be derivable from the meanings of its constituent tones, plus some generalisation about the interpretation of the star" (P&H: 301). Das heißt ein H*-Ton in H*+L oder in L+H* bzw. ein L*-Ton in L*+H oder H+L* hat einen zusätzlichen Bedeutungsaspekt, der nicht aus seiner Grundbedeutung, die für einen Sternton in einem monotonalen Akzent gilt, ableitbar ist. In Hobbs' vollständig kompositionalem Ansatz kann auf solche zusätzlichen Bedeutungsaspekte verzichtet werden. Sterntöne treten entweder alleine oder mit einem Präfix oder mit einem Suffix auf, wobei die Bedeutung von H und L für jede Funktion jeweils so spezifiziert ist, daß sich die Gesamtbedeutung aus den Teilbedeutungen ergibt. Während bei P&H L+H-Akzente und H+L-Akzente jeweils eine Klasse bilden, gruppiert Hobbs T+T*-Akzente (Präfigierung) zusammen und unterscheidet sie von T*+T-Akzenten (Suffigierung). Ich wiederhole hier die Beispiele (17) und (18) von oben als (23) und (24).

(23) A: *Alan's such a klutz.*
B: *He's a good badminton player*
L*+H L-H%

(24) Kontext: A nimnmt irrtümlich an, es sei Januar, dabei ist es noch Dezember:
A: *It's awfully warm for January.*
B: *It's even warm for December*
L+H* L-H%

P&H sehen die Gemeinsamkeit der beiden L+H-Akzente darin, daß beide eine partielle Ordnung (*scale*) evozieren. Dies läßt sich jedoch nicht aus der Grundbedeutung von H und L ableiten. Hobbs gibt dagegen die folgende Interpretation: In (23) signalisiert der L*-Ton, daß die mitgeteilte Information nicht *new* ist, d.h. sie ist *given* oder *false*. Im vorliegenden Beispiel kann davon ausgegangen werden, daß die Information *given* ist, d.h. beide Diskursteilnehmer wissen, daß Alan ein guter Badmintonspieler ist. Durch das H-Suffix wird signalisiert, daß die Proposition ‚offen' ist, ihre Relevanz hinsichtlich der von A aufgestellten Behauptung muß noch geklärt werden. Die Gesamtbedeutung kann also in etwa folgendermaßen umschrieben werden: Ein Element aus der Menge der *mutual beliefs* wird im Diskurs salient gemacht, damit seine Relevanz hinsichtlich des Themas des Diskurses überprüft werden kann. Für (24) gilt hingegen: Der H*-Ton kenn-

zeichnet die Information als *new*, während das L-Präfix die irrtümliche Annahme von A kommentiert: A denkt, die Information ist *not new* (sondern *false*, weil er der Meinung ist, es sei Januar), sie ist aber doch *new* und sollte zu den *mutual beliefs* hinzu addiert werden.

Im Falle der Phrasenakzente (und Grenztöne) unterscheiden sich die Analysen von P&H und Hobbs nur in einem Punkt. Beide Modelle gehen davon aus, daß ein H-Ton *openness* (Hobbs) bzw. die Zusammengehörigkeit zweier benachbarter Phrasen (P&H) signalisiert. Ein L-Ton betont dagegen nach P&H die Separation zweier benachbarter Phrasen. In Hobbs' Analyse wird diese Interpretation ein wenig abgeschwächt. Ein L-Ton signalisiert Unterspezifikation bezüglich *completeness* und betont also weder die Verknüpfung noch die Separation von Phrasen. Diese Analyse garantiert ein sehr einfaches und vollständig kompositionales Modell (vgl. Tabelle 2-1) und führt dennoch zu einer adäquaten Interpretation von Kontrasten wie in (19) und (20) (hier wiederholt als (25) und (26)):

(25) *Do you want apple juice or orange juice*
H* H- H* L-L%

(26) *Do you want apple juice or orange juice*
H* L- H* L-L%

In (25) signalisiert H- am Ende der ersten Phrase *openness*, während die zweite Phrase unterspezifiziert ist. Der Hörer wird also beide Phrasen als Einheit interpretieren und die exhaustive Lesart bevorzugen, wenn nicht eine weitere Äußerung folgt, die dieser Interpretation widerspricht. In (26) sind dagegen beide Phrasen unterspezifiziert, d.h. der Hörer kann die beiden Phrasen sowohl als zusammengehörig als auch als nicht-zusammengehörig interpretieren. Aufgrund der möglichen, aber nicht realisierten H-Alternative wird er die separate, also nicht-exhaustive Lesart bevorzugen.

2.3.2 Pitchakzente und Informationsstruktur

Einen etwas anderen Aspekt der intonatorischen Bedeutung – insbesondere der Bedeutung von Pitchakzenten – behandeln die Untersuchungen von Büring (1995) und Steedman (1997). Beide Arbeiten stehen in der langen Tradition der Analyse des Zusammenhangs zwischen phonologischer Prominenz und Informationsstruktur, wobei Büring diese Problematik aus einer semantischen, Steedman dagegen aus einer syntaktischen Perspektive untersucht. Phonologisch-phonetische Aspekte werden von beiden Arbeiten nicht behandelt. Ich werde in diesem Abschnitt ausschließlich und nur in Form eines kurzen Überblicks auf diejenigen Punkte eingehen, die von den beiden Autoren bezüglich der informationsstrukturellen Funktion von Pitchakzenten angeführt werden; nicht diskutieren werde ich dagegen die umfangreichen semantischen bzw. syntaktischen Teile der beiden Arbeiten. Dieser Abschnitt dient in erster Linie dazu, das Bild der von mir im Rahmen dieser Arbeit bevorzugten Herangehensweise an die Analyse intonatorischer Bedeutung zu ergänzen und zu vervollständigen. Es handelt sich jedoch hierbei nicht um eine direkte Grundlage der in Teil III vorgestellten Experimente.

Das klassische Thema der Analyse des Zusammenhangs zwischen phonologischer Prominenz und Informationsstruktur ist die Akzentzuweisung. Der zentrale Begriff dieser Forschung ist das informationstheoretische Konzept FOKUS (s. Abschnitt 6.1.2) und die einschlägigen Arbeiten setzten sich unter anderem mit den folgenden Zusammenhängen auseinander:

- Fokus und Akzentuierung
- Fokusdomäne, Fokusexponent und Akzentuierung
- Fokus, Kontext und Deakzentuierung

Der Forschungseinsatz auf diesem Gebiet ist beträchtlich und die Einschätzung liegt nahe, daß es sich hierbei mittlerweile um einen eigenen, relativ selbstständigen Zweig

der syntaktisch-phonologischen Forschung handelt. Es soll hier auch gar nicht der Versuch unternommen werden, die umfangreiche Literatur zu diesem Thema auszuwerten und zusammenzufassen, der interessierte Leser wird stattdessen auf Kuhn (1996) verwiesen, wo einige Entwicklungen der letzten Jahre kritisch diskutiert werden und insbesondere auf die Problematik von Kontexteinflüssen und Deakzentuierung eingegangen wird.

Weshalb Büring (1995) und Steedman (1997), die ebenfalls diesem Forschungszweig zuzuordnen sind, hier dennoch besondere Erwähnung finden, liegt daran, daß diese Arbeiten neben der klassischen Fragestellung – Akzent*position* – auch unterschiedliche Akzent*typen* und deren Funktion (‚Semantik') hinsichtlich der Informationsstruktur behandeln. Beide Arbeiten machen diesbezüglich sehr ähnliche Aussagen. Ich werde hauptsächlich Bürings Ansatz darstellen, da er speziell auf die Intonation im Deutschen eingeht, und dann kurz auf die Unterschiede zwischen Bürings und Steedmans Ansatz hinweisen.

Das Hauptinteresse Bürings gilt der Analyse von Satztopiks (S-Topiks[15]). Ausgehend von einem dreiteiligen Modell der Informationsstruktur eines Satzes (Topik/Fokus/Hintergrund) entwickelt Büring eine Topiksemantik, die sich sehr stark an der Rooth'schen Fokussemantik orientiert (Rooth, 1985, 1992; s. Abschnitt 6.1.2), so daß beide Komponenten – Topik und Fokus – über eine vergleichbare, parallele formale Interpretation verfügen. (27)B illustriert die Gliederung eines Satzes in Topik ($[\ldots]_T$), Fokus ($[\ldots]_F$) und Hintergrund (aus Büring, 1995: 56):

(27) A: Was hatten die Popstars an?
B: Die $[\text{weiblichen}]_T$ Popstars trugen $[\text{Kaftane}]_F$.

[15] Es ist zu unterscheiden zwischen S-Topik und Diskurstopik (D-Topik). Das D-Topik ist – vereinfacht gesagt – eine semantische Repräsentation dessen, wovon ein Diskurs handelt, d.h. das D-Topik ist eine diskursstrukturelle Kategorie (vgl. Kapitel 7) und daher der Satzsemantik gewissermaßen übergeordnet. Das S-Topik ist dagegen die Repräsentation eines (optionalen) Teils eines Satzes und als informationsstrukturelle Kategorie zu verstehen (vgl. Büring, 1995: 47f).

Die in unserem Zusammenhang interessante These von Büring ist nun, daß – zumindest im Deutschen – sowohl Fokus als auch Topik durch einen jeweils spezifischen Pitchakzent ausgezeichnet werden. Im Falle des Fokusexponenten ist dies der Standardakzent für neue Information, H*+L, während der Topikexponent durch einen steigenden Akzent, L*+H, ausgezeichnet wird (Büring, 1995: 46ff)[16]. Ohne die intonatorische Markierung mithilfe eines ‚Topikakzents', d.h. ohne kenntlich gemachtes Topik ist der Diskurs in (27) nicht wohlgeformt (Büring, 1995: 56):

(28) A: Was hatten die Popstars an?
B': Die weiblichen Popstars trugen [Kaftane]$_F$.
H*+L

Damit die unvollständige Antwort in B und B' eine adäquate Antwort auf die Frage in A darstellt, muß ein Topik markiert werden – im vorliegenden Beispiel ein sogenanntes *partial topic* (Büring, 1995: 50). Büring identifiziert insgesamt drei Topiktypen, neben *partial topic* noch *contrastive topic* und *purely implicational topic* (Büring, 1995: 49ff). (29) ist ein Beispiel für *contrastive topic*; auch in diesem Fall wäre eine Antwort ohne Topik bzw. Topikmarkierung nicht adäquat.

(29) A: Welches Buch würde Fritz kaufen?
B: [Ich]$_T$ würde ['Das Hotel New Hampshire']$_F$ kaufen.
L*+H H*+L

Parallel zur Repräsentation des Fokus als eine Menge alternativer Propositionen (Rooth, 1992 und Abschnitt 6.1.2) schlägt Büring (1995: 56ff) die Repräsentation des S-Topik als eine Menge von Fragen vor, wobei der Fokus durch eine W-Phrase repräsentiert ist,

[16] Büring verwendet eine etwas andere Notation (H-L* statt H*+L und L-H* statt L*+H), benutzt aber im übrigen durchgehend die Begriffe *falling* bzw. *rising accent*. Im Sinne einer konsistenten Darstellung und ohne dadurch Bürings Analyse zu verfälschen, behalte ich jedoch die in den vorherigen Abschnitten eingeführte Notation bei.

während das Topik durch eine (kontextuell restringierte) Menge von Alternativen ersetzt wird. Somit ergibt sich für (29)B die Topikrepräsentation bzw. der ‚Topikwert' in (30) (Büring, 1995: 57).

(30) {which book would you/I buy, which book would Bolle buy, which book would Fritz buy, which book would Fritz's brother buy, …}

Die Wohlgeformtheitsbedingung für Topikkonstruktionen der Frage/Antwort-Form formuliert Büring (1995: 58) folgendermaßen:

(31) Frage/Antwort-Bedingung:
Die Bedeutung der Frage muß mit einem Element des Topikwertes der Antwort übereinstimmen.

Diese Bedingung trifft offensichtlich auf (29)/(30) zu. Betrachten wir nun (27); der Topikwert von (27)B ist die Menge in (32) (Büring, 1995: 59f):

(32) {what did the female popstars wear, what did the male popstars wear, what did the female or male popstars wear, what did the italian popstars wear, …}

Das Element {what did the female or male popstars wear}, die *trivial property*, stimmt in diesem Fall mit der ursprünglichen Fragebedeutung überein (*what did the popstars wear*), es handelt sich also ebenfalls um eine wohlgeformte Topikkonstruktion. Das Problem von (28)B' ist, daß die Bedingung in (31) nicht erfüllt wird. Da kein S-Topik markiert ist, besteht der Topikwert von B' nur aus einem Element: {what did the female popstars wear}[17]. Dies stimmt offensichtlich nicht mit der ursprünglichen Fragebedeutung (*what did the popstars wear*) überein.

[17] Dies folgt aus Bürings Herleitung der Fragerepräsentation des Topikwertes; siehe Büring (1995: 56ff).

Ein weiterer Aspekt der Büring'schen Topiksemantik ist es, daß Topiks – egal welchen Typs – eine Implikatur auslösen. Diese Implikatur kann folgendermaßen charakterisiert werden ($[\![A]\!]^t$ repräsentiert den Topikwert von A):

(33) „Given a sentence A, containing an S-Topic, there is an element Q in $[\![A]\!]^t$ such that Q is still under consideration after uttering A.“ (Büring, 1995: 60)

Es gibt also eine Frage Q aus der Menge der Fragen in $[\![A]\!]^t$, ‚über die noch geredet werden kann/muß‘. Dieses Element Q nennt Büring *residual topic*. Für (29) – die *contrastive topic*-Konstruktion – gilt somit, daß ein Element aus (30) das *residual topic* ist. Dies trifft zu auf {which book would Fritz buy}, also das Bedeutungsäquivalent der ursprünglichen Frage, welche durch (29)B nicht beantwortet wurde. In (27)/(32) – die *partial topic*-Konstruktion – ist das *residual topic* {what did the male popstars wear}. Damit wird die Information, die zu einer vollständigen Antwort auf die ursprüngliche Frage fehlt, erfragt.

Nach diesen Ausführungen kann nun auch der dritte von Büring identifizierte Topiktyp – *purely implicational topic* – illustriert werden:

(34) A: Hat Peters Frau fremde Männer geküßt?
B: $[\text{Peters}]_T$ Frau hat $[\text{keine}]_F$ fremden Männer geküßt.
L*+H H*+L
$[\![B]\!]^t$={did Peter's wife kiss other men, did Fritz's wife kiss other men, did John's wife kiss other men, …}

Die Konstruktion ist wohlgeformt, da die Bedeutung von (34)A (*did Peter's wife kiss other men*) in $[\![B]\!]^t$ ‚enthalten‘ ist. Es ist jedoch die Besonderheit dieser Konstruktionen, daß sie auch ohne S-Topik wohlgeformt sind. Das einzige Element in $[\![B']\!]^t$ zu (35) ist

{did Peter‘s wife kiss other men}, also die Entsprechung der Fragebedeutung. Dadurch ist die Wohlgeformtheitsbedingung erfüllt.

(35) B': Peters Frau hat $[\text{keine}]_F$ fremden Männer geküßt.
H*+L

Die einzige Funktion des S-Topik bzw. des Topikakzents in (34) ist somit die Auslösung der Implikatur eines *residual topic*, daher die Bezeichnung *purely implicational topic*. (34)B beantwortet nicht nur die Frage in A, sondern impliziert gleichzeitig, daß es evtl. andere Frauen gibt, (die im gegebenen Diskurskontext von irgendeiner Bedeutung sind und) die durchaus fremde Männer küssen (z.B. Fritz' Frau, Johns Frau etc.). (35)B' hat diesen zusätzlichen Bedeutungsaspekt nicht und beantwortet ausschließlich die ursprünglich gestellte Frage.

An diesem Beispiel wird besonders deutlich, daß die von Büring postulierte Funktion des L*+H Akzents zwar einen etwas anderen Aspekt in den Vordergrund stellt, insgesamt jedoch keineswegs einen Widerspruch zu den Analysen von Pierrehumbert & Hirschberg bzw. Hobbs darstellt. Ein Topik ist stets in irgendeiner Form mit dem vorangegangenen Diskurskontext verknüpft und kann daher – im Gegensatz zu fokussiertem Material – als *not new* im Sinne von Hobbs bezeichnet werden. In (34)B ist diese Verknüpfung in ihrer deutlichsten Form realisiert: Das S-Topik ist die wörtliche Wiederholung von gegebenem Material. Diese Eigenschaft des S-Topik lizenziert den L-Sternton. Daneben ist ein fester Bestandteil der Büring'schen Topiksemantik die Implikatur eines *residual topic*, d.h. der Sprecher signalisiert, daß ein Aspekt dessen, worüber er etwas aussagt, noch ‚offen‘ ist. In (34)B sind dies die anderen Frauen, über deren Einstellung bezüglich des Küssens fremder Männer möglicherweise etwas ganz anderes ausgesagt werden könnte. Dies ist kompatibel mit der Hobbs'schen Analyse des H-Suffixes, welches *openendedness* bzw. *incompleteness* markiert.

In der Arbeit von Steedman (1997), die wie bereits erwähnt eine eher syntaktische denn semantische Perspektive einnimmt, liegt der Schwerpunkt naturgemäß weniger auf einer Analyse der Semantik von Topik und Fokus, als vielmehr auf der Untersuchung des Zusammenhangs zwischen Informationsstruktur, syntaktischer Phrasierung und phonologischer Phrasierung. Steedman beschreibt einen Weg der vollständigen Unifikation von intonatorischer Struktur und syntaktischer (Oberflächen-) Struktur und argumentiert dafür, daß beide Strukturen nur verschiedene Aspekte der selben Derivationsstruktur sind. Ausgangspunkt seiner Analyse bildet eine informationsstrukturelle Gliederung, die sich von Bürings Gliederung in einigen Punkten unterscheidet. Im Gegensatz zu Bürings ‚flachem' Modell nimmt Steedman ein 2-stufiges Modell an: Ein Satz besteht zunächst aus zwei Komponenten, Thema (*theme*, θ) und Rhema (*rheme*, ρ), welche sich jeweils weiter untergliedern lassen in Fokus (*focus, f*) und Hintergrund (*background, bg*) (Steedman, 1997: 5ff). Thema und Rhema sind beides Domänen (vergleichbar der Topik- bzw. Fokusdomäne) und der Fokus entspricht jeweils dem intonatorisch markierten, d.h. akzentuierten Exponenten einer Domäne. Nicht-akzentuiertes Material bildet den Hintergrund. Ein Beispiel aus Steedman (1997: 8):

(36) A: I know that Mary envies the man who wrote the musical. But who does she like?

B: Mary	LIKES	the woman who	DIRECTED	the musical
bg	*f*	*bg*	*f*	*bg*
θ		ρ	

(Der Begriff Fokus, wie Steedman ihn verwendet, bezieht sich tatsächlich nur auf das jeweils akzentuierte Wort, phrasale Einheiten werden ausschließlich mit den Begriffen Thema und Rhema bezeichnet.)

Bezüglich der Funktion von Pitchakzenten kommt Steedman, der im übrigen nur die Intonation des Englischen bzw. des amerikanischen Englisch betrachtet, zu einem ähnlichen Schluß wie Büring. In seinem Modell besteht ihre Funktion ebenfalls darin, das Thema (Büring: Topik) und das Rhema (Büring: Fokus) auszuzeichnen und auch die

Klassifikation der Akzenttypen ist mit der Büring'schen Klassifikation vergleichbar: Die steigenden Akzente L+H* und L*+H kennzeichnen thematische Foki, die (überwiegend) fallenden Akzente H*, L*[18], H*+L und H+L* kennzeichnen rhematische Foki (Steedman, 1997: 18). Die akzentuierten Wörter in (36) tragen also zwei unterschiedliche Akzente, einen steigenden (L+H*) auf *likes* und einen fallenden (H*+L) auf *directed* (Steedman, 1997: 8).

Das besondere Augenmerk Steedmans gilt jedoch der Korrelation zwischen Informationsstruktur einerseits und intonatorischer bzw. syntaktischer Phrasierung andererseits. Das an Grenzmarkierungen reiche Intonationssystem des amerikanischen Englisch erlaubt eine an der Informationsstruktur orientierte phonologische Phrasierung von Sätzen. Für (36) kann z.B. die folgende Phrasierung angenommen werden:

(37) (Mary LIKES) (the woman who DIRECTED the musical)
L+H* L-H% H*+L L-L%

Die Problematik dieser Phrasierung liegt darin, daß die erwünschte Übereinstimmung zwischen phonologischer und syntaktischer Phrasierung in diesem Fall nicht zu existieren scheint, da die phonologische Grenze nach *likes* das finite Verb vom direkten Objekt trennt. Dieses Problem besteht in klassischen, rechtsverzweigenden Syntaxmodellen. Lösungsmöglichkeiten bietet u.U. der Rückgriff auf die Tiefenstruktur oder die Logische Form. Steedman präsentiert dagegen eine syntaktische Analyse, die eine direkte Korrelation zwischen intonatorischer Phrasierung und syntaktischer Oberflächenstruktur erlaubt. Mithilfe des Kalküls der kombinatorischen Kategorialgrammatik (Steedman, 1987, 1996) lassen sich für einen gegebenen Satz mehrere, durch verschiedene grammatische Prinzipien restringierte Strukturen erzeugen, die sich zwar in der syntaktischen Phrasierung, nicht jedoch im Ergebnis – einer Proposition in Form einer Prädikat-Argument-Struktur – unterscheiden. Die Funktion der phonologischen Phrasierung, d.h. der intona-

[18] Für eine Diskussion des nicht-fallenden Akzents L* siehe Steedman (1997: 13).

torischen Markierung von Phrasengrenzen ist es dann, die Menge der möglichen Strukturvarianten zu disambiguieren. Ein Beispiel:

(38) Mary likes musicals.

Steedman (1997: 17) gibt für diesen Satz zwei Analysen an, die zwar in der identischen Proposition terminieren, aber zu zwei verschiedenen Phrasierungen führen: Zum einen erhält man die Standardstruktur in (39)a und zum anderen die linksverzweigende Struktur in (39)b.

(39) a. (Mary) (likes (musicals))
b. ((Mary) likes) (musicals)

Mithilfe der Intonation, die wiederum die Informationsstruktur reflektiert, kann nun zwischen diesen beiden Strukturen disambiguiert werden. Im Kontext von (40) wird das Subjekt Thema (Fokus: *Mary*) und die Verbalphrase Rhema (Fokus: *musicals*). Die intonatorische Markierung identifiziert somit die syntaktische Standardstruktur (39)a.

(40) What about MUSICALS? Who likes THEM?
(MARY) (likes MUSICALS)
H* L- L+H* L-H%

In (41) ist dagegen die ‚Konstituente' *Mary likes* Rhema und das direkte Objekt Thema. Die intonatorische Struktur reflektiert diese Gliederung und identifiziert damit die syntaktische Struktur in (39)b.

(41) What about MARY? What does SHE like?
(MARY likes) (MUSICALS)
L+H* L-H% H* L-L%

Ich werde an dieser Stelle meine Darstellung der Phonologie der Intonation und der TSM-basierten Ansätze zur Analyse intonatorischer Bedeutung beenden. Ich möchte nochmals betonen, daß das Anliegen dieses Kapitels nicht darin bestand, einen vollständigen Überblick über die Entwicklungen der phonologischen und/oder semantischen Forschung innerhalb der letzten Jahre zu geben. Insbesondere einige phonologische Ansätze im Bereich der Intonationsphonologie (IPO, KIM etc.) aber auch einige Ansätze der phonologisch-semantischen Forschung blieben unberücksichtigt. Dieses Kapitel sollte vielmehr dazu dienen, in die Thematik dieser Arbeit einzuführen und den theoretischen Rahmen für die experimentellen Kapitel in Teil III abzustecken. Dies habe ich versucht zu tun, indem ich in den Abschnitten 2.1 und 2.2 diejenige Theorie diskutiert habe, die – zumindest was den phonologischen Teil dieser Arbeit betrifft – die wichtigste Basis meiner Untersuchungen bildet – das Ton-Sequenz-Modell, und indem ich in Abschnitt 2.3 einige meines Erachtens wesentliche Analysen aus dem Bereich der Prosodie-Semantik-Schnittstelle dargestellt habe.

Damit ist Teil I meiner Arbeit abgeschlossen, der insgesamt einführenden Charakter hatte. Im nächsten Teil, in den Kapiteln 3 und 4, werde ich zwei Werkzeuge der phonologischen Analyse bzw. des experimentellen Arbeitens vorstellen, die auf dem Ton-Sequenz-Modell basieren.

Teil II

Analysewerkzeuge

3 Prosodische Annotation

Der Mangel eines Standardverfahrens für die prosodische Transkription – wie es etwa das IPA-System für die segmentale Transkription darstellt – war die Motivation für eine Gruppe von amerikanischen Forschern, ihre unterschiedlichen Erfahrungen aus den Bereichen der Prosodieanalyse, der Spracherkennungs- und Syntheseforschung und der natürlichsprachlichen Systeme zusammenzutragen, und eine Transkriptionskonvention zu entwickeln, die die wichtigsten prosodischen Parameter abdecken sollte. Teil dieser Forschergruppe waren mit Janet Pierrehumbert und Mary Beckman auch die wichtigsten Repräsentanten des Ton-Sequenz-Ansatzes. Insofern ist es nicht verwunderlich, daß das Ergebnis, das Anfang der 90er Jahre präsentiert wurde, wesentliche Erkenntnisse des Ton-Sequenz-Modells widerspiegelt und implementiert hat. TOBI – für *Tones and Break Indices* – basiert auf der TSM-Analyse der Intonation des amerikanischen Englisch und stellt, abgesehen von einigen wenigen Modifikationen, die Implementierung des Pierrehumbert'schen Intonationsmodells dar. Ergänzt wird die Intonationskomponente um eine *Break*-Komponente zur Annotation prosodischer Dissoziation und um eine Komponente für ‚Sonstiges', also lautsprachliche Phänomene, die die prosodische Analyse in irgendeiner Form überlagern.

3.1 Das TOBI-System für amerikanisches Englisch

Insgesamt besteht TOBI aus vier unabhängigen Komponenten bzw. Transkriptionsebenen, die zwar einige wesentliche, aber nicht alle Aspekte der Prosodie erfassen sollen (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg, 1992; Beckman & Ayers, 1994; Beckman & Hirschberg, 1994). Neben der Intonationskomponente – *tone tier* –, der *Break*-Komponente – *break index tier* – und der Komponente für Sonstiges – *miscellaneous tier* – ist die vierte Ebene die der orthographischen Transkription – *orthographic tier*. Zu der prosodischen Transkription im engeren Sinne zählen nur das *tone tier* und das *break index tier*, die orthographische Transkription dient dagegen in erster Linie der Referenz zum Inhalt der analysierten Äußerung, während die *miscellaneous*-Transkription Phänomene erfaßt, die zwar für die Interpretation der prosodischen Analyse u.U. relevant sind, selbst jedoch nicht Teil der Prosodie sind. Dies sind Phänomene wie Lachen, Husten, hörbares Ein- oder Ausatmen, verschiedene Arten von *disfluencies* und ähnliches, die selbst keine linguistische Funktion haben, aber die Realisierung der linguistischen Prosodie beeinflussen bzw. stören können. Das Inventar der *miscellaneous*-Komponente von TOBI erhebt nicht den Anspruch auf Vollständigkeit, es bleibt vielmehr jedem Anwender überlassen, eigene ‚Kategorien' zu definieren (Silverman *et al.*, 1992). Auch die Grundstruktur der vier parallelen Transkriptionsebenen ist offen für Erweiterungen, sie stellt lediglich das Kernsystem dar, welches für spezielle Anwendungen um weitere *tiers* ergänzt werden kann (Beckman & Ayers, 1994: 8). Eine solche Erweiterung wird im Verlaufe dieser Arbeit tatsächlich vorgeschlagen: das *register tier* (s. Kapitel 7).

3.1.1 *Break*-Indizes

Mit *break indices* wird die Stärke der prosodischen Dissoziation zwischen benachbarten Wörtern repräsentiert, wobei mit Wörtern alle auf der orthographischen Ebene gelabelten Einheiten gemeint sind. So gelten z.B. Kontraktionen (*do not* → *don't*), sofern sie orthographisch als solche transkribiert sind, als eine Einheit. Generell soll mit den *break indices* zwar die subjektiv wahrgenommene Stärke einer Dissoziation annotiert werden (Beckman & Ayers, 1994: 31), die Transkription sollte sich jedoch an den im folgenden kurz dargestellten mehr oder weniger objektiven Kriterien orientieren.

Insgesamt stehen 5 Indizes zur Verfügung, basierend auf Untersuchungen von Price, Ostendorf, Shattuck-Hufnagel & Fong (1991). Es sind dies die Indizes 0, 1, 2, 3 und 4, wobei im Prinzip 0 für die schwächste und 4 für die stärkste Dissoziation steht (Beckman & Ayers, 1994: 31). ‚Normale', phrasen-interne Wortgrenzen werden mit 1 transkribiert, Wortgrenzen innerhalb klitischer Gruppen mit 0. 3 steht für eine Wortgrenze, die mit einer intermediären Phrasengrenze zusammenfällt, 4 für eine Wortgrenze, die mit einer Intonationsphrasengrenze zusammenfällt. Die Indizes 3 und 4 stellen also keine unabhängige Repräsentation prosodischer Dissoziationsmerkmale dar (z.B. finale Längung, Pausen etc.), sondern sind sehr stark mit der intonatorischen Analyse verknüpft. Die Indizes 3 bzw. 4 können nur dann verwendet werden, wenn intonatorische und nicht-intonatorische Grenzmerkmale zusammenfallen, wobei für die Entscheidung zwischen 3 und 4 allein die intonatorischen Merkmale relevant sind. Dies bedeutet, daß z.B. eine ip-Grenze, die durch sehr starke finale Längung gekennzeichnet ist, mit dem Index 3 versehen wird, während eine IP-Grenze, die kaum finale Längung aufweist, mit 4 transkribiert wird.

Wenn die intonatorische und nicht-intonatorische Markierung einer Grenze nicht wie erwartet übereinstimmt, wird der Index 2 verwendet. Dies ist dann der Fall, wenn 1.) nicht-intonatorische Merkmale wie Längung oder Pausen vorhanden sind, jedoch kein intonatorischer Hinweis auf das Ende einer ip oder IP vorliegt oder wenn 2.) eine intonatorisch (durch Phrasenakzent oder Terminalton) deutlich markierte ip-/IP-Grenze erkennbar ist, die nicht-intonatorischen Merkmale jedoch schwächer als erwartet ausfallen.

Die Indizes 1, 2 und 3 können mit einem Diakritikum p versehen werden. Mit 1p werden Abbrüche innerhalb oder unmittelbar nach einem Wort transkribiert, z.B. vor einer Verbesserung; häufig fällt der *Break*-Index 1p mit dem tonalen Label %r zusammen, welches den Beginn einer neuen IP nach einer abgebrochenen IP markiert. 2p und 3p stehen bei Hesitationen – Pausen und/oder starken Längungen (vgl. Beckman & Ayers, 1994: 35ff).

Nach ToBI-Konvention werden *break indices* am Ende eines Wortes (i.e. einer orthographisch transkribierten Einheit) gelabelt. Es gibt keinen *default*-Index, d.h. jede orthographisch gelabelte Einheit muß auf der *Break*-Ebene mit einem Index versehen werden (Beckman & Hirschberg, 1994: 1).

3.1.2 Töne

Das *tone tier* repräsentiert die intonatorische Analyse im Sinne des Ton-Sequenz-Modells, d.h. hier werden Pitchakzente, Phrasenakzente und Grenztöne transkribiert. Das Inventar dieser Komponente ist sprachspezifisch, daher gehe ich hier nur auf die allgemeinen Labellingkonventionen und kurz auf einige Unterschiede zu früheren Versionen der TSM-Analyse des amerikanischen Englisch ein. Ein Inventar zur Annotation der Intonation des Deutschen werde ich in Abschnitt 3.2 vorschlagen.

3.1.2.1 Grenztöne

Terminaltöne und Phrasenakzente werden stets am Ende der Domäne, auf die sie sich beziehen, gelabelt, d.h. Terminaltöne am Ende einer IP und Phrasenakzente am Ende einer ip. Am Ende einer Phrase (IP oder ip) erhält man also in der Regel drei synchrone Labels in drei parallelen ToBI-Komponenten: auf der orthographischen Ebene die Transkription des letzten Wortes der Phrase, auf der *Break-Index*-Ebene einen *Break*-Index und auf der tonalen Ebene einen Phrasenakzent (ip) bzw. ein komplexes Label aus Phrasenakzent plus Terminalton (IP).

Neben den terminalen Grenztönen können in ToBI auch initiale Intonationsphrasengrenzen transkribiert werden. Hierfür stehen zwei Labels zur Verfügung: %H und %r.

Am Anfang einer Intonationsphrase startet die Tonhöhe in der Regel in der unteren Hälfte des Stimmumfangs eines Sprechers. Dies ist der unmarkierte Fall und wird in ToBI nicht transkribiert (Beckman & Hirschberg, 1994). Wenn eine Phrase dagegen relativ hoch beginnt, soll dies mit %H am Anfang der Phrase annotiert werden.

Mit %r wird der Anfang einer Phrase annotiert, wenn die vorherige Phrase unterbrochen wurde und eine komplett neue Phrase beginnt; %r fällt daher meist mit dem *Break*-Index 1p zusammen. Das Label %r verfügt über keine tonal-phonetische Interpretation.

3.1.2.2 Pitchakzente

Pitchakzente werden im Nukleus der akzentuierten Silbe gelabelt. Mit ToBI für amerikanisches Englisch können fünf verschiedene Akzenttypen annotiert werden. Aus dem Original-TSM wurden die folgenden vier Typen übernommen: H*, L*, L*+H und L+H*. Auf den Akzent H*+L, der in Pierrehumberts System in erster Linie dazu diente, *downstep* (des nächsten H-Tons) auszulösen, wurde verzichtet. In ToBI wird *downstep* stattdessen durch das Diakritikum „!“ repräsentiert, welches jedem beliebigen nicht-initialen H-Ton vorangestellt werden kann, sofern er von *downstep* betroffen ist. Während also in Pierrehumberts Modell die Sequenz H*+L H* den *downstep* des zweiten H-Tons repräsentiert, wird dies in ToBI mit H* !H* transkribiert. Der Grund für diese Modifikation ist die verbesserte phonetische Transparenz der tonalen Labels, denn ohne elaborierte Kenntnisse des Pierrehumbert'schen Modells könnte der Akzent H*+L sehr leicht mißverstanden werden als *peak*-Akzent mit unmittelbar folgendem Pitchabfall. Dieser (sehr häufige) Pitchverlauf wird jedoch sowohl von Pierrehumbert als auch in ToBI durch H* L-, also einer Kombination aus Pitchakzent und Phrasenakzent repräsentiert. Die phonetische Transparenz des Symbolinventars eines Labellingsystems für natürliche Sprache ist ein wichtiger Aspekt, sowohl hinsichtlich des Trainings von Transkribenten (bessere Erlernbarkeit, größere Konsistenz) als auch hinsichtlich der Verwendung gelabelter Korpora in der Entwicklung automatischer Erkennungssysteme (Silverman *et al.*, 1992).

Der fünfte Akzenttyp in ToBI ist H+!H*, der den ‚alten‘ H+L* Akzent ersetzt. Auch dies bedeutet einen Zugewinn an phonetischer Transparenz. H+L* repräsentiert in Pierrehumberts Modell Konturverläufe mit einem lokalen F_0-Maximum in der präakzentuierten Silbe, gefolgt von einem tieferen Pitchlevel in der akzentuierten Silbe. Das relevante Charakteristikum dieses Akzenttyps ist jedoch allein der im Vergleich zu der präakzentuierten Silbe tiefere Pitchlevel in der akzentuierten Silbe. Dieser tiefere Ton muß nicht notwendigerweise am unteren Rand des Stimmumfangs realisiert werden und stellt insofern kein echtes L-Target dar. Durch eine Sequenz aus einem H-Ton (assoziiert mit der präakzentuierten Silbe) und einem *downstepped* !H-Sternton (assoziiert mit der akzentuierten Silbe) – also H+!H* – wird dieser (dem deutschen *early peak* sehr ähnliche) Konturverlauf angemessener repräsentiert.

3.1.3 Weitere Merkmale von ToBI

Mitverantwortlich dafür, daß sich ToBI so schnell als Quasi-Standard für die prosodische Transkription etablieren konnte, sind einige weitere Merkmale des Systems, die sowohl was die technische Realisierung von Transkriptionsvorhaben als auch was die Schulung von Transkribenten betrifft von großem Vorteil sind. Dies ist z.B. die ASCII-Kompatibilität des gesamten ToBI-Inventars, die freie Verfügbarkeit von ToBI-Labellingwerkzeugen für die *Entropic/waves*™-Umgebung oder die sehr einfach zu realisierende Möglichkeit, bestimmte Korpora nur mit einer Teilmenge der ToBI-Notation oder auch mit zusätzlichen Labels zu annotieren. Außerdem kann mit den Labels *, - bzw. % tonale Unterspezifikation kodiert werden. Dies ist z.B. dann sinnvoll, wenn für ein Korpus zunächst nur die Position von tonalen Ereignissen, nicht jedoch deren Typ von Interesse ist. Ein wichtiger Aspekt für die Schulung von Transkribenten und für eine möglichst große Konsistenz von Transkriptionen stellt der Umgang mit Unsicherheiten dar. ToBI erlaubt eine relativ umfangreiche und genaue Kodierung von Unsicherheit in der tonalen Transkription, um so die Nachteile zu vermeiden, die sich aus dem Zwang zu einer Alles-oder-Nichts Entscheidung ergeben. Zwei Stufen der Unsicherheit können in ToBI unterschieden werden: 1.) Unsicherheit, ob überhaupt ein bestimmtes tonales Ereignis (Pitchakzent, Phrasenakzent, Terminalton) vorliegt oder nicht und 2.) wenn ein bestimmtes tonales Ereignis erkannt wird, Unsicherheit bezüglich des tonalen Werts dieses Ereignisses (Pitchakzenttyp, Phrasenakzenttyp, Terminaltontyp) (Beckman & Hirschberg, 1994: 4f).

Unsicherheit des ersten Typs wird mit *?, -? bzw. %? transkribiert, Unsicherheit des zweiten Typs mit X*?, X-? bzw. X%?.

3.2 Das ‚Stuttgarter System': GToBI(S)

3.2.1 GToBI

Als Konsequenz aus der rasanten Entwicklung korpusbasierter Methoden in verschiedensten Bereichen der Linguistik entstanden in den letzten Jahren eine ganze Reihe von Transkriptions- und Labellingsystemen zur Annotation der Prosodie des Deutschen. Diese zum Teil stark differierenden Ansätze wurden 1995 bei einem Workshop in Stockholm vorgestellt und diskutiert (Mayer, 1995b). Im Rahmen der Vorbereitung dieses Workshops entstand eine Zusammenarbeit zwischen den drei Arbeitgruppen, die zur Annotation der deutschen Intonation den levelbasierten ToBI-Ansatz verfolgen. Diese Zusammenarbeit zwischen der Saarbrücker Gruppe (GToBI(SB), Grice & Benzmüller, 1995), der Verbmobil-Gruppe (GToBI(VM), Reyelt & Batliner, 1994) und der Stuttgarter Gruppe (GToBI(S), Mayer, 1995a) führte zur Entwicklung eines gemeinsamen Labellingsystems für die Transkription der Intonation des Deutschen: G(erman)ToBI (Reyelt, Grice, Benzmüller, Mayer & Batliner, 1996). GToBI ist ein Konsenssystem, das auf den Gemeinsamkeiten der drei Einzelsysteme basiert, die sich aufgrund der unterschiedlichen Schwerpunkte der drei Arbeitsgruppen in einigen Punkten unterscheiden. Daß dieses Konsenssystem gelungen ist, konnte mithilfe eines Experimentes zur *Inter-Transcriber*-Variabilität nachgewiesen werden (Reyelt *et al.*, 1996; Grice, Reyelt, Benzmüller, Mayer & Batliner, 1996). An dem Experiment nahmen insgesamt 13 Transkribenten aus Saarbrücken, Braunschweig (Verbmobil) und Stuttgart teil, die jeweils mit dem lokalen System geschult worden waren. Alle Transkribenten wurden vor dem Experiment mit GToBI vertraut gemacht. Das Material, das zu labeln war, bestand aus 304 Sekunden Sprache (733 orthographische Wörter). Die Konsistenz zwischen den Transkribenten für

die Entscheidung akzentuiert/nicht-akzentuiert betrug 87%, für das Pitchakzent-Labelling 71% und für das Grenzton-Labelling 86% (Grice *et al.*, 1996). Diese Ergebnisse sind vergleichbar mit den Ergebnissen eines ähnlichen Konsistenz-Experiments für das englische ToBI (Pitrelli, Beckman & Hirschberg, 1994). Die Tatsache, daß obwohl die Transkribenten mit drei verschiedenen Systemen geschult waren, ein reliables Ergebnis für das Labelling mit GToBI erzielt wurde, zeigt, daß GToBI als Konsenssystem einen sinnvollen, anwendbaren Kompromiß darstellt.

Parallel zu der Entwicklung von GToBI wurden jedoch auch von allen drei Arbeitsgruppen die lokalen Systeme weiterentwickelt, orientiert an den lokalen Forschungsschwerpunkten und Anwendungsbereichen. Das Ergebnis dieser Entwicklung in Stuttgart – das ‚Stuttgarter System' GToBI(S) – ist Gegenstand der folgenden Abschnitte.

3.2.2 Die nicht-tonalen Komponenten

Die wesentlichen Unterschiede zwischen GToBI(S) und ToBI sind naturgemäß in der tonalen Komponente zu finden. Die wenigen Differenzen in den nicht-tonalen Komponenten, die hier beschrieben werden, sind weniger theoriebasiert, als vielmehr pragmatische Modifikationen, die während der Arbeit an den Daten entstanden sind.

In der orthographischen Komponente dient das Label <P> in GToBI(S) zur expliziten Annotation von Pausen. Zum einen können die gelabelten Daten dadurch besser als Trainingsmaterial für Erkennungssysteme verwendet werden (Rapp, 1995), zum anderen ist es dadurch möglich, Wortdauern auch von Wörtern nach Pausen aus den Labeldateien zu extrahieren[1].

Das Inventar der *Break*-Komponente ist identisch mit dem ToBI-Inventar, die Konventionen für die Verwendung der Kategorien wurden jedoch geringfügig verändert.

[1] Die ToBI-Konventionen für das Labelling in der *waves*-Umgebung schreiben vor, das Label für orthographische Einheiten immer mit dem Ende der transkribierten Einheit zu alinieren. Die Dauer eines Wortes ergibt sich dann aus der Differenz zwischen dem Zeitpunkt des Labels für dieses Wort und dem Zeitpunkt des vorangehenden Labels. Bei Wörtern nach Pausen gilt dies jedoch nur dann, wenn auch das Pausenende gelabelt wird.

Break-Index 1 für normale Wortgrenzen gilt als *default*-Wert und muß nicht explizit gelabelt werden. Dies bedeutet keinen Informationsverlust, da die Anzahl und die Position der *Break*-Indizes aus der orthographischen Transkription rekonstruiert werden kann und alle anderen Indizes (0, 2, 3 und 4) gelabelt werden; d.h. eine Wortgrenze, die mit keinem anderen Index annotiert ist, korreliert mit *Break*-Index 1. Nur wenn *Break*-Index 1 mit dem Diakritikum p versehen wird, muß dies explizit gelabelt werden.

Für die Verwendung der Indizes 3 und 4 werden dem Labeler größere Freiheiten eingeräumt. Die strenge Koppelung an den Phrasentyp (3 nur bei ips, 4 nur bei IPs) erscheint uns redundant und wird folgendermaßen abgeschwächt: Die Indizes 3 und 4 sollen nach wie vor nur am Ende einer Phrase gelabelt werden, jedoch unabhängig vom Phrasentyp; für die Unterscheidung zwischen 3 und 4 ist allein der Grad der Dissoziation relevant. Daten, die auf der Grundlage dieser Konvention gelabelt werden, sind informativer was die Stärke der *Breaks* betrifft, können aber auch sehr einfach rückkonvertiert werden, so daß sie mit den ursprünglichen ToBI-Konventionen kompatibel sind: Da die Information über den Phrasentyp in der tonalen Transkription enthalten ist, können alle ip-korrelierten *Break*-Indizes durch 3 und alle IP-korrelierten Indizes durch 4 ersetzt werden.

Eine weitere Modifikation betrifft sowohl die tonale als auch die *Break*-Index-Komponente. Neben den oben erwähnten zwei Stufen der Unsicherheit erlaubt GToBI(S) die Kodierung einer dritten Stufe: An jedes Label der beiden prosodischen Komponenten kann ein „?“ angehängt werden. Dies bedeutet, daß einige Kriterien für das verwendete Label sprechen, daß aber dennoch Zweifel an seiner Richtigkeit bestehen. Dadurch können z.B. zweifelhafte oder sehr untypische Instanzen bestimmter Kategorien vor dem Training von prosodischen Erkennungssystemen herausgefiltert werden.

Die Modifikationen, die in diesem Abschnitt beschrieben wurden, haben sich in unserer Arbeitgruppe als sinnvoll erwiesen und bewährt, insbesondere im Bereich der automatischen Prosodieerkennung (Rapp, i.V.) und der Entwicklung eines Sprachsynthesesystems (Möhler, i.V.). Ob sie jedoch in das Standard-ToBI übernommen werden sollten, kann und soll hier nicht entschieden werden. Diese Entscheidung muß aber auch nicht

getroffen werden, da die Modifikationen keinerlei Kompatibilitätsprobleme verursachen, und daher auch bei einer Koexistenz von GToBI(S) und Standard-ToBI die Vergleichbarkeit und der Austausch gelabelter Daten gewährleistet ist.

3.2.3 Die tonale Komponente

Die tonale Komponente von GToBI(S) ist – wie bei allen sprachspezifischen ToBI-Adaptionen – der zentrale Teil des Stuttgarter ToBI-Derivats. Sie basiert wesentlich auf Férys TSM-Analyse der Intonation des Deutschen (Féry, 1993; s. Abschnitt 2.2).

3.2.3.1 Grenztöne

In GToBI(S) wird in der Regel nur die Position, nicht die tonale Spezifikation von terminalen ip- oder IP-Phrasengrenzen annotiert. Hierfür stehen die Labels „-“ (für ip-Grenzen) und „%“ (für IP-Grenzen) zur Verfügung. Die tonale Spezifikation kann aus dem Trailington des phrasenfinalen Pitchakzents abgeleitet werden. Diese Reduktion reflektiert den von Féry vorgeschlagenen Verzicht auf Phrasenakzente in der Analyse der deutschen Intonation und den Verzicht auf den Terminalton L%. Die Repräsentation des Pitchverlaufs zwischen dem phrasenfinalen Akzent und der Phrasengrenze, also die wichtigste Aufgabe des Phrasenakzents im Originalmodell, übernimmt in Férys Modell der Trailington. Nach einem steigenden Pitchakzent (H-Trailington) fällt die F_0-Kurve nicht mehr nennenswert ab. Es reicht daher aus, die Phrasengrenze mit „-“ (interpretiert als H-) bzw. mit „%“ (interpretiert als H%) zu annotieren. Nach einem fallenden Akzent (L-Trailington) bleibt die Kontur entweder tief – in diesem Fall wird wieder „-“ (→ L-) bzw. „%“ (→ L%) annotiert – oder es kann ein steigender Terminalton in der letzten Silbe der Phrase realisiert werden. Für diesen einzigen ‚echten‘ Terminalton in Férys Analyse steht das Label H% zur Verfügung. Das GToBI(S)-Inventar an terminalen Grenz‚tönen‘ besteht also nur aus den folgenden drei Kategorien:

(1) Terminale Grenzkategorien in GToBI(S):

a. ip-Grenzen: -

b. IP-Grenzen ohne Grenzton: %

c. IP-Grenzen mit steigendem Grenzton (nur nach fallendem nuklearem Akzent): H%

Daneben wurden aus dem originalen ToBI-Inventar die initialen Grenzkategorien %H und %r übernommen (s.o.).

AUSBLICK — Aufgrund der bisherigen Erfahrungen mit der Transkription größerer Datenmengen[2] und angeregt durch die Diskussion mit anderen GToBI-Entwicklern, soll die folgende Ergänzung vorläufig in GToBI(S) aufgenommen und getestet werden: Die Annotation eines zusätzlichen Anstiegs in der phrasenfinalen Silbe nach einem H-Trailington mit H% (*upstep*).

Die ToBI-Sequenz H-H% repräsentiert einen postakzentualen Konturverlauf an der oberen Grenze des F_0-Umfangs der Phrase, gefolgt von einem zusätzlichen Anstieg in der letzten Silbe. Die ToBI-Entwickler nennen dieses Phänomen *upstep*: „the H- phrase accent causes 'upstep' on the following boundary tone, so that the H% after a H- rises to a very high value“ (Beckman & Hirschberg, 1994: 3). Solche Konturverläufe können auch im Deutschen beobachtet werden (Abb. 3-1). Die Voraussetzung für eine Repräsentation dieses Phänomens in einem phonologisch motivierten Transkriptionssystem wie GToBI(S) ist jedoch, daß der Unterschied zwischen einem hohen Pitchverlauf ohne finalen Anstieg und einem hohen Pitchverlauf mit finalem Anstieg interpretierbar ist. In Hobbs' Modell der Bedeutung von Tönen (s. Abschnitt 2.3.1) signalisieren H-Suffixe (i.e. H-Trailingtöne und H-Terminaltöne) generell *openness* bzw. *incompleteness*. Ein

[2] Mit GToBI(S) wurden bisher ca. 2 Stunden Sprachdaten manuell prosodisch annotiert. Das Korpus besteht hauptsächlich aus Radionachrichten (1 h 26 min.), aber auch aus anderen gelesenen und spontansprachlichen Daten. Ein ausführliche Beschreibung des Korpus findet sich in Dogil, Kuhn, Mayer & Rapp (1997), sowie im *World Wide Web* unter der Adresse http://www.ims.uni-stuttgart.de/projekte/sfb/c4/databases.html.

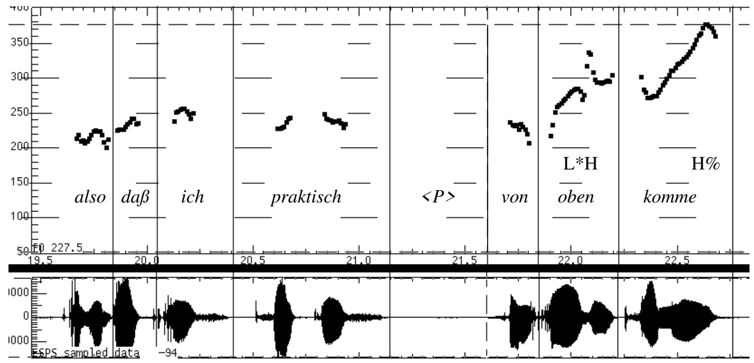
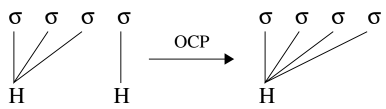


Abbildung 3-1: *Terminaler Pitchanstieg nach H-Trailington (L*H H%). (Das L-Target von L*H ist zwar perzipierbar, aber aufgrund der Laryngalisierung zwischen* von *und* oben *(s. Sprachsignal) in der* F_0*-Kontur nicht erkennbar.)*

Aspekt von *openness* ist *other-directedness* (Hobbs, 1990: 320). Dieser Aspekt steht z.B. bei Entscheidungsfragen (die in aller Regel mit einem H-Suffix enden) im Vordergrund, da in diesem Fall ein anderer Diskursteilnehmer als der Sprecher aufgefordert ist, den ‚offenen‘ Diskurs fortzusetzen. *Other-directedness* ist jedoch kein obligatorischer Bestandteil der Semantik von H-Suffixen. Dies läßt sich daran erkennen, daß auch ‚*continuation rise*‘ mit einem H-Suffix repräsentiert wird, wobei in diesem Fall die mit *openness* verknüpfte Initiative zur Fortsetzung des Diskurses auf den Sprecher selbst gerichtet ist, also nicht *other-directed*. Eine mögliche Hypothese für die Interpretation von *upstep*, die sich auf diese Analyse stützt, lautet, daß durch *upstep* die *other-directed*-Qualität explizit spezifiziert wird, während sie ohne *upstep* unterspezifiziert bleibt. Basierend auf dieser Hypothese kann die folgende phonologische Analyse entwickelt werden: Wenn man davon ausgeht, daß der tonale Gehalt terminaler Grenztöne durch *spreading* des Trailingtons des finalen Pitchakzents repräsentiert wird, stellt die Sequenz aus einem H-Trailington gefolgt von H% ein Problem dar. Die H-Spezifikation des

Grenztons H% ist in der autosegmentalen Repräsentation der H-Spezifikation des Trailingtons direkt benachbart und kann gemäß dem *Obligatory Contour Principle* (OCP; Yip, 1988) getilgt werden:

(2)



Wenn jedoch ein finaler H-Ton eine andere Funktion hat als ein ‚normaler', d.h. nicht finaler H-Ton, z.B. als Merkmal für *upstep*, ist die Repräsentation weniger problematisch. Es handelt sich dann nach wie vor um einen H-Ton im *tonal tier* mit der für H-Töne charakteristischen phonetischen Interpretation, jedoch erweitert um eine spezifische phonologische Funktion, die ebenfalls phonetisch interpretiert werden kann: durch die Realisierung von „*very high values*". Entsprechend gilt für die semantische Repräsentation, daß es sich bei H% um einen H-Ton mit der generellen H-Tonsemantik handelt, wobei in einem bestimmten Kontext – nach einem H-Trailington – ein Aspekt dieser generellen Semantik explizit spezifiziert wird.

3.2.3.2 Pitchakzente

Das Akzentinventar von GTOBI(S) entspricht den von Féry (1993) identifizierten Kategorien. Es besteht aus den fünf Akzenten H*L, L*H, HH*L, L*HL und H*M.[3]

[3] Die Notation für Akzente in GTOBI(S) verzichtet auf das „+"-Zeichen. Dem Ton, der mit der akzentuierten Silbe assoziiert ist (‚Sternton'), wird wie in der Standardnotation ein „*" nachgestellt, alle anderen Töne sind abhängig von ihrer Position (vor oder nach dem Sternton) entweder *leading tones* (Präfixe) oder *trailing tones* (Suffixe). H*L ist also äquivalent zu H*+L usw.

H*L ‚Gipfelakzent'; ein H-Target (lokales F_0-Maximum) in der akzentuierten Silbe, gefolgt von einem *steilen* Pitchabfall in das untere Drittel des Stimmumfangs (L-Traget). Wenn die akzentuierte Silbe die letzte Silbe einer Phrase ist, liegt das L-Target in der akzentuierten Silbe, ansonsten wird es – abhängig von Sprechgeschwindigkeit, Silbenstruktur und segmentaler Struktur (Stimmhaftigkeit) – in der Regel erst nach der akzentuierten Silbe erreicht.

L*H Ein L-Target (lokales F_0-Minimum) in der akzentuierten Silbe, gefolgt von einem *steilen* Pitchanstieg in das obere Drittel des Stimmumfangs (H-Target). Wenn die akzentuierte Silbe die letzte Silbe einer Phrase ist, liegt das H-Target in der akzentuierten Silbe, ansonsten wird es in der Regel erst nach der akzentuierten Silbe erreicht.

HH*L ‚*early peak*'; ein H-Target in der präakzentuierten Silbe, gefolgt von einem steilen oder stufenförmigen Pitchabfall. Dieser Akzent muß mindestens in zwei Silben – präakzentuierte und akzentuierte – realisiert werden; notwendige Bedingung ist ein Pitchabfall zwischen der präakzentuierten und der akzentuierten Silbe. Der Pitchabfall kann entweder stufenförmig, d.h. mit einem downstepped H-Target in der akzentuierten Silbe realisiert werden oder steil, d.h. wie bei einem nach vorne verschobenen H*L Akzent. Die erste Variante wird in der Regel dann realisiert, wenn die akzentuierte Silbe nicht phrasenfinal ist, die zweite Variante dagegen in der Regel dann, wenn die akzentuierte Silbe phrasenfinal ist. HH*L kann nur dann verwendet werden, wenn die präakzentuierte Silbe metrisch schwach ist, d.h. nicht akzentuierbar.

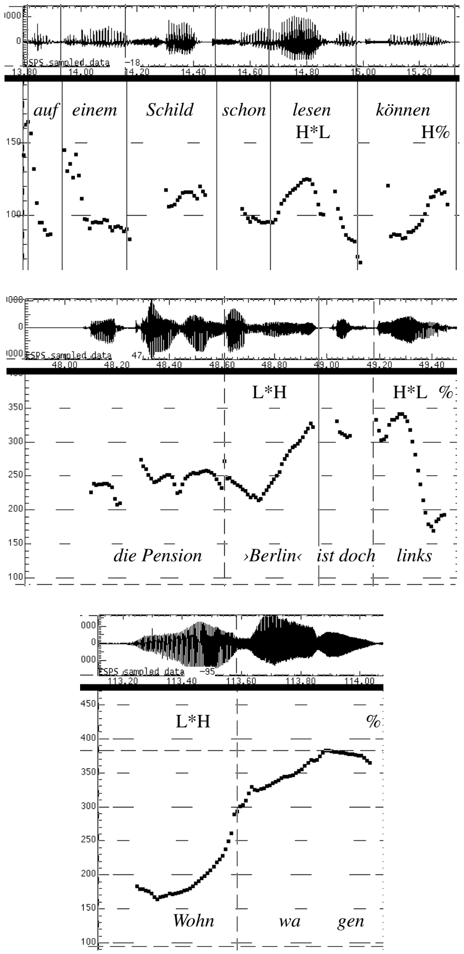


Abbildung 3-2: *Illustration der Akzente H*L und L*H in verschiedenen Kontexten. Vertikale Linien kennzeichnen Wort- bzw. Silbengrenzen. (Die Konturen sind nicht vollständig annotiert).*

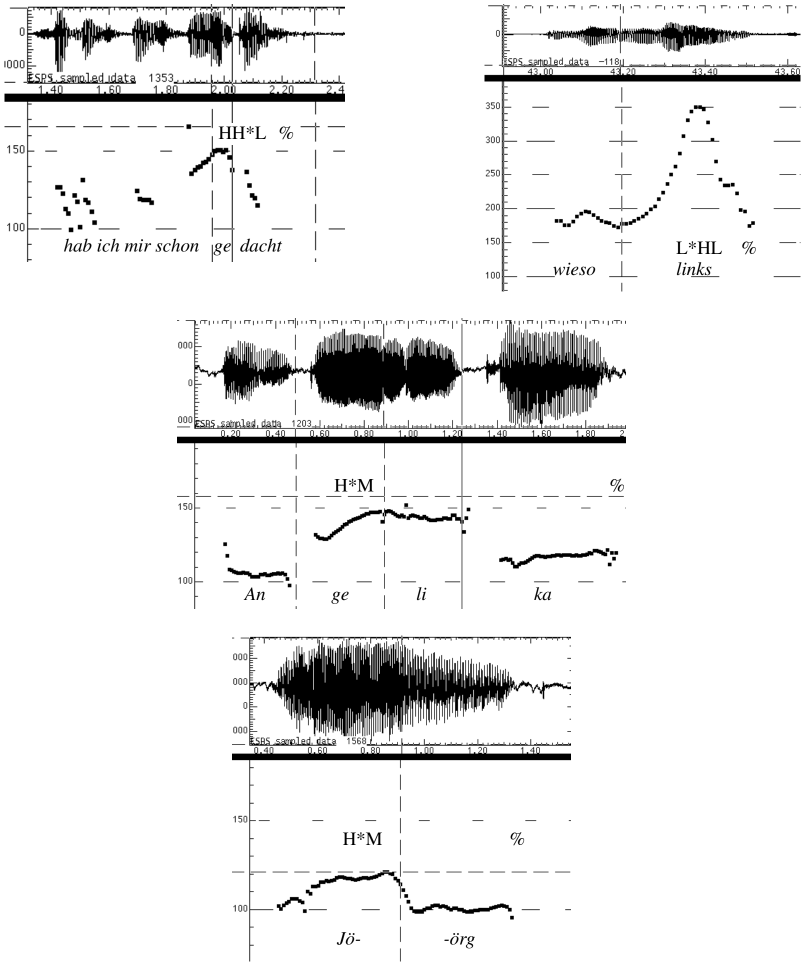


Abbildung 3-3: *Illustration der Akzente HH*L, L*HL und H*M. Vertikale Linien kennzeichnen Wort- bzw. Silbengrenzen.*

L*HL ‚*rise-fall*‘; ein L-Target früh in der akzentuierten Silbe, gefolgt von einem Pitchgipfel (H-Target) und einem steilen Pitchabfall (L-Target). Wenn die akzentuierte Silbe nicht phrasenfinal ist (was selten der Fall ist, da nach einem L*HL-akzentuierten Wort meist zumindest eine intermediäre Phrasengrenze realisiert wird), können die drei tonalen Targets auf drei aufeinanderfolgende Silben verteilt sein. Meist wird L*HL in zwei Silben oder einer (langen) Silbe realisiert, wobei die Position des H-Targets stark von der segmentalen Struktur der Silben abhängt.

H*M ‚*stylization*‘; H*M kann nur an nuklearer Position stehen, die Pitchkontur endet in der Mitte des Stimmumfangs. Nach dem H-Target in der akzentuierten Silbe folgt ein Levelton, entweder auf dem H-Niveau – dann wird das M-Target in der finalen Silbe realisiert – oder auf dem M-Niveau. Wenn die H*M-akzentuierte Silbe phrasenfinal ist, wird der Nukleus der Silbe in der Regel stark gelängt bzw. dupliziert, so daß beide Targets deutlich realisiert werden können.

Für die Repräsentation von Linking-Prozessen stehen die Labels H*, L*, ..L und ..H zur Verfügung.

H*/..L partielles Linking; ein H-Target in der akzentuierten Silbe, gefolgt von einem *flachen* Pitchabfall. Der Pitchabfall endet idealerweise in der Silbe vor der nächsten akzentuierten Silbe. H* wird in der akzentuierten Silbe gelabelt, ..L in der Silbe unmittelbar vor der nächsten akzentuierten Silbe.

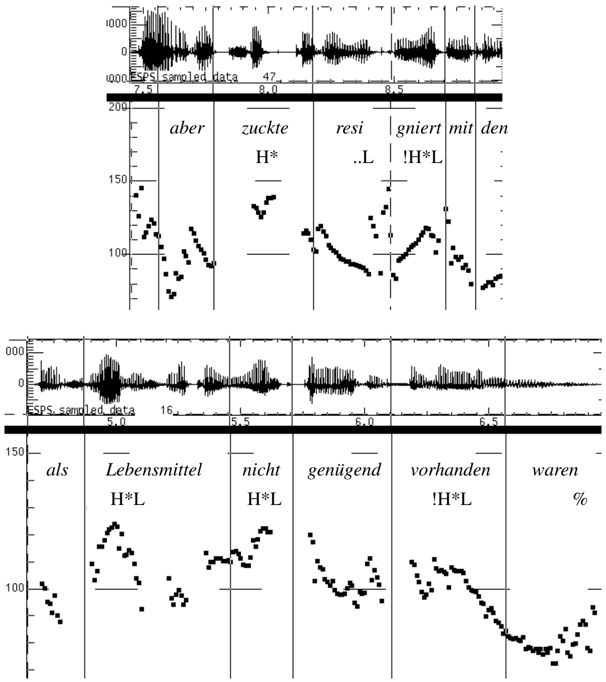


Abbildung 3-4: *Linking und* downstep. *Vertikale Linien kennzeichnen Wort- bzw. Silbengrenzen.*

L*/..H partielles Linking; ein L-Target in der akzentuierten Silbe, gefolgt von einem *flachen* Pitchanstieg. Der Pitchanstieg endet idealerweise in der Silbe vor der nächsten akzentuierten Silbe. L* wird in der akzentuierten Silbe gelabelt, ..H in der Silbe unmittelbar vor der nächsten akzentuierten Silbe.

H* komplettes Linking; ein H-Target in der akzentuierten Silbe. Der Pitchverlauf zwischen H* und dem nächsten Akzent gleicht einer Interpolation zwischen den benachbarten Tönen.

L* komplettes Linking; ein L-Target in der akzentuierten Silbe. Der Pitchverlauf zwischen L* und dem nächsten Akzent gleicht ebenfalls einer Interpolation zwischen den benachbarten Tönen.

Wie in ToBI dient das Diakritikum „!“, das allen nicht-phraseninitialen H-Tönen vorangestellt werden kann, der Repräsentation von *downstep*.

3.2.3.3 Schlußbemerkung

Vielleicht mehr noch als das originale ToBI-System ist GToBI(S) ein *phonologisches* Transkriptionssystem. Im Mittelpunkt steht nicht eine möglichst adäquate Beschreibung von Grundfrequenzkonturen. Für diesen Zweck sind F_0-Algorithmen und diverse Glättungsverfahren zweifellos die besseren Werkzeuge. GToBI(S) versucht stattdessen, nur solche intonatorischen Ereignisse zu annotieren, die in einem phonologischen Sinne kategorial, d.h. interpretierbar sind. Die phonetische Transparenz der Labels, die oben kurz angesprochen wurde, ist in diesem Zusammenhang ein sicherlich wünschenswerter, aber eben nur sekundärer Aspekt. Damit wird auch deutlich, wie dieses Kapitel und die übrigen Teile dieser Arbeit zusammenhängen: Einerseits ist GToBI(S) tatsächlich ‚nur‘ ein Werkzeug zur Annotation gesprochener Sprache bzw. ein Hilfsmittel zur Durchführung von Analysen z.B. im Bereich der Prosodie-Semantik-Schnittstelle wie im vorliegenden Fall. Andererseits kann sich ein phonologisches Transkriptionssystem wie GToBI(S) nur im Zusammenhang mit Untersuchungen zur Phonologie und Bedeutung der tonalen Kategorien, wie sie im Hauptteil dieser Arbeit, im Teil III, vorgestellt werden, entwickeln.

4 Regelbasierte Synthese

Sowohl für die Überprüfung phonologischer Hypothesen im Bereich der Prosodie und der Intonation, als auch für die Durchführung kontrollierter Perzeptionsexperimente (vgl. Kapitel 6) stellt die regelbasierte Synthese der prosodischen Parameter des Sprachsignals ein wichtiges Hilfsmittel dar: 1.) Durch die Implementierung eines phonologischen Modells in Form eines Synthesesystems können phonologische Kategorien basierend auf dem *Analysis-by-Synthesis*-Paradigma experimentell evaluiert werden. 2.) Durch die gezielte Variation und Synthese der drei wichtigsten prosodischen Parameter – Intensität, Dauer und Grundfrequenz – können Sprachsignale mit exakt kontrollierbaren prosodischen Eigenschaften hergestellt werden.

In Stuttgart wurde parallel zu der Entwicklung von GToBI(S) ein solches Synthesesystem entwickelt (Möhler i.V.). Dieses Synthesesystem resynthetisiert ein gegebenes Sprachsignal auf der Basis der veränderten prosodischen Parameter, d.h. die segmentalen Signalcharakteristika des (natürlichen) Originalsignals wie z.B. die Formantenstruktur bleiben erhalten, während die prosodischen Parameter Intensität, Dauer und F_0 neu generiert werden. Das Syntheseverfahren des Möhler'schen Systems basiert auf der PSOLA-Technik (Moulines & Charpentier, 1990). Der wichtigste Vorteil der PSOLA-Resynthese gegenüber einer Vollsynthese ist die Natürlichkeit des resultierenden Sprachsignals. Da bei der Resynthesetechnik die segmentalen Eigenschaften nicht künstlich erzeugt, sondern von einem natürlichen Signal übernommen werden, klingt das Syntheseergebnis kaum anders als das Ausgangssignal – abgesehen von den beabsichtigten Veränderungen der Prosodie[4]. Die gesamte Funktionalität des Stuttgarter Synthese-Werkzeugs wurde in die ESPS/*waves*-Umgebung integriert.

[4] Dies gilt bei einer moderaten Veränderung der prosodischen Parameter. Bei größeren Eingriffen, z.B. bei starker Längung oder extremer Expansion des Grundfrequenzumfangs, verringert sich von Fall zu Fall die Natürlichkeit des Syntheseergebnisses.

4.1 Systematische Variation prosodischer Parameter

Das Möhler'sche Synthesesystem erlaubt dem Benutzer die gezielte Modifikation der Intensität, der Dauer und/oder der Grundfrequenz in frei definierbaren Bereichen des Originalsignals. Hierfür stehen dem Benutzer drei unabhängige Panels zur Verfügung (s. Abbildung 4-1):

- *intensity panel*: eine Verschiebung der Nullinie nach oben resultiert in größerer Intensität, eine Verschiebung nach unten in geringerer Intensität.
- *duration panel*: eine Verschiebung der Nullinie nach oben längt das Segment, eine Verschiebung nach unten hat eine kürzere Dauer zur Folge.
- F_0 *panel*: die Originalkontur kann beliebig verändert werden, das Signal wird mit der eingegebenen Kontur synthetisiert.

In den einzelnen Panels kann entweder mit der Maus editiert werden oder die gewünschten Werte werden direkt in die ESPS-Dateien eingetragen. Auf Tastendruck wird das Signal dann mit den editierten Parametern resynthetisiert, im ESPS-Dateiformat gespeichert und wiedergegeben. Auf diese Weise kann z.B. überprüft werden, ob die Variation der Dauer und die Variation der Grundfrequenz tatsächlich zu unterschiedlichen und unabhängigen Prominenzeffekten führen (Wortbetonung vs. Satzakzent) und welche Rolle die Variation der Intensität spielt (vgl. Möhler & Dogil, 1995).

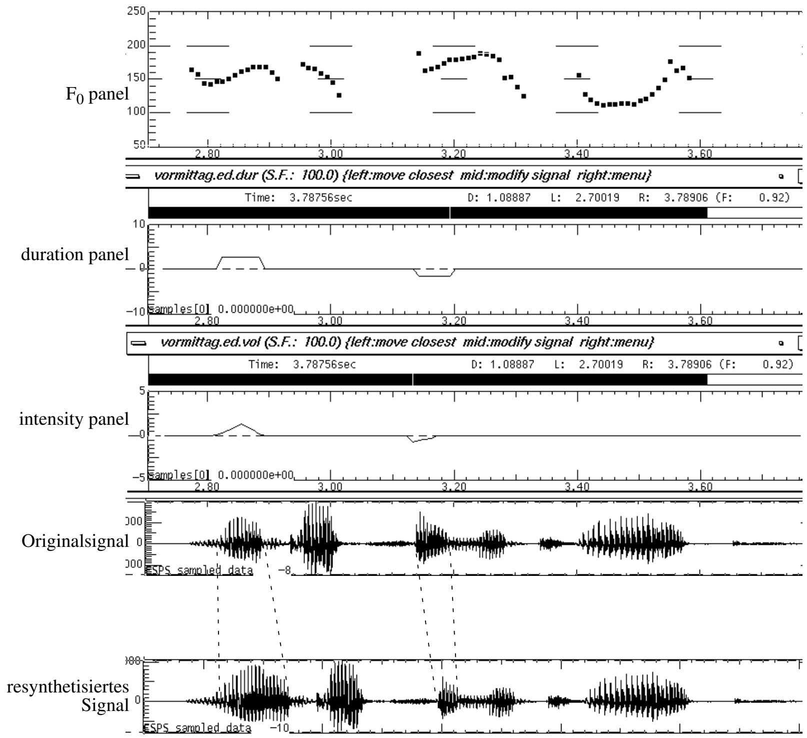


Abbildung 4-1: *Originalsignal mit editierten prosodischen Parametern und Resyntheseprodukt. Äußerung: „*Montag Vormittag*“. Die gestrichelten Linien illustrieren die Dauerunterschiede zwischen Original- und Resynthesesignal (vgl.* duration panel*).*

4.2 Regelbasierte Generierung von Grundfrequenzkonturen

Einen Schritt weiter geht die regelbasierte Generierung von prosodischen Parametern. Die Parameter werden in diesem Fall nicht mehr von Hand editiert, sondern mithilfe von Generierungsregeln – vergleichbar den phonetischen Realisierungsregeln eines phonologischen Modells – aus der prosodischen Annotation abgeleitet und automatisch erzeugt. In Möhlers Synthesesystem ist diese Option für die Variation der Grundfrequenz bereits verwirklicht (Möhler, 1996). Diese regelbasierte Konturgenerierung basiert auf dem Symbolinventar von GTOBI(S). In dem Regelwerk ist für jede GTOBI(S)-Kategorie ein Regelsatz formuliert, der, abhängig vom tonalen Kontext, charakteristische Konturverläufe generiert. Die Regeln reflektieren den Level- bzw. Targetansatz des Ton-Sequenz-Modells, d.h. für jeden Ton wird ein F_0-Target bestimmt und die Targets werden durch Interpolation verbunden. Die Targets werden in y-Richtung (Tonhöhe) relativ zu einem frei konfigurierbaren *range* definiert, welcher durch eine *topline* (maximale F_0-Werte) und eine *baseline* (minimale F_0-Werte) repräsentiert ist. Ein H-Target erhält also entweder einen F_0-Wert auf der *topline* (*default*) oder einen F_0-Wert, der um einen bestimmten Prozentsatz über der *topline* liegt (z.B. H% nach H-Trailington; s. Abschnitt 3.2.3) oder einen F0-Wert, der um einen bestimmten Prozentsatz unter der *topline* liegt (z.B. !H). Analog dazu werden L-Targets relativ zur *baseline* definiert. Die temporale Komponente der Targets (x-Richtung) wird relativ zu verschiedenen segmentalen Ankerpunkten definiert, z.B. relativ zum Silbenanfang/Silbenende, zum Vokalanfang/Vokalende oder zum Beginn/Ende der Stimmhaftigkeit innerhalb einer Silbe[5]. Wie bereits erwähnt, berücksichtigt das Regelwerk auch den tonalen Kontext und die Phrasierung, so daß auch unmittelbar aufeinanderfolgende Akzente oder Akzente in der letzten Silbe einer Phrase korrekt modelliert werden. Dies betrifft insbesondere die temporale Alinierung der tona-

---

[5] Die segmentalen Informationen erhält das System aus der (automatisch erzeugten; vgl. Rapp, i.V.) Phonemtranskription. Diese Phonemtranskription (und die Annotation der Silbenstruktur) kann als zusätzliche GTOBI(S)-Spur den Sprachdaten beigefügt werden.

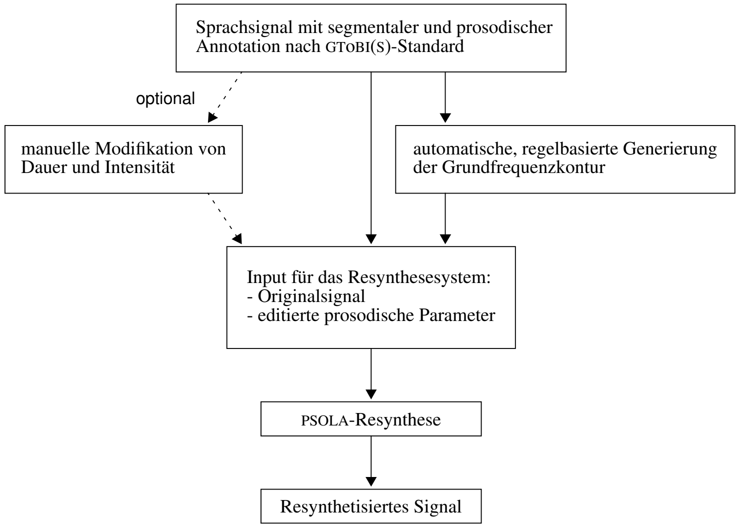


Abbildung 4-2: *Die Integration des Konturgenerierungsmoduls in das PSOLA-Resynthesesystem.*

len Targets mit der segmentalen Struktur; z.B. wird bei einem H*L Akzent, der mit einer phrasenfinalen Silbe assoziiert ist, das H-Target früher in der akzentuierten Silbe realisiert und das L-Target fällt ebenfalls, mangels einer weiteren Silbe, in die akzentuierte Silbe. Das gesamte Regelwerk ist in C++ implementiert, kann jedoch auch ohne C++-Kenntnisse leicht modifiziert und im Rahmen der *Analysis-by-Sythesis* angepaßt werden. Abbildung 4-2 zeigt die Architektur des Konturgenerierungs- und Resynthesesystems.

In der experimentellen Arbeit liegt der entscheidende Vorteil der automatischen Konturgenerierung gegenüber der manuellen Modifikation in der größeren Konsistenz der Syntheseergebnisse. Wenn die Feineinstellung des Regelwerks abgeschlossen und der Zielsprache (in unserem Fall Deutsch) angepaßt ist, werden alle phonologischen Kategorien – Pitchakzente und Grenztöne – im gleichen Kontext stets identisch realisiert, d.h. in

F_0-Werte umgesetzt. Eine mögliche Anwendung dieses Verfahrens, nämlich die Konstruktion experimenteller Items für ein Perzeptionsexperiment, wird in Kapitel 6 demonstriert.

Ich möchte dieses Kapitel abschließen mit einer Illustration des Konturgenerierungsverfahrens. Abbildung 4-3 zeigt oben die Originalkontur mit einem H*L Akzent in *UNgeduldig*. Die mittlere Kontur zeigt das Ergebnis der automatischen Konturgenerierung, sie bildet den Input für das Resynthesesystem. Dieses Resultat basiert allein auf der unten angegebenen tonalen Annotation H*L %.[6] Kleine Abweichungen vom linearen Konturverlauf weisen darauf hin, daß auch mikroprosodische Aspekte bei der Konturgenerierung berücksichtigt werden (Möhler, i.V.). Die untere Kontur zeigt schließlich das Ergebnis der F_0-Analyse des resynthetisierten Signals. Die geringen, auditiv nicht wahrnehmbaren Abweichungen von der generierten Kontur sind zurückzuführen auf den Resynthesealgorithmus. Mit diesem Beispiel soll demonstriert werden, daß die regelbasierte Synthese sehr gut in der Lage ist, natürliche Konturen zu modellieren. Das Beispiel in Abbildung 4-4 zeigt dagegen, daß mithilfe dieses Werkzeugs auch Konturen erzeugt werden können, die von der Originalkontur mehr oder weniger stark abweichen. Auf diese Weise können beliebige Äußerungen mit allen phonologisch möglichen Intonations- und Prominenzmustern resynthetisiert werden – in der Regel ohne entscheidende Einbußen hinsichtlich der Signalqualität.

[6] Das Label L% in der Abbildung ist äquivalent zu %. Das Label L% wird in der aktuellen GToBI(S)-Version nicht mehr benutzt (vgl. Kapitel 3).

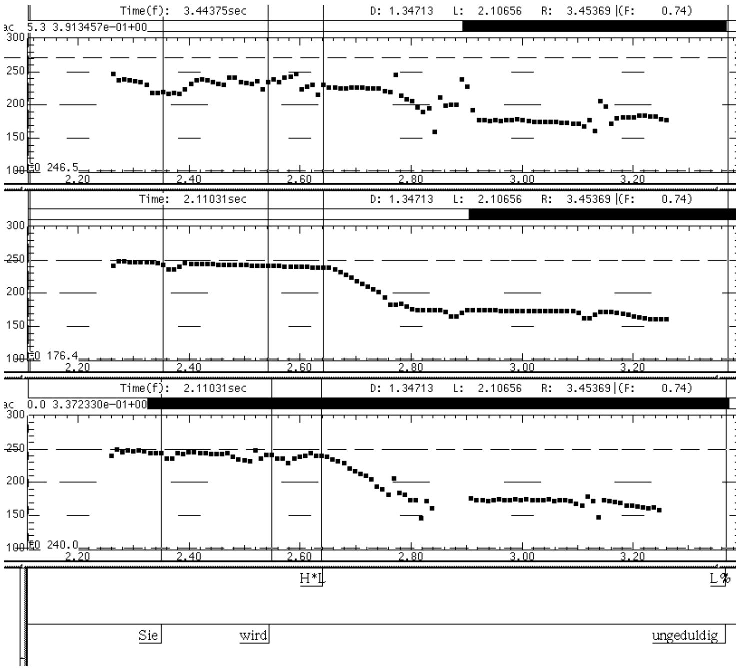


Abbildung 4-3: *Labelspuren (unten) und F_0-Konturen zu dem Satz „*Sie wird ungeduldig*“. Die obere Kontur ist das Ergebnis der F_0-Analyse des Originalsignals, die mittlere Kontur ist automatisch generiert auf der Basis der tonalen Labels, die untere Kontur ist das Ergebnis der F_0-Analyse des resynthetisierten Signals.*

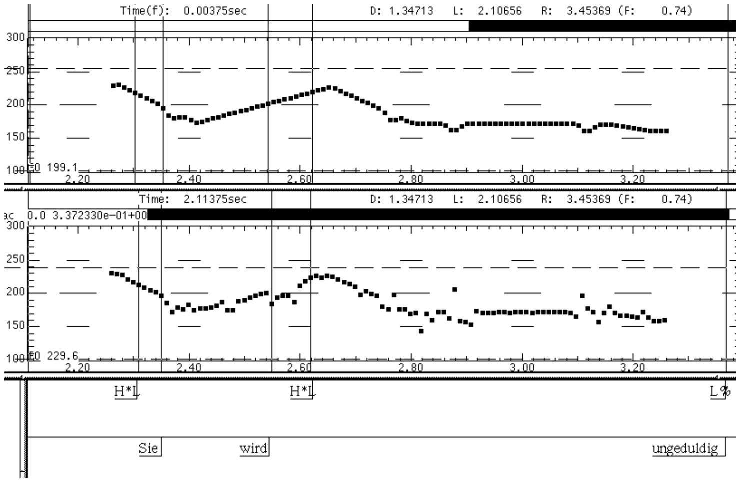


Abbildung 4-4: *Generierte Kontur (oben) und Ergebnis der F_0-Analyse des resynthetisierten Signals (unten) zu dem Satz „*Sie wird ungeduldig*“. Grundlage der regelbasierten Konturgenerierung ist die prosodische Annotation im Bild ganz unten (L% entspricht %).*

Teil III

Experimente

In diesem Teil werden drei unterschiedliche Aspekte des Zusammenspiels von Intonation und Bedeutung mit experimentell-phonetischen Methoden untersucht, d.h. die in Teil I vorgestellten Phänomene werden anhand von drei Beispielen mit den in Teil II eingeführten Werkzeugen und anderen empirischen Verfahren exemplarisch untersucht und diskutiert. Die erste Studie beschäftigt sich mit lokalen intonatorischen Phänomenen und ihrem Beitrag zur Disambiguierung eines Adverbs auf der Satzebene. Behandelt werden die drei Lesarten des Adverbs *erst* und ihre distinktiven prosodischen Korrelate. Die zweite Studie geht über die Satzebene hinaus, sie hat die Interpretation von anaphorischen Einheiten in Zwei-Satz-Diskursen zum Gegenstand und problematisiert den ›lokalen Charakter‹ lokaler intonatorischer Phänomene. Die dritte Studie schließlich behandelt die diskurssemantische Strukturierung größerer Diskurseinheiten und deren globalen intonatorischen Korrelate. Neben der Pausendauer steht dabei im Mittelpunkt die Analyse der Tonregister.

5 Lokale Prominenz: Disambiguierung des Adverbs *erst*

Durch die Untersuchung des ambigen deutschen Adverbs *erst* im Rahmen der Diskursrepräsentationstheorie (DRT) (Kamp & Reyle, 1993) zeigt Eberle (1996), daß zur Auflösung ambiger Strukturen oder ambiger lexikalischer Elemente in bestimmten Fällen auf eine tiefe semantische Analyse der Kontextinterferenzen verzichtet werden kann, zugunsten einer weniger aufwendigen Analyse der Informationsstruktur. Computerlinguistische Anwendungen wie automatisches Textverstehen oder qualitativ hochwertige maschinelle Übersetzung sind angewiesen auf solche schnellen und zuverlässigen Disambiguierungsverfahren. Die in diesem Kapitel vorgestellte Studie soll erste Hinweise liefern, ob in einem solchen Szenario die Berücksichtigung intonatorischer Phänomene gesprochener Sprache die flache semantische Analyse ergänzen und unterstützen kann.

5.1 Drei Lesarten von *erst*

Eberle (1996) unterscheidet die folgenden drei Lesarten von *erst*:

- *First of a Sequence* (FS)
  „the recipient understands the introduced event as the first of a sequence of events that he expects to be completed by the following text.“

- *Exclusion of Preceding Alternatives* (EPA)
  „the recipient understands *erst* as a signal of the speaker/writer that the occurrence of the reported event is not preceded by the occurrence of similar (alternative) events.“

- *Retardation* (R)
  „the recipient understands the event as element of a sequence of events, and the realization of the sequence, in particular the reported realization of the event at the textual perspective time, seems to be in retardation, with regard to some (previous) expectation about the realization dates of the sequence.“
  (alle Zitate aus Eberle (1996), S.1)

Das folgende Beispiel, ebenfalls entnommen aus Eberle (1996), bietet drei Kontexte für den *erst*-Satz in (1), die jeweils eine der oben genannten Lesarten von *erst* präsupponieren.

(1) Peter zeigte erst auf die vierte Glückszahl.

(1a) Und dann auf die zweite. (FS)

(1b) Nicht zuvor auf die erste, zweite oder dritte. (EPA)

(1c) Noch nicht auf die fünfte. (R)

Die Funktion von *erst* im FS-Kontext ist die eines Temporaladverbs, im EPA- und R-Kontext die eines Fokusadverbs.

5.2 Die Lesarten und ihre intonatorische Realisierung

Die folgende Untersuchung soll aufzeigen, ob und wenn ja, wie die Akzentstruktur eines Satzes zur Disambiguierung lexikalischer Einheiten beitragen kann. Die Studie hat rein deskriptiven, exemplarischen Charakter und liefert insofern keine tiefgehende Analyse der Schnittstelle zwischen Intonation und Semantik. Stattdessen beleuchtet sie anhand eines einfachen Beispiels die Thematik, die dann im experimentellen Hauptteil der Arbeit in Kapitel 6 eingehender behandelt wird. Außerdem bildet diese Studie den Ausgangspunkt einer Reihe weiterer Arbeiten, die sich speziell mit der Prosodie fokussensitiver Partikeln im Deutschen beschäftigen. Diese Arbeiten sind jedoch noch im Anfangsstadium, so daß sie hier noch nicht dokumentiert werden können. Im einzelnen handelt es sich dabei um eine korpusbasierte Untersuchung von *auch* in Zusammenarbeit mit Prof. Marga Reis, Universität Tübingen, und um eine experimentelle Untersuchung verschiedener Gradpartikeln des Deutschen im Rahmen einer Diplomarbeit von Karin Müller, Universität Stuttgart. Gegenstand der vorliegenden Pilotstudie ist das Adverb *erst* und die intonatorische Realisierung seiner drei Lesarten. Die Studie ist als Produktionsexperiment angelegt.

5.2.1 Material, Sprecher und Durchführung

Als Material dienten die Kontexte aus (1). Die Satzpaare in (2) - (4) wurden von zwei Sprecherinnen und einem Sprecher je dreimal vorgelesen, so daß von jedem Satzpaar drei Realisierungen von drei verschiedenen Sprechern vorlagen.

(2) *Peter zeigte erst auf die vierte Glückszahl. Und dann auf die zweite.*

(3) *Peter zeigte erst auf die vierte Glückszahl. Nicht zuvor auf die erste, zweite oder dritte.*

(4) *Peter zeigte erst auf die vierte Glückszahl. Noch nicht auf die fünfte.*

Die Sprecher waren Teilnehmer an einem Seminar zur Intonation des Deutschen und insofern vertraut mit der Phonologie und Phonetik der Intonation sowie mit Modellen der Interaktion zwischen Intonation und Semantik. Sie wurden auch über die Fragestellung dieser Studie aufgeklärt und mit den Lesarten von *erst* vertraut gemacht. Die Möglichkeiten der intonatorischen Realisierung der drei Satzpaare wurde jedoch nicht diskutiert.

Den Versuchspersonen wurden alle drei Satzpaare in schriftlicher Form vorgelegt. Die Versuchspersonen sollten sich mit dem Material vertraut machen und dann isoliert voneinander die Satzpaare in der in (2) - (4) angegebenen Reihenfolge dreimal hintereinander produzieren.

Die insgesamt 27 Satzpaare wurden in einem normalen Raum ohne besondere qualitätsverbessernde Vorkehrungen auf DAT-Band aufgenommen und auf eine SGI Indy Workstation überspielt (16 kHz, 16 Bit). Der bei allen drei Satzpaaren identische erste Satz wurde dann isoliert und mit ESPS/*waves* und dem in Kapitel 3 vorgestellten Labellingsystem annotiert und analysiert.

5.2.2 Ergebnisse und Diskussion

Bei allen drei Sprechern kamen in dem Satz *Peter zeigte erst auf die vierte Glückszahl* in den oben angegebenen Kontexten als Akzentpositionen nur *erst* und *vierte* in Frage. Alle anderen Wörter waren unbetont. Tabelle 5-1 gibt einen Überblick über die realisierten Akzentmuster.

Zunächst ist zu beobachten, daß *erst* in der Funktion des Temporaladverbs (FS-Lesart) tendenziell nicht akzentuiert wird, wohingegen *erst* in der Funktion des Fokusadverbs (EPA- und R-Lesart) in der Regel einen Akzent trägt und zwar vom Typ L*H. Ein weiterer Unterschied zwischen der FS-Lesart auf der einen Seite und der EPA- und R-Lesart auf der anderen Seite ist der Akzenttyp, der auf dem fokussierten Element (*vierte*) realisiert

Tabelle 5-1: Akzentmuster von 9 Realisierungen pro Kontext; in Klammer die absolute Häufigkeit des jeweiligen Akzentmusters.

	Peter zeigte	*erst*	*auf die*	*vierte*	*Glückszahl*
FS		- (7)		L*H (9)	%
		H* (2)			
EPA		L*H (9)		H*L (8)	%
				L*H (1)	
R		L*H (6)		H*L (9)	%
		- (3)			

wird. Es folgt zunächst die Diskussion des mit der FS-Lesart assoziierten Akzentmusters, dann der Akzentmuster der EPA- und R-Kontexte und schließlich eine kurze Zusammenfassung der Forschungsperspektive, die sich aus diesem Experiment ergibt.

Der *erst*-Satz erscheint in der semantischen Repräsentation, der Diskursrepräsentationsstruktur (DRS), jeweils als Ereignis (*event*) *e* im Sinne der DRT (Kamp & Reyle, 1993). Im Falle der FS-Lesart präsupponiert *erst* ein Referenzereignis *e'*, so daß *e* als das erste Ereignis einer Elaborations-Sequenz *e'* verstanden wird (Eberle, 1996). Der *erst*-Satz behauptet *e* und präsupponiert *e'*, wobei *e'* aus *e* und mindestens einem weiteren Ereignis e_2 zusammengesetzt ist. In unserem Beispiel gilt:

e' = {*e*: zeigen_auf(peter,vierte_glückszahl),
e_2: zeigen_auf(peter,zweite_glückszahl)}

Wenn also ein Hörer den *erst*-Satz in der FS-Lesart interpretiert, erwartet er die Fortsetzung und Vervollständigung der mit der Behauptung von *e* begonnenen Ereignissequenz. Der L*H-Akzent mit progredienter Terminalkontur (L*H %), also die Konstellation, die von allen Sprechern bei allen Realisierungen dieser Lesart produziert wurde, korreliert sehr gut mit dieser semantischen Analyse. Wie in Abschnitt 2.3 gezeigt, geht das Modell

von Hobbs (1990) davon aus, daß ein L*-Akzent signalisiert, daß die Proposition nicht neu ist, während ein H-Suffix (*trailing*-Ton) Unvollständigkeit signalisiert. Wenn nun das Satzpaar in (2), wie im Experiment geschehen, *out of the blue* realisiert wird, liegt es nahe, die gesamte Proposition, die von beiden Teilsätzen transportiert wird – also *e'* –, mit einem H*L-Akzent im zweiten Teilsatz als neu (H*-Akzent) und vollständig (L-Suffix) zu kennzeichnen. Das prominente Element im ersten Teilsatz muß dagegen aufgrund der präsupponierten Weiterführung der Proposition Unvollständigkeit bzw. Progredienz signalisieren: *e* ist, wenn *erst* in der FS-Lesart gebraucht wird, nur das erste Ereignis in einer zusammengehörigen Sequenz von Ereignissen und kann insofern nicht mit einem Akzent, der Vollständigkeit signalisiert, assoziiert werden. Ist *e'* aus mehr als zwei Ereignissen zusammengesetzt, so ist zu erwarten, daß der H*L-Akzent erst im letzten Teilsatz realisiert wird:

(5) Peter zeigte erst auf die vierte Glückszahl. Und dann auf die zweite
L*H L*H

und die dritte. Und zum Schluß auf die fünfte.
L*H H*L

Ein H*L-Akzent im ersten Teilsatz ist nur als kontrastiver Akzent möglich, nur dann ist der *erst*-Satz für sich alleine kohärent, vgl. (6) und (7).

(6) *out of the blue* Realisierung:
Peter zeigte erst auf die vierte Glückszahl. Und dann auf die zweite.
L*H H*L

(7) Kontrastakzent:
Zeigte Peter zuerst auf die zweite Glückszahl?
Nein, Peter zeigte erst auf die vierte Glückszahl. Und dann auf die zweite.
H*L H*L

In (7) ist *vierte* mit H*L-Akzent und engem Fokus tatsächlich die neue Information und der erste Teilsatz ist eine adäquate und vollständige Proposition in diesem restriktiven Kontext. Der zweite Teilsatz wird in diesem Fall nicht präsupponiert, es handelt sich um eine optionale Ergänzung, die einen zusätzlichen Aspekt des Kontrastes ausdrückt, nämlich die tatsächliche temporale Relation zwischen dem im ersten Teilsatz ausgedrückten Ereignis und dem im zweiten Teilsatz ausgedrückten Ereignis. Daher ist der Akzent auch nicht auf dem Adjektiv, sondern auf *dann* zu erwarten.

Die Akzentposition auf *vierte* bei der *out of the blue* Realisierung ist, wie auch bei den anderen Lesarten, kontextabhängig. Bei einem anderen Kontext, wie z.B. in (8), sind andere Akzentpositionen zu erwarten (in (8): *Glückszahl*), der Akzenttyp L*H ist dagegen – so meine Hypothese – fest mit der FS-Lesart von *erst* verbunden.

(8) Peter zeigte erst auf die vierte Glückszahl. Und dann auf den vierten Kandidaten.

Nun zu den Akzentmustern der EPA- und R-Lesart. Nur eine Sprecherin unterscheidet die beiden Lesarten mithilfe intonatorischer Mittel. Diese Sprecherin realisierte alle Satzpaare vom Typ (3) (EPA-Lesart) mit L*H auf *erst* und H*L auf *vierte*. Die Satzpaare vom Typ (4) (R-Lesart) realisierte sie dagegen mit nur einem Akzent, nämlich H*L auf *vierte*. Die beiden anderen Versuchspersonen realisierten beide Lesarten mit dem selben Muster, L*H auf *erst* und H*L auf *vierte*. Die Erklärung, für die ich hier argumentieren möchte, ist erstens, daß sich beide Lesarten tatsächlich intonatorisch nicht unterscheiden und zweitens, daß der H*L-Fokusakzent aufgrund der Semantik des Fokusadverbs *erst*

gefordert wird, während der L*H-Akzent auf *erst* optional ist und nur durch die spezifische Konstruktion des experimentellen Materials so häufig auftritt.

Die Gemeinsamkeit in der DR-theoretischen Analyse der EPA- und R-Lesart von *erst* liegt darin, daß mit dem *erst*-Satz eine Summe von Ereignissen e_1 - e_k präsupponiert wird, wobei die Ereignisse e_1 - e_k durch die Menge der Prädikate P_1 - P_k restringiert sind. P_1 - P_k sind Prädikate der Form des Ausdrucks im Skopus von *erst*, mit Alternativen an der Position des fokussierten Elements (Eberle, 1996; zur sog. ‚Alternativensemantik' vgl. Rooth, 1985, 1992). Diese Analyse wird in (9) veranschaulicht.

(9) *Peter zeigte erst auf die vierte Glückszahl.*
H*L

Skopus von *erst*: vierte_glückszahl(x)
e: zeigen_auf(peter,x)
P_1 - P_k: {erste_glückszahl, zweite_glückszahl … }[1]
e_1 - e_k: {zeigen_auf(peter,P_1), zeigen_auf(peter,P_2) … zeigen_auf(peter,P_k)}

Erst als Fokusadverb präsupponiert also kein Referenzereignis e', wovon e ein Teil ist, sondern im Sinne der Alternativensemantik von Rooth eine Menge von Alternativen zu e. Dies bedeutet, daß e genau dann kohärent ist, wenn die präsupponierten Alternativen im konkreten Diskursmodell verfügbar sind, die Äußerung eines *erst*-Satzes wie in (9) ist dann adäquat.

Ein weiterer wichtiger Aspekt der Bedeutung von *erst* in der EPA- und R-Lesart ist die intrinsische Ordnung der Menge der Fokusalternativen (Eberle, 1996). Das Fokusadverb *erst* präsupponiert die lineare Ordnung der Alternativen zu dem fokussierten Element in seinem Skopus und der konkrete *erst*-Satz wählt aus der sich daraus ergebenden möglichen oder erwartbaren Sequenz von Ereignissen ein bestimmtes aus, während er gleich-

[1] Die Menge der Alternativen wird durch den Kontext beschränkt, in diesem Fall also durch die Anzahl der Glückszahlen, auf die gezeigt werden kann (vgl. Rooth, 1985, 1992).

zeitig die übrigen Ereignisse als nicht realisiert (obwohl möglich) bzw. als noch nicht realisiert (obwohl erwartbar) auszeichnet. In unserem konkreten Beispiel ergibt sich die interne Ordnung der Fokusalternativen aus der Fokussierung des numerischen Adjektivs: es ist die kanonische Ordnung der Zahlen. Angenommen, Peter hat die Möglichkeit, auf sechs Glückszahlen zu zeigen, dann ist das Zeigen auf die vierte Glückszahl das vierte Ereignis einer Sequenz von sechs alternativen Ereignissen. Die Fokussierung von *vierte* in den Äußerungen der Versuchspersonen ist, wie bereits erwähnt, zurückzuführen auf den Kontext, der mit den zweiten Teilsatz geliefert wird. Durch die Variation des zweiten Teilsatzes wie in (10) könnte die Akzentposition auch auf *Glückszahl* verschoben werden.

(10) *Peter zeigte erst auf die vierte Glückszahl.*
H*L

(10a) Noch nicht auf die vierte Gewinnzahl.

(10b) Noch nicht auf die erste Gewinnzahl.

(10c) Für die Gewinnzahl hatte er sich noch nicht entschieden.

Die Ordnung der Fokusalternativen ist in diesem Fall nicht mehr so offensichtlich wie bei akzentuiertem Adjektiv, sie müßte aus anderen Quellen erschlossen werden (Weltwissen, Kontext etc.)[2].

Der Unterschied zwischen der EPA- und der R-Lesart ist in der oben angedeuteten Unterscheidung zwischen möglichen und erwartbaren Ereignissen zu suchen. Die EPA-Lesart ist so zu verstehen, daß *erst* eine semantische Relation *r* einführt, welche die Ereignisse e_1 - e_k als mögliche Ereignisse der Form *e* (zeigen_auf(Peter,*P*)) charakterisiert.

[2] Außerdem muß – wie das Beispiel zeigt – bei der Konstitution der Menge der Fokusalternativen in diesem Fall auch die Fokusprojektion berücksichtigt werden.

Das behauptete Ereignis *e* wird dann verstanden als das erste tatsächlich realisierte Ereignis aus der geordneten Sequenz der möglichen Ereignisse, vorangehende Realisierungen alternativer Ereignisse werden ausgeschlossen – *Exclusion of Preceding Alternatives* (Eberle, 1996). In der R-Lesart sind die Ereignisse e_1 - e_k dagegen nicht als mögliche Instantiationen von Ereignissen der Form *e* charakterisiert, sondern als „events … of an expectation about the ongoing of the world" (Eberle, 1996: 3). Die R-Lesart präsupponiert also eine bestimmte Einstellung (*attitudinal state*) gegenüber e_1 - e_k und der *erst*-Satz teilt eine Verzögerung – *Retardation* – der erwarteten Ereignissequenz mit. Außerdem wird die Realisierung von Ereignissen, die dem behaupteten Ereignis *e* vorangehen, mit der R-Lesart nicht ausgeschlossen.

Meine Hypothese ist nun, daß solche feinen semantischen Unterschiede, die erst sehr tief in der DR-Struktur zum Tragen kommen, nicht mit den Mitteln der Intonation signalisiert werden können. Dagegen verlangt die grobe semantische Struktur, die bei beiden Lesarten identisch ist, einen Fokusakzent im Skopus von *erst*[3]. Dieser Akzent hat die Funktion, das fokussierte Element auszuzeichnen, um so die korrekte Konstitution der Menge der alternativen Prädikate zu gewährleisten. Dies ist die Voraussetzung für die Identifikation und Interpretation sowohl der EPA- als auch der R-Lesart.

Zu erwarten ist ein Akzent des Typs H*L, der im vorliegenden Experiment, mit einer Ausnahme, in allen Produktionen der Satzpaare (3) und (4) realisiert wurde (vgl. Tabelle 5-1). Weder EPA- noch R-Lesart präsupponieren ein Referenzereignis, d.h. in beiden Fällen ist der *erst*-Satz auch ohne elaborierende Weiterführung kohärent. Dadurch ist ein Akzent mit L-Suffix lizenziert. Der H*-Anteil des Akzents korreliert mit der Auszeichnung des fokussierten Elements im Skopus von *erst* als die behauptete Instantiation aus einer Menge von kontextuell verfügbaren Alternativen (Büring, 1995: 46ff).

[3] Dies gilt selbstverständlich nicht bei kontrastiver Verwendung eines *erst*-Satzes mit einer der beiden Lesarten (vgl. die Diskussion zur FS-Lesart). In diesem Fall ist der Fokus jedoch aus dem Kontext erschließbar und wird deshalb nicht mehr mit einem Pitchakzent markiert (vgl. Rooth (1997) zur Analyse von *second occurrence focus*).

Der L*H-Akzent auf *erst*, der bei allen 9 Produktionen der EPA-Lesart und bei 6 Produktionen der R-Lesart realisiert wurde, ist nicht – so meine Hypothese – auf die spezifische Semantik von *erst* zurückzuführen, sondern signalisiert die (optionale) Elaboration der durch den *erst*-Satz ausgedrückten Proposition. Wie ich oben versucht habe zu zeigen, ist diese Elaboration nicht obligatorisch, aufgrund der Vorgabe von Satzpaaren jedoch im experimentellen Material faktisch vorhanden. Da den Versuchspersonen kein unbekanntes Material satzweise präsentiert wurde, sondern sie im Gegenteil gebeten wurden, sich vor der Aufnahme mit dem gesamten Material vertraut zu machen, ist es nicht verwunderlich, daß die elaborierende Fortsetzung schon im ersten Teilsatz intonatorisch angekündigt wird. Wenn diese Vermutung zutrifft, ergäbe sich für den L*(H) Akzent auf *erst* in etwa die folgende Interpretation: Aus einer präsupponierten Menge von Ereignissen wird ein Ereignis ausgezeichnet (dies ist die Funktion des Fokusakzents H*L). Über die Relation dieses Ereignisses zu den übrigen Ereignissen wird etwas ausgesagt – mit dem *erst*-Satz. Die Verwendung des Adverbs *erst* (und nicht z.B. des Adverbs *schon*) charakterisiert diese Relation – daher die (optionale) L*-Akzentposition auf *erst*. Die notwendige Information ist damit zwar mitgeteilt, aber es folgt noch eine zusätzliche Elaboration (im vorliegenden Beispiel durch eine Aussage über den Status der übrigen Ereignisse) – daher das Unvollständigkeit (*incompleteness*) signalisierende H-Suffix.

Um es nocheinmal zu wiederholen: Diese intonatorische Markierung ist optional. Sie trägt nichts zur Disambiguierung und semantischen Interpretation des *erst*-Satzes bei. Sie verstärkt aber die Kohärenz der gesamten Diskursäußerung. Einen wichtigen Beitrag zur Interpretation des *erst*-Satzes liefert dagegen, wie gezeigt, der H*L-Akzent. Und auch zur Disambiguierung der Lesarten – um auf die ursprüngliche Fragestellung dieser Studie zurückzukommen – kann nur auf diesen Akzent zurückgegriffen werden. Er unterscheidet zwar nicht zwischen der EPA- und R-Lesart, zeichnet diese beiden jedoch gegenüber der FS-Lesart aus, die bevorzugt mit einem L*H-Akzent im Skopus von *erst* realisiert wird.

Die Studie hat gezeigt, daß die Berücksichtigung der Akzentstruktur, also lokaler intonatorischer Merkmale von Äußerungen, durchaus zur Disambiguierung von semantischen Repräsentationen beitragen kann und auch, daß umgekehrt aus der DR-Struktur eines Satzes oder eines Diskurses Akzentpositionen und Akzenttypen vorhergesagt werden können. Für zukünftige vergleichbare Arbeiten ergeben sich aus der vorliegenden Studie unter anderem die folgenden Punkte zur Berücksichtigung bzw. zur Überprüfung und genaueren Untersuchung:

- Die Behandlung von optionalen Akzenten, die sowohl vom Standpunkt der Erkennung (Auswahl der relevanten Informationen) als auch vom Standpunkt der Synthese (Natürlichkeit, Redundanz) eine Herausforderung darstellen.
- Die Frage, welche Strukturen der semantischen Repräsentation sich in der intonatorischen Realisierung einer Äußerung widerspiegeln und welche nicht. Diese Fragestellung ergibt sich aus der Unterscheidung zwischen FS-Lesart einerseits und EPA- und R-Lesart andererseits und der fehlenden Unterscheidung zwischen EPA- und R-Lesart.

6 Relative Prominenz: Auflösung ambiger pronominaler Referenzen

Die wichtigste Verknüpfung zwischen einem anaphorischen Pronomen und dem zugehörigen Diskursreferenten, seinem Antezedens, ist die morpho-syntaktische Beziehung der Kongruenz in Numerus und Genus. Nicht immer kann jedoch allein anhand dieser im Deutschen obligatorischen grammatischen Relation ein Diskursreferent eindeutig identifiziert werden, nämlich dann nicht, wenn mehr als ein Referent mit Numerus- und Genus-Kongruenz im Diskurs verfügbar ist. Das Beispiel (1) zeigt eine solche ambige pronominale Referenz.

(1) Marlow erhält von dem Unbekannten ein Paket. Kurz darauf wird er verhaftet.

Das Pronomen *er* im zweiten Satz kann sowohl auf den Unbekannten als auch auf Marlow referieren, beide Referenten tragen, wie das Pronomen selbst, die Merkmale Singular und Maskulinum. Neben der obligatorischen grammatischen Kongruenzbeziehung wird die pronominale Referenz jedoch auch über andere, semantische und pragmatische Restriktionen gesteuert. Während die Ambiguität in (1) selbst für Kenner der Chandler-Stories nicht aufzulösen ist, gilt dies für Beispiel (2) nicht.

(2) Marlow erhält von dem Unbekannten ein Paket. Kurz darauf wird er ermordet.

Ein Hörer (oder Leser) dieses Satzes, der über das Wissen verfügt, daß Marlow der Held einer ganzen Serie von Detektivromanen ist und daher – zumindest in den Romanen Chandlers – in der Regel überlebt, wird als Mordopfer eindeutig den Unbekannten identifizieren. Restriktiver als die Disambiguierung mit pragmatischen Mitteln (Weltwissen, situatives Wissen; vgl. (2)) steuern lexikalisch-semantische Beschränkungen die Pronominalreferenz.

(3) a. Der Kater entdeckt den Vogel. Er fliegt weg.

b. Der Kater entdeckt den Vogel. Er schnurrt.

In diesem Beispiel trägt das Verb im zweiten Satz zur Auflösung der ambigen Pronominalreferenz bei. Die denotative Bedeutung von *Kater* umfaßt nicht die Eigenschaft, fliegen zu können. Daher kommt in (3)a nur der Vogel als Antezedens in Frage, während es in (3)b der Kater ist, da Vögel für gewöhnlich nicht schnurren. Ein weiteres Beispiel für eine lexikalisch-semantische Restriktion ist Präsupposition.

(4) Hans begrüßt Karl. Maria begrüßt ihn auch.

(4) kann nicht so verstanden werden, daß die Maria den Hans begrüßt. Aufgrund des Adverbs *auch* präsupponiert der zweite Satz eine vorangehende Proposition etwa der Form „eine andere Person als Maria begrüßt den Referenten von *ihn*“. Eine kohärente Interpretation des Diskurses ist daher nur möglich, wenn *ihn* auf *Karl* referiert, denn nur dann findet man im Kontext die passende Proposition „Hans (= eine andere Person als Maria) begrüßt Karl (= Referent von *ihn*)“.

Es gibt jedoch Fälle wie in Beispiel (1), in denen trotz semantischer und den erwähnten pragmatischen Überlegungen eine Auflösung der Ambiguität nicht möglich scheint. Das in Abschnitt 6.2 vorgestellte Experiment wird zeigen, inwiefern die Prosodie in solchen Fällen zur Disambiguierung beitragen kann. Zuvor, in Abschnitt 6.1, werden zwei Ansätze vorgestellt, die, basierend auf zusätzlichen pragmatischen Prinzipien, ein umfassendes Modell der Interpretation ambiger Pronomina entwickeln: die *Centering Theory* und das *Dynamic Preference Model*.

Das gesamte Kapitel beschäftigt sich ausschließlich mit der *anaphorischen* Verwendung von Pronomina, also mit Pronomina, deren Referenz durch den Bezug auf ein Antezedens im vorangegangenen Kontext bestimmbar ist. Um den Gegenstand der vorliegenden Untersuchung auf ein vernünftiges Maß einzuschränken, werden Pronomina, die sich auf den indexikalischen Kontext der jeweiligen Äußerungssituation beziehen (deiktische Verwendung) oder auf Entitäten eines implizit gemeinsamen Weltmodells der Diskurspartner, nicht berücksichtigt.

6.1 Strategien zur Disambiguierung pronominaler Referenzen

6.1.1 Centering und Präferenzen

Eines der einflußreichsten Modelle zur Interpretation anaphorischer Ausdrücke ist die *Centering Theory* (Grosz, Joshi & Weinstein, 1983; Grosz, Joshi & Weinstein, 1995). Centering beschreibt die dynamische Repräsentation von Diskursinhalten innerhalb einer spezifischen Komponente der von Grosz & Sidner (1986) vorgeschlagenen Diskursstruktur. Diese Komponente – *attentional state* – modelliert die ›kognitive Präsenz‹ von bestimmten Diskursinhalten zu einem bestimmten Zeitpunkt des Diskurses (Grosz *et al.*, 1995). Mit jeder neuen Äußerung, die Teil des Diskurses ist, wird das Modell der kognitiven Präsenz aktualisiert und insofern dynamisch verwaltet. Im Gegensatz zur propositionalen Komponente der Diskursstruktur ist das Präsenzmodell am Ende eines Diskurses leer (Grosz & Sidner, 1986), d.h. die einzelnen Diskursinhalte befinden sich nicht mehr im Zentrum der Aufmerksamkeit der Diskursteilnehmer. Die propositionale Komponente, die die expliziten und implizierten Propositionen des Diskurses modelliert, bleibt dagegen als Repräsentantin des semantischen und pragmatischen Diskurs-Resultats erhalten (Cahn, 1995).

Die zentrale Einheit des Centering-Formalismus sind *centers of attention*, kurz: *centers*. Sie sind folgendermaßen charakterisiert:

> „We use the term *centers* of an utterance to refer to those entities serving to link that utterance to other utterances in the discourse segment that contains it. It is an utterance (i.e., the uttering of a sequence of words at a certain point in the discourse) and not a sentence in isolation that has centers. The same sentence uttered in different discourse situations may have different centers. Centers are thus discourse constructs. Furthermore, centers are semantic objects, not words, phrases, or syntactic forms" (Grosz *et al.*, 1995: 208; Hervorhebung im Original).

Die Repräsentation der Centers im Präsenzmodell basiert auf zwei Verarbeitungsstrukturen: *forward-looking centers* und *backward-looking center*. Zu jeder Äußerung wird eine Menge (mit mindestens einem Element) von forward-looking centers (Cf) und genau ein backward-looking center (Cb) generiert (mit einer Ausnahme, s.u.). Cf ist eine Menge von Kandidaten für Bezugspunkte, zu denen die nachfolgende Äußerung eine Verknüpfung herstellen kann. Cb ist die Repräsentation eines linguistischen Ausdrucks, welche tatsächlich mit einem Cf-Element der vorangegangenen Äußerung verknüpft ist. Daher bildet die erste Äußerung eines Diskurses eine Ausnahme, sie generiert kein Cb, da es im Präsenzmodell noch keine Bezugskandidaten gibt (Grosz *et al.*, 1995: 208). Die Verknüpfung von Centers ist eine Relation innerhalb des Präsenzmodells, also eine Relation zwischen Repräsentanten linguistischer Ausdrücke, nicht zwischen den linguistischen Ausdrücken selbst.

Cf ist eine *geordnete* Menge, d.h. wenn mehrere Elemente in Cf aufgenommen werden, ist das Resultat nicht eine Menge von gleichberechtigten Kandidaten sondern eine nach Prominenzkriterien hierarchisch geordnete Liste. An der Spitze steht ein Element, welches – relativ zu den anderen Elementen der Menge – zum Zeitpunkt der Generierung von Cf über die größte kognitive Präsenz verfügt (Grosz *et al.*, 1995: 208). Deshalb ist das prominenteste Mitglied von Cf_n der beste Kandidat für Cb_{n+1}, den backward-looking center der nächsten Äußerung. Die Faktoren, die die Hierarchisierung von Cf-Elementen steuern, werden von Grosz *et al.* (1995) nicht erschöpfend diskutiert, es wird jedoch darauf hingewiesen, daß die grammatische Funktion eines linguistischen Ausdrucks eine bedeutende Rolle zu spielen scheint: Repräsentanten von Äußerungs*subjekten* sind in besonderem Maße prädestiniert, die Spitzenposition in Cf einzunehmen (Grosz *et al.*, 1995: 211ff). Eine ausführlichere Darstellung der Ordnungsfaktoren findet sich in Kameyama (1994a). Ich werde weiter unten in diesem Abschnitt darauf zurückkommen.

Die Centering Theory unterscheidet desweiteren drei Typen von Cf- bzw. Cb-Transitionen zwischen aufeinanderfolgenden Äußerungen desselben Diskurses: *center continuation* (CONT), *center retaining* (RET) und *center shifting* (SHIFT) (Grosz *et al.*, 1995: 210).

> CONT: Der backward-looking center einer Äußerungen $Cb(utt_n)$ ist identisch mit dem backward-looking center der vorangegangenen Äußerung $Cb(utt_{n-1})$ und die Entität, die von Cb repräsentiert wird, nimmt auch wieder die Spitzenposition in $Cf(utt_n)$ ein.

> RET: $Cb(utt_n)$ ist identisch mit $Cb(utt_{n-1})$, aber die Entität, die von Cb repräsentiert wird, nimmt nicht mehr die Spitzenposition in $Cf(utt_n)$ ein.

> SHIFT: $Cb(utt_n)$ ist nicht identisch mit $Cb(utt_{n-1})$.

Bevor anhand einiger Beispiele die Voraussagen der Centering Theory illustriert werden, hier zunächst eine kurze Zusammenfassung und schematische Darstellung (Abb. 6-1) der zentralen Konzepte.

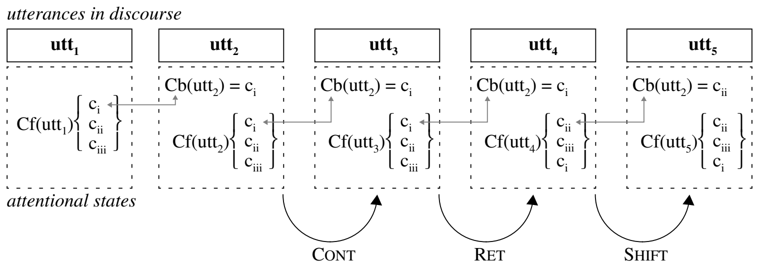


Abbildung 6-1: *Centering: Schematische Darstellung*

- CENTERS OF ATTENTION — semantische Objekte, die Teil des Aufmerksamkeits- oder Präsenzmodells eines Diskurses sind, nicht Teil des propositionalen Modells
- FORWARD- UND BACKWARD-LOOKING CENTERS — Strukturen zur Repräsentation und Verarbeitung von Centers
- HIERARCHISIERUNG VON Cf — die Elemente in Cf sind geordnet nach ihrer relativen ›kognitiven‹ Präsenz im Diskurs
- TRANSITIONEN — wohldefinierte Übergänge zwischen benachbarten Äußerungen bezüglich Cf und Cb

Nun zurück zum zentralen Punkt dieser Untersuchung: ambige anaphorische Pronomina. Pronomina sind linguistische Ausdrücke, die aufgrund der Referenzrelation zwischen Pronomen und Antezedens in prototypischer Weise eine Diskursäußerung mit anderen Äußerungen im Diskurskontext verknüpfen. Daher bilden Pronomina, wie alle Anaphern, bevorzugt den backward-looking center der Äußerung, in der sie auftauchen. Legt man nun den Centering-Formalismus zugrunde, erhält man selbst im Falle von ambigen Pronomina (auf die ich mich hier beschränken möchte) eine eindeutige Voraussage hinsichtlich des infragekommenden Antezedens’: diejenige Entität, die durch das prominenteste Element in Cf der vorangegangenen Äußerung repräsentiert wird. Damit kommt den Hierarchisierungsprinzipien, die die Elemente in Cf ordnen, eine entscheidende Bedeutung zu.

In (5) ist HANS[4] das einzige Element[5] in Cf(utt_a) und daher das passende Antezedens für ER, den Cb(utt_b). Cf(utt_b) enthält jedoch zwei Elemente, HANS (das Antezedens von *er*) und KARL, und damit zwei potentielle Antezedenten für das Pronomen in (5)c.

(5) a. Letzten Sonntag hat sich Hans gelangweilt.

b. Am Nachmittag hat er Karl angerufen, um ins Kino zu gehen.

c. Er wollte „Tote schlafen fest“ sehen.

Die Intuition, daß es Hans ist, der „Tote schlafen fest“ sehen möchte, stimmt überein mit der Voraussage der Centering Theory. Das Pronomen in (5)c ist der geeignete Kandidat für Cb(utt_c). Der beste Kandidat für die Spitzenposition in Cf(utt_b) ist aufgrund der Präferenz für Subjekte HANS, das Subjekt bzw. das Antezedens des Subjekts in (5)b. Durch die bestehende Verknüpfung zwischen dem prominentesten Element in Cf(utt_b) – HANS – und dem Cb(utt_c) erhält man die erwünschte Koreferenz zwischen *Hans* und dem Pronomen *Er* in (5)c und damit die oben intuitiv motivierte Interpretation. Ein weiteres in der Centering Theory formuliertes Präferenzprinzip unterstützt diese Interpretation: die Bevorzugung von CONT-Sequenzen gegenüber RET- oder SHIFT-Sequenzen (Grosz *et al.*, 1995: 215). Dieses Prinzip reflektiert die Intuition, daß die Beibehaltung eines attentional centers über mehrere Äußerungen hinweg entscheidend zur lokalen Kohärenz eines Diskurses beiträgt. Wendet man dieses Prinzip auf (5) an, erhält man auf der Grundlage der beiden alternativen Analysen (6)a und (6)b wiederum die erwünschte Interpretation.

(6) a. Cf(utt_a){HANS} ⟶ Cb(utt_b) = HANS, Cf(utt_b){HANS, KARL} $\xrightarrow{\text{CONT}}$ Cb(utt_c) = HANS, Cf(utt_c){HANS}

b. Cf(utt_a){HANS} ⟶ Cb(utt_b) = HANS, Cf(utt_b){KARL, HANS} $\xrightarrow{\text{SHIFT}}$ Cb(utt_c) = KARL, Cf(utt_c){KARL}

Eine Weiterentwicklung der Centering Theory stellt Kameyamas (1994a) *Dynamic Preference Model* dar. Kameyama testet und diskutiert verschiedene Strategien, die angewandt werden können, um ambige Pronominalreferenzen aufzulösen. Ihre Arbeit konzentriert sich auf den Aspekt der Gewichtung von Elementen in Cf und der formalen Repräsentation dieses Prozesses. Das Dynamic Preference Model formuliert eine Reihe von Prinzipien und Interaktionsregeln, um die empirisch nachgewiesenen Interpretationspräferenzen zu beschreiben. Diese Regeln und Prinzipien sind Teil eines formalen pragmatischen Systems und – im Gegensatz zu lexikalisch-semantischen und morphosyntaktischen Regeln – nicht obligatorisch, d.h. sie können außer Kraft gesetzt und überschrieben werden. Ich werde mich im folgenden auf die Beschreibung der von Kameyama (1994a) nachgewiesenen Präferenzeffekte und ihrer Interferenzen konzentrieren.

Kameyama unterscheidet 4 Haupteffekte: *commonsense preference, subject antecedent preference, grammatical parallelism preference* und *pronominal chain preference*. Die Effekte selbst sowie die Interferenzen weist sie mit Hilfe eines Leseexperimentes nach. Die Versuchspersonen mußten Testdiskurse lesen (z.B. *John hit Bill. Mary told him to go home.*) und danach entscheiden, wer was getan hat.

[4] Die Objekte im Präsenzmodell sind, wie oben bereits erwähnt, Repräsentationen linguistischer Ausdrücke. Sie stehen daher in Kapitälchen.

[5] Eine Diskussion darüber, welche der im Diskurs behandelten Entitäten im Präsenzmodell repräsentiert werden, ist für die hier vorgestellte Untersuchung nicht relevant und würde den Rahmen dieser Arbeit sprengen; es sei hierfür auf die Arbeit von Grosz *et al.* (1995) verwiesen. Ich werde im folgenden nur diejenigen Entitäten berücksichtigen, die im konkreten Beispiel als Antezedenten für vorhandene Pronomina in Frage kommen.

Der stärkste Effekt, d.h. der für Interferenzen durch andere Effekte am wenigsten anfällige, ist commonsense preference (Kameyama, 1994a: 13f), der auch schon für die Auflösung der Ambiguität in (2) verantwortlich war. Das situative Wissen bzw. Weltwissen, über das (idealerweise) alle beteiligten Diskurspartner verfügen, diktiert die Interpretation der pronominalen Referenz. Diese Regel kann nur durch obligatorische Restriktionen der anderen Teilsysteme (Morpho-Syntax, lexikalische Semantik) überschrieben werden, nicht jedoch durch eine der übrigen pragmatischen Präferenzregeln. (7), als Minimalpaar zu (2), zeigt diese Unterlegenheit von commonsense preference gegenüber obligatorischen Restriktionen am Beispiel der morpho-syntaktischen Kongruenz zwischen Pronomen und Antezedens.

(7) Marlow erhält von der Unbekannten ein Paket. Kurz darauf wird er ermordet.

Das einzig verfügbare Antezedens in mask./sing. und mithin das einzig mögliche Mordopfer ist in diesem Falle Marlow. Diese Koreferenz muß jeder Leser/Hörer anerkennen, auch entgegen besseren (Welt-) Wissens.

Der schwächste Effekt, der hier nur der Vollständigkeit halber erwähnt sei, ist pronominal chain preference. Damit gemeint ist die bevorzugte Koreferenz einer Kette von Pronomina über mehrere Sätze hinweg wie in (8).

(8) Marlow war pünktlich am Treffpunkt. Er erhielt von dem Unbekannten ein Paket. Kurz darauf wurde er verhaftet.

Kameyama (1994a: 10ff) weist nach, daß ein Leser eines solchen Diskurses eindeutig einen gemeinsamen Referenten für beide Pronomina bevorzugt. Da sich das Pronomen im zweiten Satz nur auf *Marlow* beziehen kann, kommt somit auch für das Pronomen im dritten Satz bevorzugt *Marlow* als Antezedens in Frage. Sie liefert damit die experimentelle Bestätigung der in der Centering Theory postulierten Präferenz für CONT-Sequenzen (s.o.). Da wir jedoch in dem unten beschriebenen Experiment nur 2-Satz-Diskurse verwendet haben, ist dieser Effekt für die vorliegende Untersuchung nicht relevant.

Subject antecedent preference ist eine Reformulierung des Centering-Prinzips, daß grammatische Subjekte bevorzugt an der Spitze von Cf stehen. Kameyama weist außerdem explizit nach, daß es die grammatische Funktion und nicht die thematische Rolle ist, die die interne Ordnung der Menge der verfügbaren Centers determiniert (Kameyama, 1994a: 10ff). Beispiele (9) und (10) sind Übersetzungen von Kameyamas Testdiskursen:

(9) Hans schlug Karl. Maria schickte ihn nach Hause.

(10) Karl wurde von Hans geschlagen. Maria schickte ihn nach Hause.

In (9) bevorzugen die Versuchspersonen *Hans* als Antezedens, in (10) *Karl*, also in beiden Fällen das Subjekt des ersten Satzes, unabhängig von der jeweiligen θ-Rolle (in beiden Sätzen ist *Hans* Agens und *Karl* Patiens).

Ein Interferenzeffekt entsteht bei zusätzlicher commonsense preference, die Subjektpräferenz kann unter dem Einfluß von Weltwissen überschrieben werden wie das folgende Beispiel zeigt. Das Wissen um kausale Zusammenhänge verschiebt in (11) die referentielle Präferenz zu Gunsten von *Karl*.

(11) Hans schlug Karl. Er wurde schwer verletzt.

Der vierte von Kameyama beschriebene Effekt ist grammatical parallelism preference, „a general preference for two adjacent utterances to be grammatically parallel“ (Kameyama, 1994a: 13). Grammatikalische Parallelität bedeutet Koreferenz zwischen den jeweiligen Subjekten zweier benachbarter Sätze und – sofern vorhanden – Koreferenz zwischen den jeweiligen Objekten. Ein Beispiel:

(12) Hans schlug Karl. Er mag ihn nicht.

Die parallele Interpretation der beiden Pronomina im zweiten Satz von (12) führt zur Koreferenz zwischen *Er* und *Hans* (Subjekte) und zwischen *ihn* und *Karl* (Objekte). Die Parallelitätspräferenz interferiert stets mit der Subjektpräferenz, da beide Prinzipien auf der grammatischen Funktion der verfügbaren Antezedenten basieren. Führen beide Prinzipien zum selben Ergebnis wie in (12) und (10), so wird der Effekt verstärkt. Führen sie jedoch zu einem widersprüchlichen Ergebnis, gibt es keinen eindeutigen ›Gewinner‹. Betrachten wir zunächst den Fall mit einem Pronomen (Beispiel (9) von oben):

(13) Hans schlug Karl. Maria schickte ihn nach Hause.

Folgt man dem Parallelitätsprinzip, ist aufgrund der Objekt-Koreferenz *Karl* das bevorzugte Antezedens. Die Versuchspersonen in Kameyamas Studie wählten jedoch ausschließlich *Hans* als Antezedens (Kameyama, 1994a: 11), folgen also dem Subjektprinzip. Infolgedessen ist nach Kameyamas Analyse die Subjektpräferenz der Parallelitätspräferenz überlegen. Betrachtet man jedoch den folgenden Fall mit zwei Pronomina, wird diese Überlegenheit in Frage gestellt:

(14) Hans schlug Karl. Ihn mag er nicht.

Wenn man von einer seriellen Verarbeitung der Pronomina ausgeht, sagt das Subjektprinzip die Interpretation „der Karl mag den Hans nicht“ voraus, da es ausschließlich auf der grammatischen Funktion der potentiellen Antezedenten basiert und die grammatische Funktion der Anaphern unberücksichtigt läßt: Das erste Pronomen *Ihn* präferiert das Subjekt *Hans* als Antezedenten, während für das zweite Pronomen *er* nur das Objekt *Karl* übrigbleibt. Dagegen ist die meines Erachtens bevorzugte Interpretation von (14), daß der Hans den Karl nicht mag, also Koreferenz zwischen den Subjekten *Hans* und *er* sowie zwischen den Objekten *Karl* und *Ihn* – entsprechend der Voraussage des Parallelitätsprinzips[6].

Zusammenfassung: Als eine Theorie der Diskurskohärenz entwickelt die Centering Theory ein umfassendes Modell zur Interpretation anaphorischer Ausdrücke. Die Analyse basiert auf der Modellvorstellung einer ›Aufmerksamkeitsstruktur‹ (*attentional structure*) und nimmt Bezug auf die ›kognitive Präsenz‹ von Diskursentitäten. Der entscheidende Aspekt – zumindest für die Behandlung ambiger Pronomina – ist die hierarchische Ordnung von Cf-Elementen, die sich an der relativen kognitiven Präsenz der Elemente orientiert. Ergänzend zum Subjektprinzip der Centering Theory formuliert Kameyama (1994a) vier Präferenzprinzipien, die Rückschlüsse auf die Prominenz von Cf-Elementen und deren interne Ordnung erlauben.

Keines der beiden Modelle berücksichtigt prosodisch markierte Pronomina. Im nächsten Abschnitt wird gezeigt, *daß* Prosodie die Interpretation von Pronomina beeinflussen kann und wie prosodisch markierte Pronomina im Rahmen von Centering analysiert werden können.

6.1.2 Akzent, Fokus und komplementäre Präferenzen

Das klassische Beispiel für prosodisch markierte Pronomina und ihren Einfluß auf die Interpretation eines Satzes bzw. Diskurses sind Minimalpaare wie in (15) (nach Lakoff, 1971).

[6] Diese Interpretation erhält man sowohl mit neutralem Satzakzent auf *mag* als auch mit Akzent auf dem Pronomen *Ihn*. Der problematische Fall, der gegen die eindeutige Überlegenheit des Subjektprinzips spricht, ist jedoch nur derjenige mit neutralem Satzakzent. Mit akzentuiertem Pronomen würde auch Kameyamas Modell (zumindest die erweiterte Version in Kameyama, 1994b) die hier vorgeschlagene Interpretation voraussagen (s. Abschnitt 6.1.2).

(15) a. Hans nannte Karl einen Republikaner
und dann beLEIdigte er ihn.[7]

b. Hans nannte Karl einen Republikaner
und dann beleidigte ER IHN.

Die bevorzugte Interpretation von (15)a ist, daß *er* mit *Hans* und *ihn* mit *Karl* koreferiert. Diese Präferenz ist kompatibel mit dem Centering-Formalismus und den Voraussagen von Kameyamas Präferenzprinzipien. Das Subjektprinzip prognostiziert die Verknüpfung *er* := *Hans*, das Parallelitätsprinzip die Verknüpfung *ihn* := *Karl*; commonsense-Präferenz und Pronominalketten-Präferenz spielen in diesem Fall keine Rolle. In (15)b dagegen wird – entgegen der Voraussage der erwähnten Prinzipien – Koreferenz zwischen *er* und *Karl* bzw. zwischen *ihn* und *Hans* bevorzugt.

Das Beispiel (15) zeigt, daß bei der Untersuchung gesprochener Sprache die Prosodie, hier speziell die Plazierung des Satzakzents, in der Analyse ambiger Pronomina und ihrer Interpretation berücksichtigt werden muß. Ein Modell, welches die kognitive Präsenz semantischer Objekte nur allein aufgrund von textuell repräsentierten Merkmalen errechnet, wird dieser Anforderung nicht gerecht und kann, wie in (15)b, zu falschen Voraussagen führen. Ich werde im folgenden einige neuere Ansätze skizzieren, die versuchen, prosodisch markierte Pronomina und die damit verbundenen Phänomene im Rahmen des Centering-Modells zu beschreiben.

Akzentuierte Pronomina stellen eine Ausnahme dar, d.h. sie sind markiert im Sinne der linguistischen Markiertheitstheorie (Trubetzkoy, 1958; Jakobson, 1990; Battistella, 1990). Im unmarkierten Fall sind Pronomina wie die meisten sogenannten Funktionswörter nicht akzentuiert. Im Rahmen einer diskurstheoretischen Analyse kann dies folgendermaßen begründet werden: (1.) Pitchakzente zeichnen insbesondere den *propositionalen* Gehalt der akzentuierten linguistischen Einheiten aus (vgl. Abschnitt 2.3).

[7] Akzentuierte Wörter oder Silben stehen entsprechend der üblichen Konvention in Versalien. Sofern der Akzent nicht näher spezifiziert ist, kann ein Peak-Akzent (H*[X]) angenommen werden.

Daher scheint es im Hinblick auf die Qualität der übermittelten Information sinnvoller, informationsreiche nicht-pronominale Einheiten zu akzentuieren als semantisch verarmte Pronomina (Cahn, 1995). (2.) Brown (1983) und andere Studien haben gezeigt, daß tendenziell eher linguistische Einheiten, die neue (*new*) Information in den Diskurs einführen, akzentuiert werden als Einheiten, die im Diskurskontext bereits vorerwähnt sind (*given*). Anaphorische Pronomina beziehen sich stets auf einen Diskursreferenten des vorangehenden Kontexts und führen infolgedessen keine neue Information ein. Sie sind daher nicht sehr geeignet, einen Pitchakzent zu tragen.

Wie ist nun aber der markierte Fall – akzentuierte Pronomina –, der in (15) illustriert ist, zu erklären? Neben der propositionalen Komponente von Pitchakzenten kann auch eine Aufmerksamkeitskomponente im Sinne der Centering Theory angenommen werden, d.h. es gibt einen spezifischen Beitrag von Pitchakzenten zur Konstitution von attentional states. Stellt man diese Komponente in den Mittelpunkt der Analyse, löst sich der Widerspruch zwischen Akzentuierung und Pronominalisierung auf. Die Annahme von Hirschberg & Pierrehumbert (1986), daß Pitchakzente die kognitive Präsenz von Repräsentationen linguistischer Einheiten steigern, also einen direkten Beitrag zur Plazierung der akzenttragenden Einheit im Präsenzmodell leisten, liefert allerdings noch nicht die gewünschte Erklärung. Pronominale Referenten verfügen ohnehin – wie in 6.1.1 erwähnt – über die maximale kognitive Präsenz im Präsenzmodell der Äußerung, in der sie stehen, sie sind die optimalen Kandidaten für den backward-looking center Cb. Eine Steigerung der Präsenz durch Akzentuierung wäre also erstens redundant und würde zweitens nach wie vor nicht die veränderte Interpretation von (15)b gegenüber (15)a erklären. Verschiedene neuere Arbeiten schlagen dagegen eine eher indirekte Analyse der Aufmerksamkeitskomponente von Pitchakzenten vor (Cahn, 1990; Kameyama, 1994b; Cahn, 1995; Nakatani, 1997). Allen diesen Arbeiten gemeinsam ist die Grundannahme, daß Akzente die hierarchische Ordnung von Cf-Elementen in einer für die Interpretation der aktuellen Äußerung relevanten Cf-Liste manipulieren[8]. Die Funktion eines Pitchakzents im Hinblick auf die Aufmerksamkeitskomponente der Diskursstruktur ist also nicht auf die akzenttragende Einheit und ihre Repräsentation selbst bezogen, sondern auf eine unabhängige, zum Zeitpunkt der Äußerung gegebene Struktur innerhalb des Präsenzmodells. Betrachtet man Beispiel (15), so sind die Cf-Listen, die vom jeweils ersten Teil der Äußerung (*Hans nannte Karl einen Republikaner*) konstituiert werden, in (15)a und (15)b identisch: {HANS > KARL}. In (15)a werden die pronominalen Referenzen im zweiten Teil (*und dann beleidigte er ihn*) entsprechend dieser Hierarchie aufgelöst. In (15)b dagegen verursachen die akzentuierten Pronomina zunächst die Reorganisation der Cf-Liste und werden dann entsprechend der neuen Hierarchie {KARL > HANS} aufgelöst.

Kameyama (1994b) verbindet diesen Ansatz der Reorganisation von Cf mit der Fokustheorie von Rooth (1992) und entwickelt daraus ein integriertes Modell der Interpretation akzentuierter und nicht-akzentuierter Pronomina im Diskurs. Dieses Modell, das ich im folgenden kurz skizzieren werde, bildet die theoretische Basis unserer experimentellen Arbeit.

Die Rooth'sche Theorie der *restricted alternative semantics* (Alternativensemantik) schlägt folgende Analyse der Fokusinterpretation vor: Eine Fokusphrase α präsupponiert eine semantische Restriktion $\sim C$, wobei C eine anaphorische Variable ist, deren Wert eine kontextuell determinierte Menge von Diskursentitäten mit mindestens zwei Elementen ist. Ein Element von C ist die Denotation von α, daneben enthält C mindestens ein weiteres Element, welches mit der Denotation von α kontrastiert. Diese Menge der kontextuell verfügbaren Alternativen (C) heißt *focus semantic value* (fokussemantischer Wert) von α, $[\![\alpha]\!]^f$; die Denotation von α heißt *ordinary semantic value* (regulärer semantischer Wert) von α, $[\![\alpha]\!]^o$. Es gilt also $[\![\alpha]\!]^o \in [\![\alpha]\!]^f = C$ mit $C \subseteq E$. E steht für die Domäne der alternativen semantischen Werte. Ist z.B. ein Satz oder eine Phrase fokussiert, so ist C eine (kontextuell verfügbare) Teilmenge der Menge der Propositionen. Ist dagegen ein Eigenname fokussiert, so ist C eine Teilmenge der Menge der Individuen. Ein Beispiel:

(16) Maria mag Hammet.

Der reguläre semantische Wert des Satzes in (16) ist eine Proposition der Form *mögen*(*Maria*, *Hammet*), formalisiert in (17)

(17) $[\![\text{Maria mag Hammet}]\!]^o = \textbf{mögen}(\textbf{m},\textbf{h})$

Für die fokussierte Variante in (18) erhält man die Analyse in (19):

(18) F: Liest Maria gerne Chandler?

A: Nein. Maria mag $[\text{Hammet}]_F$.

(19) $[\![\text{Maria mag } [\text{Hammet}]_F]\!]^f = \{\textbf{mögen}(\textbf{m}, x) \mid x \in C \subseteq E\}$

Die Menge C in (19) enthält die kontextuell verfügbaren Individuen *Chandler* und *Hammet*. Der fokussemantische Wert ist daher die Menge der Propositionen {mögen(Maria, Hammet), mögen(Maria, Chandler)}, also der reguläre semantische Wert plus ein kontrastierendes Element.

Vereinfacht gesagt besteht die Fokusinterpretation nach Rooth (1992) also darin, aus einer kontextuell zu bestimmenden Menge von Alternativen genau ein Element zu isolieren, um damit die durch die Fokussierung präsupponierte Restriktion $\sim C$ aufzulösen und Kohärenz herzustellen. Ausgehend von der Beobachtung, daß bei der Interpretation anaphorischer Pronomina ganz ähnliche Arbeitsschritte involviert sind – aus einer kontextuell zu bestimmenden Menge von möglichen Referenten (Cf) muß genau ein Element aus-

[8] Zur Modellierung der lokalen Diskurskohärenz und in Übereinstimmung mit dem klassischen Centering-Ansatz (Grosz *et al.*, 1995) ist für die Interpretation der aktuellen Äußerung ausschließlich die Cf-Liste relevant, die durch die unmittelbar vorangehende Äußerung konstituiert wird. Nakatani (1993, 1994, 1997) zeigt jedoch, daß bei der Analyse längerer Diskurse auch „ältere" *attentional state*-Elemente berücksichtigt werden müssen, die nicht in der aktuellen Cf-Liste stehen. Solche Aspekte der globalen Diskurskohärenz spielen aber bei der in diesem Kapitel vorgestellten Untersuchung von 2-Satz-Diskursen keine Rolle und werden daher an dieser Stelle nicht ausführlicher behandelt.

gewählt werden, um Kohärenz herzustellen – entwickelt Kameyama (1994b) ihr Modell der Anaphernauflösung. Die zentralen Thesen ihres Ansatzes sind folgende:

- Die Interpretation eines akzentuierten (=fokussierten) Pronomens folgt aus der Interpretation seines nicht-akzentuierten (=nicht-fokussierten) Gegenstücks (bei identischem Diskurskontext)
- Die Präferenz eines akzentuierten Pronomens ist komplementär zu der Präferenz seines nicht-akzentuierten Gegenstücks

Da wir nun bei den Grundlagen des unten beschriebenen Experiments angelangt sind, werde ich Kameyamas Argumentation mit einem Beispiel aus dem experimentellen Material (vgl. 6.2.2) illustrieren.

(20) a. Der Ober verhandelt mit dem Gast.

b. ER / Er hat kein Kleingeld.

Betrachten wir zunächst die nicht-akzentuierte Variante. Analog zur Rooth'schen Analyse schlägt Kameyama die folgende Repräsentation vor: $[\![Er]\!]^{f*} = \{x \mid x \in B \subseteq E\}$, wobei B für eine ›kognitiv präsente‹, kontextuell determinierte Menge von möglichen Referenten steht. Im Unterschied zu $[\![\alpha]\!]^{f}$ mit mindestens zwei Elementen, hat $[\![\alpha]\!]^{f*}$ nur mindestens ein Element, nämlich $[\![\alpha]\!]^{o}$, also die Denotation, d.h. den tatsächlichen Referenten des Pronomens. Die entscheidenden Schritte zur Interpretation einer derartigen Repräsentation sind nun (1.) die Lokalisation von B, d.h. die Bestimmung einer relevanten Menge möglicher Referenten aus dem Kontext, und (2.) die Auswahl eines Elementes aus B, d.h. des tatsächlichen Referenten. Ein geeigneter Formalismus hierfür ist, wie im letzten Abschnitt gezeigt wurde, die Centering Theory und die Präferenzprinzipien von Kameyama (1994a).

Für die akzentuierte Variante von (20)b erhält man die folgende Repräsentation: $[\![ER]\!]^{f} = \{x \mid x \in C \subseteq E\}$. Kameyama postuliert, daß C dieselben Elemente enthält wie B: „Stressed and unstressed counterparts choose their values from the *same salient subset* of the domain of individuals“ (Kameyama, 1994b: 478; Hervorhebung im Original), es gilt also $C = B \subseteq E$. Der einzige Unterschied zwischen beiden Varianten ist die Ordnung der Elemente in C bzw. B. Während die Hierarchie in B regulär gemäß der Präferenzprinzipien errechnet wird, ist die Hierarchie in C – so die These Kameyamas – genau komplementär: „A focused pronoun takes the complementary preference of the unstressed counterpart“ (Kameyama, 1994b: 481). Daraus ergibt sich das folgende System pragmatischer Subroutinen zur Interpretation eines akzentuierten Pronomens in utt_n (vgl. Kameyama, 1994b: 482):

(21) a. isoliere den attentional state $B_{n\text{-}1}$ im Kontext $K_{n\text{-}1}$

b. berechne die Präferenzhierarchie $H_{n\text{-}1}$ der nicht-akzentuierten Variante für $B_{n\text{-}1}$

c. berechne die komplementäre Hierarchie $H'_{n\text{-}1}$ aus $H_{n\text{-}1}$

d. löse die Präsupposition $\sim C_n$

e. etabliere die Kohärenz von $[\![utt_n]\!]^{o}$

Wendet man diesen Algorithmus auf (20)b an, ergibt sich die folgende Interpretation:

(22) a. $[[\text{ER}]_F$ hat kein Kleingeld$]_n \sim C$, $B_{n\text{-}1} = \{\text{OBER, GAST}\}$

b. $H_{n\text{-}1} = \{\text{OBER} > \text{GAST}\}$

c. $H'_{n\text{-}1} = \{\text{GAST} > \text{OBER}\}$

d. $[\![\text{ER}]\!]^{f} = \{\text{Gast, Ober} \mid [\![\text{ER}]\!]^{o} = \text{Gast}\}$

e. $[\![utt_n]\!]^{o}$ = hat-kein-Kleingeld(Gast)

Zusammenfassung: Kameyamas Analyse, basierend auf der Integration von Fokussemantik (Rooth, 1992), Centering (Grosz *et al*., 1995) und Präferenzprinzipien (Kameyama, 1994a), begründet die Hypothese, daß akzentuierte Pronomina ihren Referenten aus der gleichen Menge von Alternativen auswählen wie ihr nicht-akzentuiertes Gegenstück (identische syntaktische Position und identischer Diskurskontext vorausgesetzt), daß diese Menge jedoch komplementär geordnet ist, was eine Umkehrung der Präferenz zur Folge hat. Kameyama (1994b) hat diese Hypothese nicht experimentell überprüft, dies soll mit dem folgenden Experiment nachgeholt werden.

6.2 Experiment

6.2.1 Vorbemerkungen

Das in diesem Abschnitt vorgestellte Experiment ist als Wahrnehmungs- und Entscheidungsexperiment gestaltet. Ein Leseexperiment, wie es von Kameyama (1994a) zur Untersuchung von nicht-akzentuierten Pronomina eingesetzt wurde, erscheint uns aus folgenden Gründen problematisch: Bei der vergleichenden Untersuchung von nicht-akzentuierten und akzentuierten Pronomina kommt der exakten Kontrolle der prosodischen Eigenschaften der Testitems eine entscheidende Bedeutung zu. Diese Kontrolle ist in einem Leseparadigma nicht gewährleistet, wenn die Versuchspersonen angewiesen werden, visuell vorgegebene Stimuli zu lesen, mit einem bestimmten prosodischen Muster wiederzugeben und dann zu interpretieren. Bei diesem Verfahren sind zum einen zu viele Verarbeitungsschritte involviert, so daß der Ursprung der letztlich protokollierten Interpretation nicht eindeutig bestimmbar ist. Zum anderen ist die adäquate Realisation des vorgegebenen prosodischen Musters nicht bzw. nur nach ausführlicher Instruktion und Übung sichergestellt. Außerdem ist die Evaluation der produzierten prosodischen Muster häufig nur Offline und mit erheblichem Aufwand möglich. Mit der auditiven Darbietung von vorab aufbereitetem, einheitlichem Stimulusmaterial werden diese Nachteile umgangen. Die prosodischen Muster sämtlicher Testitems – also auch derjenigen mit nicht-akzentuierten Pronomina – können exakt kontrolliert und evaluiert werden und die Reaktionen der Versuchspersonen basieren auf einem einheitlichen und weniger komplexen Verarbeitungsprozeß.

In dem Experiment werden, wie bereits erwähnt, ausschließlich anaphorische Pronomina in 2-Satz-Diskursen untersucht. Wir haben versucht, die Diskurse so zu konstruieren, daß der Einfluß von Weltwissen (*commonsense*-Effekt) auf die Interpretation möglichst ausgeschlossen wird, da dieser Einfluß kaum kontrollierbar und damit kaum ›meßbar‹ ist. Das Experiment dient also ausschließlich der Evaluation des Zusammenspiels zwischen den in Abschnitt 6.1.1 beschriebenen grammatikalischen Präferenzprinzipien – Subjektpräferenz und Parallelitätspräferenz – und der in Abschnitt 6.1.2 formulierten Hypothese der komplementären Präferenz. Die Kontrolle der prosodischen Variation in dem experimentellen Material beruht auf dem Tonsequenzmodell (s. Teil I), da sich dieser Ansatz der Modellierung prosodischer Prominenz auf Satzebene derzeit international am weitesten durchgesetzt hat. Das heißt, daß prosodische Variation in diesem Experiment ausschließlich durch die Manipulation des Tonhöhenverlaufs realisiert wurde, da der Tonhöhenverlauf das wichtigste und stabilste Korrelat der Satzakzentuierung darstellt. Alle anderen Parameter, die als Korrelate der Wortbetonung im Deutschen angesehen werden können, wie z.B. Segment- oder Silbendauer, Intensität und *spectral tilt* (vgl. Neppert & Pétursson, 1986: 144ff; Claßen, 1996) blieben unverändert. Ebenso blieben prosodische Eigenschaften, die der globalen Strukturierung von Äußerungen dienen wie z.B. Pausenlänge und Tonregister, unverändert (vgl. Kapitel 7).

Dasselbe Experiment wurde auch mit Patienten mit neurologischen Beeinträchtigungen durchgeführt (Bender, 1997). Daraus ergeben sich einige Besonderheiten in Bezug auf die Zusammenstellung des Materials und den Modus der Durchführung, auf die ich an entsprechender Stelle hinweisen werde.

6.2.2 Material

6.2.2.1 Diskurse

Für das Experiment wurden neun zugrundeliegende Diskurse konstruiert. Durch geringfügige Variationen (s.u.) ergaben sich insgesamt 32 Diskurse, die sich textuell unterschieden. Alle Diskurse bestanden aus 2 Sätzen; der erste Satz führte jeweils zwei Diskursreferenten ein, im zweiten Satz standen ein oder zwei Pronomina. 14 der insgesamt 32 Diskurse waren ambig, d.h. es bestand morpho-syntaktische Kongruenz zwischen dem Pronomen bzw. den Pronomina im zweiten Satz und den beiden Referenten des ersten Satzes. Bei den restlichen 18 Diskursen waren die Referenten im ersten Satz von unterschiedlichem grammatischem Geschlecht, so daß eine eindeutige Beziehung zwischen Pronomina und Referenten bestand. Diese nicht-ambigen Diskurse dienten in erster Linie der Leistungsbeurteilung aphasischer Versuchspersonen bei der oben erwähnten Durchführung des Experimentes mit neurologischen Patienten. Im Rahmen des hier vorgestellten Experimentes mit gesunden Versuchspersonen wurden diese Diskurse als Füllmaterial betrachtet und nicht ausgewertet.

Die neun zugrundeliegenden Diskurse lassen sich nach 2 Merkmalen klassifizieren: (1.) ein vs. zwei Pronomina im zweiten Satz und (2.) Austauschbarkeit der Referenten im ersten Satz. Tabelle 6-1 gibt eine Übersicht über die neun Diskurse, geordnet nach Klassen. Die lexikalisch-semantischen Restriktionen der Verben im ersten Satz der Diskurse 1 bis 5 erlauben die Vertauschung der Reihenfolge der Referenten, bei den Diskursen 6 bis 9 ist dies dagegen nicht möglich. Von den Diskursen 1 bis 5 wurden beide Versionen verwendet, z.B. *Der Ober verhandelt mit dem Gast. Er hat kein Kleingeld.* und *Der Gast verhandelt mit dem Ober. Er hat kein Kleingeld*. Durch diese Verdopplung der Diskurse 1 bis 5 erhält man die Gesamtzahl von 14 ambigen Diskursen (vgl. Anhang A).

Tabelle 6-1: Zugrundeliegende Diskurse

	ein Pronomen	zwei Pronomina
Referenten austauschbar	**1.** Die Enkelin verreist nicht mit ihrer Oma. Sie ist krank. **2.** Der Ober verhandelt mit dem Gast. Er hat kein Kleingeld.	**3.** Die Nichte besucht ihre Tante. Sie zeigt ihr neue Fotos. **4.** Der Motorradfahrer und der Radfahrer haben einen Unfall. Er hat ihm die Vorfahrt genommen. **5.** Die Enkelin wäscht bei der Oma. Sie muß ihr die Maschine erklären.
nicht austauschbar	**6.** Die Verkäuferin berät die Kundin. Sie wird ungeduldig.	**7.** Der Maurer schimpft mit dem Jungen. Er zeigt ihm den Vogel. **8.** Der Maler malt den Sänger. Er ärgert sich über ihn. **9.** Die Masseurin behandelt die Patientin im Wohnzimmer. Sie ist zu ihr nach Hause gekommen.

Tabelle 6-2 listet auf, durch welche Referenten die Originalreferenten der ambigen Diskurse ersetzt wurden, um die anaphorische Beziehung zu disambiguieren. Bei Diskursen mit einem Pronomen wurden zwei Versionen verwendet: Eine Version, in der sich das Pronomen auf den einen Referenten bezog (z.B. *Die Kellnerin verhandelt mit dem Gast. Sie hat kein Kleingeld.*) und eine andere Version, in der sich das Pronomen auf den anderen Referenten bezog (*Die Kellnerin verhandelt mit dem Gast. Er hat kein Kleingeld.*). Bei Diskursen mit zwei Pronomina wurden ebenfalls zwei Versionen verwendet: Eine grammatikalisch parallele Konstruktion (z.B. *Die Nichte besucht ihren Onkel. Sie zeigt ihm neue Fotos.*) und eine komplementäre Konstruktion (*Die Nichte besucht ihren Onkel. Er zeigt ihr neue Fotos.*). Dies ergibt die erwähnten 18 nicht-ambigen Diskurse.

Tabelle 6-2: Substituierte Referenten zur Disambiguierung der pronominalen Referenz.

Diskurs #	original	Substitut
1	Oma	Opa
2	Ober	Kellnerin
3	Tante	Onkel
4	Radfahrer	Radfahrerin
5	Enkelin	Enkel
6	Verkäuferin	Verkäufer
7	Junge	Mädchen
8	Sänger	Sängerin
9	Masseurin	Masseur

6.2.2.2 Prosodie

Die 32 textuell unterschiedlichen Diskurse wurden in zufälliger Reihenfolge von einer Sprecherin gesprochen und auf DAT-Band aufgenommen. Die Aufnahme fand in einem normalen Raum statt, ohne besondere qualitätsverbessernde Vorkehrungen. Die Sprecherin wurde gebeten, die Diskurse möglichst deutlich vorzulesen; weitere Instruktionen wurden nicht gegeben. Die Aufnahmen wurden dann auf eine SGI Indy Workstation überspielt (16 kHz, 16 Bit) und mit ESPS/*waves* und dem Stuttgarter PSOLA-System (s. Kapitel 4) weiterverarbeitet. Den folgenden Arbeitsschritten wurden alle 32 Diskurse unterworfen:

a. Trennung der Diskurse zwischen dem ersten und zweiten Satz; Speichern der Segmente in separaten Dateien (32*2 = 64 Dateien)

b. automatisches Labelling auf Wort- Silben- und Phonemebene (vgl. Rapp, 1995), manuell korrigiert

c. Berechnung der Grundfrequenz F_0 (ESPS *get_f0*)

d. automatisches Pitch-synchrones Labelling (ESPS *epochs*)

e. PSOLA-Resynthese des ersten Satzes von jedem Diskurs ohne Veränderung der originalen Grundfrequenzkontur

f. PSOLA-Resynthese des zweiten Satzes von jedem Diskurs mit jeweils 3 bzw. 7 Varianten (s.u.)

g. Rekonstruktion der kompletten Diskurse, d.h. Zusammensetzen des Outputs der Schritte e und f

Die Schritte a-d dienten der Vorbereitung des Materials für die PSOLA-Resynthese. Der erste Satz der Diskurse, der die jeweiligen Referenten einführt und prosodisch nicht verändert werden sollte, wurde dennoch resynthetisiert (Arbeitschritt e), um die Signalqualität innerhalb eines Diskurses konstant zu halten. Der PSOLA-Algorithmus kann eine geringfügige, u.U. jedoch hörbare Verschlechterung der Signalqualität verursachen. Wären also die kompletten Diskurse aus dem Originalsignal des jeweils ersten Satzes und einer resythetisierten Version des zweiten Satzes zusammengesetzt worden (vgl. Arbeitschritt g), hätte dies evtl. zu einer wahrnehmbaren Veränderung der Signaleigenschaften zwischen den beiden Sätzen geführt. Um diesen Effekt zu vermeiden, wurden die ersten Sätze – mit den originalen F_0-Werten als PSOLA-Input – resynthetisiert. Daß dies keine gravierende Veränderung der Grundfrequenzkontur zur Folge hat, zeigt Abbildung 6-2 in einer Gegenüberstellung der F_0-Analyse des Originalsignals und der F_0-Analyse des resynthetisierten Signals.

Der entscheidende Schritt zur Erstellung der Versuchsitems mit unterschiedlichen Akzentuierungsmustern war der Arbeitschritt f. Von allen Sätzen mit einem Pronomen (n=5) wurden jeweils drei, von den Sätzen mit zwei Pronomina (n=9) jeweils sieben verschiedene Varianten erzeugt. Die Varianten unterschieden sich sowohl bezüglich der Pla-

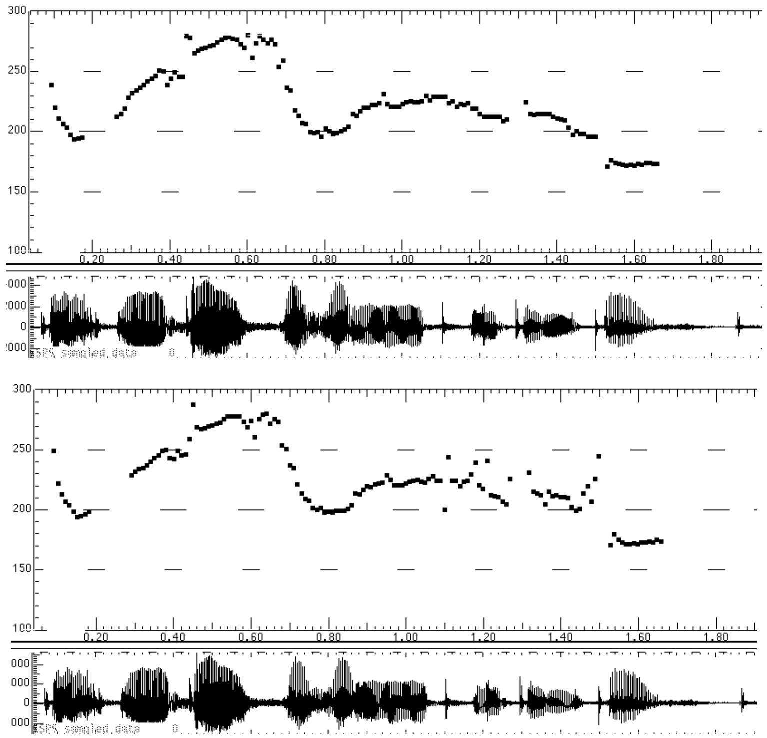


Abbildung 6-2: *F0-Berechnung aus dem Originalsignal (oben) und aus dem resynthetisierten Signal (unten). Äußerung:* Der Ober verhandelt mit dem Gast.

zierung des Pitchakzents als auch bezüglich des Akzenttyps. Für die Akzentplazierung kamen die folgenden Möglichkeiten infrage:

1. Akzent nur auf dem Element, das den neutralen Satzakzent trägt (neutrales Akzentmuster)

2. Akzent nur auf einem Pronomen

3. Akzent auf beiden Pronomina (in Sätzen mit zwei Pronomina)

4. Akzent auf einem Pronomen und dem Element, das den neutralen Satzakzent trägt

Das Element, das den neutralen Satzakzent trägt, wurde in Anlehnung an den Algorithmus von Cinque (1993) folgendermaßen lokalisiert: der semantische Kopf der syntaktisch am tiefsten eingebetteten Phrase, sofern es sich dabei um ein lexikalisches Element handelte. Handelte es sich dabei um eine pronominalisierte Form, so wurde der neutrale Satzakzent dem Verb zugewiesen. Tabelle 6-3 gibt eine Übersicht über die Plazierung des neutralen Satzakzents im zweiten Satz der neun zugrundeliegenden Diskurse.

Tabelle 6-3: Plazierung des neutralen Satzakzents.

Diskurs #	zweiter Satz mit neutralem Satzakzent
1	Sie ist KRANK.
2	Er hat kein KLEINGELD.
3	Sie zeigt ihr neue FOTOS.
4	Er hat ihm die VORFAHRT genommen.
5	Sie muß ihr die MASCHINE erklären.
6	Sie wird UNGEDULDIG.
7	Er zeigt ihm den VOGEL.
8	Er ÄRGERT sich über ihn.
9	Sie ist zu ihr nach HAUSE gekommen.

Bezüglich der Akzenttypen haben wir uns in diesem Experiment auf die zwei Standardakzente H*L und L*H beschränkt. Der wichtigste Grund hierfür war, die Anzahl der Versuchsitems in einem vernünftigen Rahmen zu halten. Schon die Kombination dieser zwei Akzenttypen führte zu einer Gesamtzahl von 173 Items (s.u.), was angesichts der daraus resultierenden Versuchsdauer von über 60 min. die Grenze des Zumutbaren darstellt. Andererseits wollten wir uns nicht ausschließlich auf den Peak-Akzent (H*L) beschränken, um zu überprüfen, ob sich verschiedene Akzenttypen bezüglich der Stärke ihres interpretativen Effekts unterscheiden (vgl. Cahn, 1995).

Aus der Kombination der beiden Akzenttypen und den erwähnten Plazierungsmöglichkeiten ergeben sich die in den Tabellen 6-4 und 6-5 jeweils an einem Beispiel dargestellten Akzentmuster. Das Akzentmuster 1 entspricht jeweils der neutralen Akzentuierung. Beispiele für die resultierende Grundfrequenzkontur der einzelnen Akzentmuster finden sich in Anhang B.

Tabelle 6-4: Akzentmuster bei Sätzen mit zwei Pronomina.

Akzentmuster #	*Sie*	*zeigt*	*ihr*	*neue*	*Fotos*	
1					H*L	L%
2	H*L					L%
3			H*L			L%
4	L*H		H*L			L%
5	H*L		H*L			L%
6	L*H				H*L	L%
7			L*H		H*L	L%

Tabelle 6-5: Akzentmuster bei Sätzen mit einem Pronomen.

Akzentmuster #	*Er*	*hat*	*kein*	*Kleingeld*	
1				H*L	L%
2	H*L				L%
6	L*H			H*L	L%

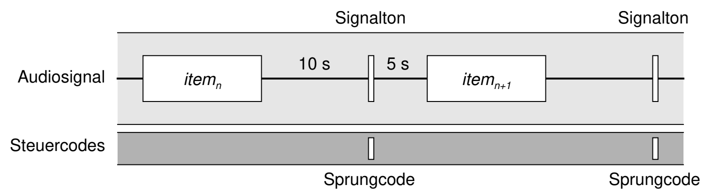


Abbildung 6-3: *Anordnung der Testitems auf dem DAT-Band.*

6.2.2.3 Testitems

Alle Akzentmuster wurden mit dem in Kapitel 4 vorgestellten System erzeugt. Die resynthetisierten Signale wurden dann in randomisierter Reihenfolge auf ein DAT-Band überspielt, mit einer Pause von 15 Sekunden zwischen jedem Item und einem Signalton 5 Sekunden vor jedem Item. Synchron zu jedem Signalton wurde ein Sprungcode erzeugt, so daß bei der Wiederholung eines Items die Präsentation stets mit dem Signalton startete. Diese Anordnung ist in Abbildung 6-3 dargestellt. Insgesamt ergaben sich 173 Items, davon 76 ambige und 97 eindeutige. Tabelle 6-6 gibt einen Überblick über die Zusammensetzung der Items.

Erläuterungen zu Tabelle 6-6: Ein Item aus der Klasse der eindeutigen Diskurse ging bei der Erstellung des DAT-Bandes verloren. Daher fanden in dem Experiment nur 173 statt 174 Items Verwendung. Dies ist jedoch für das vorliegende Experiment ohne Belang, da die Items aus der Klasse der eindeutigen Diskurse ohnehin nur als Füllmaterial dienten. Eine zweite Anmerkung betrifft den Diskurs #8. Im zweiten Satz dieses Diskurses (*Er ärgert sich über ihn*) ließen sich nur 5 der 7 Akzentmuster realisieren. Das Akzentmuster #6 konnte nicht realisiert werden, da dieses Muster aufgrund der direkten Nachbarschaft der akzentuierten Silben unnatürlich klang. Eine Verbesserung hätte nur erreicht werden können durch eine Manipulation der Segmentdauern, was jedoch durch die Vorgabe, ausschließlich die Grundfrequenzkonturen zu verändern, ausgeschlossen war. Die Reihenfolge der syntaktischen Konstituenten war der Grund für den Ausschluß

Tabelle 6-6: Die Zusammensetzung der Testitems.

Diskursklasse	Anzahl Diskurse	Anzahl Akzentmuster	Items gesamt
ambig, 1 Pronomen	5	3	15
ambig, 2 Pron.	8	7	56
ambig, 2 Pron. (Diskurs #8)	1	5	5
ambig gesamt	14		**76**
eindeutig, 1 Pron.	6	3	18
eind., 2 Pron.	10	7	70
eind., 2 Pron. (Diskurs #8)	2	5	10
eind. gesamt	18		**98**
	32		**174**

des Akzentmusters #7. In diesem Falle wäre L*H der nukleare Akzent vor dem Terminalton L% gewesen. Diese Abfolge ist im Deutschen nicht möglich (Féry, 1993). Im experimentellen Material lag Diskurs #8 einmal in ambiger (*Er ärgert sich über ihn*) und zweimal in eindeutiger Form vor (*Er ärgert sich über sie / Sie ärgert sich über ihn*) (vgl. Zeile 3 und Zeile 6 in Tabelle 6-6).

6.2.3 Versuchspersonen

Das Experiment wurde mit 10 Versuchspersonen durchgeführt, 5 weibliche und 5 männliche. Das Alter der Versuchspersonen lag zwischen 51 und 71 Jahre, der Altersdurchschnitt betrug 59,8, der Median 60 Jahre. Keine der Versuchspersonen verfügte über besondere sprachwissenschaftliche Kenntnisse.

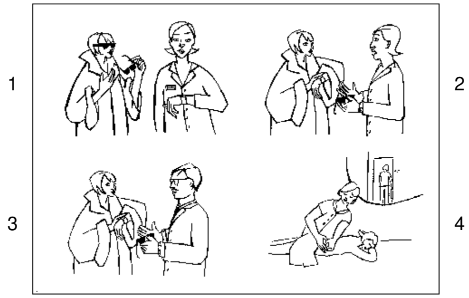


Abbildung 6-4: *Bildtafel mit 4 Situationsbildern zum Diskurs #6 (*Die Verkäuferin berät die Kundin. Sie wird ungeduldig*). Bild 1: parallele Interpretation (Sie := Verkäuferin), Bild 2: komplementäre Interpretation (Sie := Kundin), Bild 3 und 4: Ablenker (ähnlich und unähnlich).*

6.2.4 Durchführung

Die Testitems wurden auditiv über Kopfhörer dargeboten. Es wurde immer das selbe DAT-Band verwendet, so daß alle Versuchspersonen die Testitems in der gleichen (randomisierten) Reihenfolge hörten. Gleichzeitig mit dem auditiven Stimulus wurde den Versuchspersonen eine Bildtafel mit jeweils 4 Situationsbildern (Strichzeichnungen) vorgelegt (vgl. Abb. 6-4). Auf jeder Bildtafel war eine Situation abgebildet, die zu einer parallelen Interpretation des Items paßte (Subj. Pronomen := Subj. Referent bzw. Obj. Pron. := Obj. Ref.; vgl. Abb. 6-4, Situation 1) und eine Situation, die zu einer komplementären Interpretation paßte (Subj. Pron. := Obj. Ref. bzw. Obj. Pron. := Subj. Ref.; vgl Abb. 6-4, Situation 2). Daneben waren noch zwei Ablenker abgebildet, einer, der eine ähnliche Situation zeigte, und einer, der keine Ähnlichkeit mit der auditiv präsentierten Situation aufwies. Die Ablenker wurden in das Material aufgenommen, um das Experiment auch für die Anforderungen der klinischen Anwendung (s.o.) zu präparieren. Die Bildtafeln, die zu den eindeutigen Items präsentiert wurden, zeigten entsprechend nur eine passende Situation, daneben zwei ähnliche und einen unähnlichen Ablenker. Die

Anordnung der Situationsbilder auf den Bildtafeln wurde variiert, d.h. sowohl die passenden Situationen als auch die Ablenker konnten an jeder der vier möglichen Positionen auf der Bildtafel erscheinen.

Die Versuchspersonen wurden gebeten, auf diejenige Situation zu zeigen, die von dem auditiv dargebotenen Stimulus am besten beschrieben wird. Nach jedem Item hatten die Versuchspersonen mindestens 10 Sekunden Zeit (bis zum nächsten Signalton), sich zu entscheiden. Jedes Item konnte jedoch auf Wunsch auch beliebig oft wiederholt werden. Die Reaktion ›nicht entscheidbar‹ war zugelassen.

6.2.5 Ergebnisse und Diskussion

Insgesamt wurden 1730 Entscheidungen erfaßt (10 VPs x 173 Items), davon waren 970 eindeutige und 760 ambige Items. Die Reaktionen der Versuchspersonen wurden folgendermaßen klassifiziert:

0: nicht entscheidbar
1: parallele Interpretation
2: komplementäre Interpretation
3: ähnlicher Ablenker
4: unähnlicher Ablenker

Reaktion 4 war erwartungsgemäß in den Daten nicht vertreten. Reaktion 0 war nur einmal vertreten und zwar bei einem ambigen Item (Diskurs #4, Akzentmuster #3). Bei den eindeutigen Items wurde insgesamt 15 mal (1,3% der Fälle) auf einen ähnlichen Ablenker gezeigt. Diese geringe Anzahl an ›fehlerhaften‹ Reaktionen bzw. 0-Reaktionen spricht für die gute Verständlichkeit und relative Natürlichkeit der synthetisierten auditiven Stimuli ebenso wie für die ausreichende Interpretierbarkeit der Situationsbilder. Die Reaktionen 0, 3 und 4 werden bei der folgenden Auswertung nicht mehr berücksichtigt. Ebenfalls unberücksichtigt bleiben die eindeutigen Items, diskutiert wird im weiteren

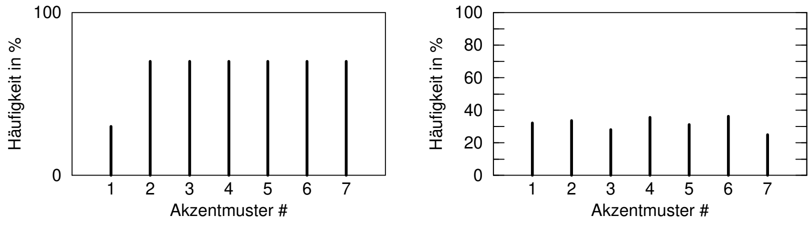


Abbildung 6-5: *Erwartete (links) und tatsächliche (rechts) relative Häufigkeit von Reaktion 2 (komplementäre Interpretation) in Abhängigkeit von Akzentmuster #1 (neutraler Satzakzent) bis #7.*

Verlauf ausschließlich die zur Beurteilung der experimentellen Hypothese relevante Verteilung der Reaktionen 1 und 2 bei ambigen Items.

Die in Abschnitt 6.2.1 formulierte Hypothese läßt erwarten, daß bei Items mit neutralem Satzakzent (Akzentmuster #1) sehr viel häufiger die parallele Interpretation (Reaktion 1) gewählt wird als die komplementäre Interpretation (Reaktion 2). Umgekehrt ist bei Items, die eines der Akzentmuster #2 bis #7 realisieren – bei denen also mindestens ein Pronomen akzentuiert ist –, zu erwarten, daß häufiger die komplementäre Interpretation, also Reaktion 2 gewählt wird (möglicherweise mit unterschiedlich starkem Effekt bei verschiedenen Akzentmustern). Abbildung 6-5 zeigt auf der linken Seite die zu erwartende Häufigkeit der Reaktion 2 für das Akzentmuster #1 (gering) und für die Akzentmuster #2 bis #7 (hoch). Die tatsächliche Häufigkeit der Reaktion 2 für die einzelnen Akzentmuster wird in Abbildung 6-5 auf der rechten Seite gezeigt. Es wird deutlich, daß die Erwartungen durch das Ergebnis des Experiments nicht bestätigt werden können. Der prozentuale Anteil von Reaktion 2 liegt bei allen Akzentmustern auf vergleichbarem Niveau, nämlich durchgehend bei etwa 30% (Mittelwert 31,7%). Der Anteil von Reaktion 1 liegt entsprechend bei allen Akzentmustern um 70%. Dies bedeutet, daß ein Einfluß der Akzentmuster auf die Interpretation der Diskursitems im Sinne der in Abschnitt 6.2.1 formulierten Hypothese mit dem von uns gewählten experimentellen

Setting nicht nachgewiesen werden konnte. Die Entscheidung für eine parallele oder komplementäre Interpretation scheint nicht von den prosodischen Eigenschaften der Diskursitems abhängig zu sein, die Akzentuierung oder De-Akzentuierung eines Pronomens scheint für dessen Interpretation keine Rolle zu spielen. Bevor ich in den verbleibenden Abschnitten dieses Kapitels einen Ansatz vorstelle, der das ›Scheitern‹ dieses Experimentes zu erklären versucht und die Intuition von der Relevanz prosodischer Eigenschaften für die Interpretation anaphorischer Beziehungen rehabilitiert, möchte ich zunächst einige Aspekte des Ergebnisses dieses Experiments detaillierter diskutieren.

Der wichtigste Aspekt betrifft die Inhomogenität der 14 ambigen Diskurse. Abbildung 6-6 macht deutlich, daß sich die Diskurse bezüglich der Ausprägung der abhängigen Variablen – Reaktion 1 vs. Reaktion 2 – z.T. erheblich voneinander unterscheiden. Abbildung 6-6 zeigt – wie Abbildung 6-5 – den prozentualen Anteil von Reaktion 2 an der Gesamtheit der Reaktionen, und zwar aufgeschlüsselt nach Akzentmustern und Diskursen. Jedes Diagramm gibt die relative Häufigkeit von Reaktion 2 für jeden der 14 ambigen Diskurse an. Die Diagramme unterscheiden sich dadurch, daß jeweils nur Items, die eines der sieben Akzentmuster realisiert haben, berücksichtigt werden. Da nicht alle Akzentmuster auf allen Diskursen realisiert wurden (vgl. Abschnitte 6.2.2.2 und 6.2.2.3), weisen die Diagramme für die Akzentmuster #3 bis #7 Lücken auf. Die gestrichelte Linie gibt jeweils die über alle Diskurse gemittelte relative Häufigkeit von Reaktion 2 für die einzelnen Akzentmuster an (vgl. Abbildung 6-5, rechts). Die Diskursnummern entsprechen der Numerierung in Tabelle 6-1. Wie oben erwähnt, wurden die Diskurse #1 bis #5 jeweils einmal mit der in Tabelle 6-1 angegebenen Reihenfolge der Referenten aufgenommen (1a, 2a, 3a, 4a, 5a) und jeweils einmal mit umgekehrter Reihenfolge (1b, 2b, 3b, 4b, 5b).

Deutlich zu sehen ist, daß bei den Diskursen #4 und #5 die Häufigkeit der Reaktion 2 bei allen Akzentmustern über dem Durchschnitt liegt. Bei Diskurs #4 ist eine wahrscheinliche Erklärung darin zu finden, daß die beiden Referenten im ersten Satz nicht an der Subjekt- bzw. Objektposition eingeführt werden, sondern beide gemeinsam inner-

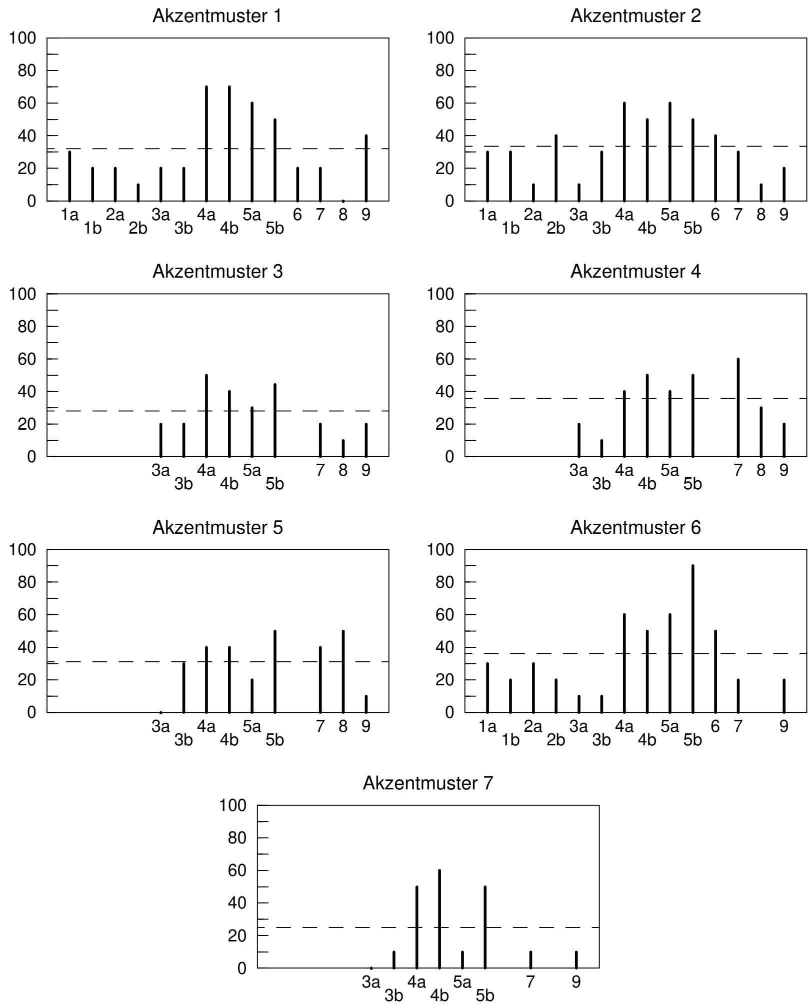


Abbildung 6-6: *Relative Häufigkeit von Reaktion 2 bei den Akzentmustern #1 bis #7 (jeweils Y-Achse in %). Abgetragen ist die relative Häufigkeit für alle 14 Diskurse (jeweils X-Achse), sowie der Mittelwert für jedes Akzentmuster (jeweils gestrichelte horizontale Linie).*

halb einer komplexen Subjekt-NP, d.h. die Ambiguität der beiden Pronomina im zweiten Satz entsteht nicht dadurch, daß ein Subjekt- und ein Objektreferent verfügbar ist, sondern dadurch, daß zwei Subjektreferenten verfügbar sind.

Diskurs #4

Der Motorradfahrer und der Radfahrer haben einen Unfall. Er hat ihm die Vorfahrt genommen.

Die Präferenz für einen der beiden Referenten ist in diesem Fall – unabhängig vom Akzentmuster – zufällig verteilt, der Anteil von Reaktion 1 und Reaktion 2 beträgt jeweils etwa 50% (Mittelwert für Reaktion 2: 52,1%). Diese Verteilung unterstützt die Bezugnahme der anfangs des Kapitels vorgestellten Präferenzprinzipien auf die grammatische Funktion der Referenten anstatt auf deren Reihenfolge. Da beide Referenten in Diskurs #4 die selbe grammatische Funktion erfüllen, greift keines der erwähnten Präferenzprinzipien und die Wahl eines Antezedenten für die Pronomina bleibt dem Zufall überlassen. Hätte die Reihenfolge der Referenten eine Relevanz bezüglich der Präferenz, wäre ein Überwiegen der Reaktion 1 zu erwarten gewesen.

Die überdurchschnittliche Häufigkeit von Reaktion 2 bei Diskurs #5 liegt wahrscheinlich darin begründet, daß dieser Diskurs nicht über die erwünschte Neutralität bezüglich des Einflusses von Weltwissen verfügt.

Diskurs #5

a. *Die Enkelin wäscht bei der Oma. Sie muß ihr die Maschine erklären.*

b. *Die Oma wäscht bei der Enkelin. Sie muß ihr die Maschine erklären.*

Die Interpretation dieses Diskurses kann in dem Sinne verschoben sein, daß die Objektreferentin des ersten Satzes – als Eigentümerin der Waschmaschine – eher in Betracht

kommt, die Maschine zu erklären, als die Subjektreferentin, die nur zu Gast ist. Diese Interpretation führt zu der Koreferenz zwischen Subjektpronomen und Objektreferent, also zu einer Präferenz für Reaktion 2 im Experiment. Da es jedoch nicht – außer mit Akzentmuster #6 – zu einem signifikant höheren Anteil von Reaktion 2 im Vergleich zu Reaktion 1 kommt (Mittelwert von Reaktion 2 bei Diskurs #5: 47,6%), scheint der Effekt des Weltwissens nicht stark genug zu sein, um die konkurrierende Präferenz – grammatische Parallelität, also Reaktion 1 – völlig auszuschalten. Dadurch kommt es, wie bei Diskurs #4, zu einer etwa gleichen Verteilung der beiden Reaktionen.

Zusammenfassend läßt sich die Inhomogenität der Diskurse folgendermaßen beschreiben: Bei den Diskursen #1 bis #3 überwiegt stets Reaktion 1, sowohl mit dem neutralen Akzentmuster #1 als auch mit den Akzentmustern #2 bis #7. Bei diesen Diskursen hat also die prosodische Variation keinerlei Effekt. Ebenfalls ohne Einfluß bleibt die prosodische Variation bei Diskurs #4, Reaktion 1 und Reaktion 2 sind zufällig verteilt unter allen Bedingungen. Bei Diskurs #5 liegt der Anteil von Reaktion 2 ebenfalls unter allen Bedingungen über dem Durchschnitt, im Mittelwert nahe der Zufallsverteilung von 50%, mit Akzentmuster #6 kommt es jedoch zu einem deutlich höheren Spitzenwert. Die Diskurse #6 bis #9 unterscheiden sich darin von den Diskursen #1 bis #3, daß bei einigen Akzentmustern ein Effekt der prosodischen Variation angedeutet ist. Da jedoch jeder Diskurs auf ein anderes Akzentmuster zu reagieren scheint, läßt sich aus dieser Beobachtung kein einheitlicher Effekt im Hinblick auf die experimentelle Hypothese ableiten.

Neben der erwähnten Heterogenität der Diskurse und den prosodischen Aspekten, die in den folgenden Abschnitten diskutiert werden, sollen zur Einschätzung des Experimentes und seiner Ergebnisse die folgenden kritischen Punkte, die während und nach der Durchführung des Experimentes aufgefallen sind, kurz genannt werden:

Materialumfang. Das Material war sehr umfangreich, vor allem aufgrund der zahlreichen Füllitems (eindeutige Diskurse). Eine Versuchsdurchführung dauerte dadurch mindestens 60 min. Während dieser Zeit wurden den Versuchspersonen immer die selben 9

Diskurse präsentiert, zwar mit geringen textuellen und/oder prosodischen Variationen, doch ein Gewöhnungseffekt ist unter diesen Umständen nicht auszuschließen.

Diskurskonstruktion. Die Diskurse waren alle sehr kurz und strukturell ähnlich. Möglicherweise wäre eine Mischung aus kurzen und längeren Diskursen sinnvoller, um auch hier einem Gewöhnungseffekt entgegenzuwirken.

6.3 Relative Prominenz

Mit dem oben beschriebenen Experiment konnte kein systematischer Einfluß prosodischer Markierung auf die Interpretation von Pronomina nachgewiesen werden. Wie in Abschnitt 6.1.2 gezeigt wurde, entspricht dieses Ergebnis jedoch nicht der linguistischen Intuition, auf die sich die einschlägigen Arbeiten zur Prosodie-Semantik-Schnittstelle seit vielen Jahren berufen. Wie kann diese Diskrepanz erklärt werden? In diesem Abschnitt soll ein Ansatz skizziert werden, der die aus dem Tonsequenzmodell abgeleitete experimentelle Prämisse, daß allein das *Vorhandensein* eines Pitchakzents die hinreichende Bedingung für prosodische Prominenz auf der Satzebene ist, kritisch hinterfragt. Daraus ergibt sich eine mögliche Erklärung für das Scheitern des oben beschriebenen Experimentes und es wird eine Perspektive aufgezeigt für die zukünftige experimentelle Untersuchung der hier behandelten Problematik.

Wie in verschiedenen Arbeiten gezeigt werden konnte, ist die Prominenz von Pitchakzenten graduell unterschiedlich (‘t Hart, 1981; Liberman & Pierrehumbert, 1984; Rietveld & Gussenhoven, 1985; Hermes & van Gestel, 1991). Durch die Variation des Umfangs der Grundfrequenzbewegung, die mit einem Pitchakzent assoziiert ist, können verschiedene Grade der Prominenz realisiert werden. Desweiteren konnte in den Arbeiten zum sog. *Gussenhoven-Rietveld-Effekt* gezeigt werden, daß der Umfang der Tonhöhenbewegung bzw. die Höhe des Peaks eines Akzents die wahrgenommene Prominenz eines nachfolgenden Akzents beeinflussen kann (Gussenhoven & Rietveld, 1988; Ladd,

Verhoeven & Jacobs, 1994). Und obwohl nach wie vor zahlreiche theoretische und empirische Aspekte der Interaktion zwischen benachbarten Pitchakzenten ungelöst sind, so darf doch aufgrund des aus der metrischen Phonologie bekannten Konzepts der *relativen Prominenz* erwartet werden, daß auch die Prominenz von Pitchakzenten relativ zu anderen Elementen innerhalb einer zu bestimmenden phonologischen Domäne kalkuliert wird. Ausgehend von der Beobachtung, daß die Prominenz von Pitchakzenten mit dem Umfang der assoziierten Tonhöhenbewegung korreliert, wäre also mit der Veränderung des Tonhöhenumfangs eines vorhandenen Pitchakzents zu rechnen, wenn ein weiterer Pitchakzent hinzugefügt wird, um die Prominenz eines Akzents relativ zum anderen zu justieren. In den folgenden beiden Abschnitten werden die Daten einer kleinen Produktionsstudie vorgestellt und diskutiert, um die Relevanz des Konzepts der relativen Prominenz für das von uns verwendete experimentelle Material zu überprüfen.

6.3.1 Daten

Methode — Die hier vorgestellten Daten wurden folgendermaßen erfaßt: Eine Sprecherin (29 Jahre, Standardaussprache) wurde gebeten, die Diskurse #1, #5, #6 und #8 vorzulesen. Der Sprecherin wurden zu jedem Diskurs zwei Situationsbilder vorgelegt, jeweils eines, das eine zur parallelen Interpretation der Pronomina passende Situation darstellte (S1), und eines, das eine zur komplementären Interpretation passende Situation darstellte (S2) (vgl. Bild 1 und 2 in Abb. 6-4). Die Sprecherin wurde angewiesen, die Diskurse jeweils zweimal zu produzieren, einmal passend zu S1 und einmal passend zu S2. Da beide Bilder gleichzeitig mit dem vorzulesenden Diskurs präsentiert wurden und die beiden kontrastierenden Produktionen unmittelbar hintereinander erfolgen sollten, war die Sprecherin direkt mit dem zu realisierenden Kontrast konfrontiert. Die Sprecherin erhielt jedoch keine direkten Vorgaben, wie dieser Kontrast zu realisieren sei. Die Äußerungen wurden auf DAT aufgenommen (48 kHz) und zur weiteren Analyse mit ESPS/*waves* auf eine SGI Indy Workstation überspielt (16 kHz, 16 bit).

Ergebnisse — Bei den Diskursen #1, #5 und #6 waren die Pronomina bei der Beschreibung von S2 eindeutig prosodisch markiert und unterschieden sich deutlich von ihrem Gegenstück bei der Beschreibung von S1. Bei Diskurs #8 (Satz 2: *Er ärgert sich über ihn.*) war bei der Beschreibung von S2 nur auf dem zweiten Pronomen (*ihn*) eine Akzentuierung hör- und erkennbar. Bei allen Diskursen war bei der Beschreibung von S1 der neutrale Satzakzent an der erwarteten Position (#1: *krank*, #5: *Maschine*, #6: *ungeduldig*, #8: *ärgert*) deutlich hörbar und im Grundfrequenzverlauf erkennbar. An der selben Position befand sich auch jeweils bei der Beschreibung von S2 ein Akzent, zusätzlich zu den akzentuierten Pronomina. Die Produktionen zu S2 wiesen also jeweils mindestens 2 Pitchakzente auf: einen bzw. zwei assoziiert mit dem Pronomen/den Pronomina und einen an der ›neutralen Position‹.

Zur Untersuchung des Effekts der relativen Prominenz wurde dann der Umfang der Grundfrequenzbewegung auf den Pronomina und auf der Silbe, die den neutralen Satzakzent trägt, ermittelt und die Ergebnisse der jeweils zwei Produktionen (Beschreibung von S1 und Beschreibung von S2) einander gegenübergestellt. Der Umfang der Grundfrequenzbewegung ergab sich aus der Differenz zwischen dem höchsten und tiefsten F_0-Wert innerhalb der Zielsilbe. Abbildung 6-7 zeigt die mittlere F_0-Differenz von 3 Diskurspaaren (Diskurs #8 wurde nicht berücksichtigt, da die prosodische Markierung hier nicht eindeutig war). Auf den unbetonten Pronomina war der Grundfrequenzverlauf erwartungsgemäß flach, die mittlere F_0-Differenz betrug 10,4 Hz. Dagegen war bei den akzentuierten Pronomina eine ausgeprägte Bewegung der Grundfrequenzkontur zu beobachten, die mittlere Differenz zwischen höchstem und tiefstem F_0-Wert betrug 36,8 Hz. Der für die Frage nach dem Effekt der relativen Prominenz interessantere Vergleich ist der zwischen den beiden Differenzwerten an der ›neutralen Position‹. In der S1-Bedingung ist der Akzent an der neutralen Position der einzige Akzent in der Intonationsphrase, der Umfang der Grundfrequenzbewegung, die mit diesem Akzent assoziiert ist, ist mit durchschnittlich 82,3 Hz sehr hoch. In der S2-Bedingung, mit zusätzlichen Akzenten auf dem Pronomen/den Pronomina, ist der Umfang der Grundfrequenzbewe-

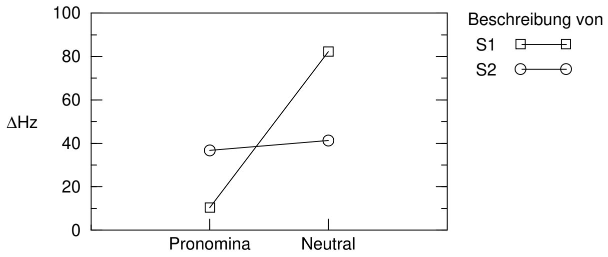


Abbildung 6-7: *Umfang der Grundfrequenzbewegung auf den Pronomina und an der ›neutralen Position‹ in Hz (Mittelwerte von jeweils 3 Produktionen). Die Pronomina sind in einer Bedingung nicht akzentuiert (Quadrat) und in einer Bedingung akzentuiert (Kreis), die ›neutrale Position‹ ist in beiden Bedingungen akzentuiert.*

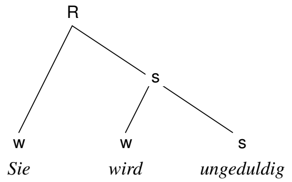
gung an der ›neutralen Position‹ deutlich reduziert, er beträgt nur noch durchschnittlich 41,3 Hz. In den Multiakzentphrasen (S2 Bedingung) wird also die Auslenkung der Pitchakzente auf einem mittleren Niveau justiert, während der Pitchakzent in Einzelakzentphrasen mit einer sehr großen Auslenkung realisiert wird.

6.3.2 Diskussion

Zunächst fällt auf, daß durchgängig auch in der S2-Bedingung ein Pitchakzent an der ›neutralen Position‹ realisiert wurde. Dies kann damit erklärt werden, daß der gesamte zweite Satz in beiden Fällen eine neue Information einführt und somit einen Akzent auf demjenigen Element zuläßt, welches die maximale Fokusprojektion ermöglicht. Der Unterschied zwischen den beiden Bedingungen ist lediglich der, daß in der S2-Bedingung *zusätzliche,* kontrastive Akzente realisiert werden, um die Default-Interpretation der Pronomina zugunsten der komplementären Interpretation zu unterdrücken.

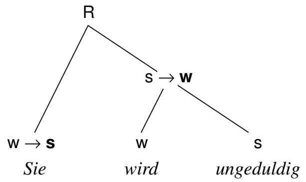
Außerdem zeigt der Vergleich der Pitchakzente an der neutralen Position, also an der Position, die bei neutraler Satzbetonung den Akzent trägt, daß der Umfang der akzentbedingten Grundfrequenzbewegung abhängig ist von den Prominenzverhältnissen innerhalb der Intonationsphrase. In der metrischen Baumstruktur (vgl. Liberman & Prince, 1977) von *Sie wird ungeduldig* in (23) ist *ungeduldig* dasjenige Element, das den Pfad mit *s*-Knoten (*s* = *strong*) terminiert. Da die restlichen Elemente metrisch schwach sind (*w* = *weak*), verfügt *ungeduldig*, bzw. die betonte Silbe *un*, über größere Prominenz als alle anderen Silben innerhalb der Phrase und ist als einzige mit einem Pitchakzent assoziiert.[9]

(23)



Die metrische Analyse des selben Satzes mit einem kontrastiven Akzent auf *Sie* (S2-Bedingung) in (24) zeigt die veränderten Prominenzverhältnisse innerhalb der Phrase. Die Akzentuierung von *Sie* verlangt das Pronomen an einer metrisch starken Position. Dies hat zur Folge, daß gemäß den metrischen Prinzipien der Schwesterknoten zu einem *w*-Knoten wird (*contrastive relabelling*, vgl. Dogil, 1979: 64ff).

(24)

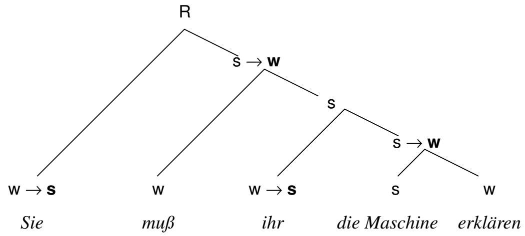


Damit wird der s-Knoten von ungeduldig von einem w-Knoten dominiert, was einen Verlust an Prominenz bedeutet (vgl. Liberman & Prince, 1977), wobei die Realisierung

[9] Für die metrische Analyse der Prominenz auf der Ebene phonologischer Wörter (also die Bestimmung der betonten Silbe eines Wortes) verweise ich auf die Arbeiten von Halle & Vergnaud, 1987 und Halle & Idsardi, 1995.

eines Pitchakzents an dieser Position nach wie vor zulässig ist. Diese Analyse liefert eine Erklärung für die Reduktion des Tonhöhenumfangs, wie sie bei den Produktionsdaten zu beobachten war. Das metrisch starke *Sie* erhält einen Pitchakzent und das ebenfalls metrisch starke *ungeduldig* behält einen Pitchakzent, jedoch mit geringerer Ausprägung aufgrund der geringeren Prominenz. Die analoge Anwendung der *contrastive relabelling*-Prinzipien auf Sätze mit zwei betonten Pronomina in (25) führt zu einem vergleichbaren Ergebnis.

(25)



Der *s*-Pfad, der bei neutraler Betonung zu *Maschine* führt, wird infolge der metrischen Hervorhebung der Pronomina durch *w*-Knoten unterbrochen, die relative Prominenz von *Maschine* wird reduziert.

6.4 Abschließende Diskussion

Die Produktionsstudie hat gezeigt, daß bei der Realisierung von prosodischen Minimalpaaren, wie sie in dem Perzeptionsexperiment verwendet wurden, die Beachtung der Prominenzrelationen innerhalb der Intonationsphrase von Bedeutung ist. Außerdem konnte gezeigt werden, daß auch im Deutschen das phonologische Konzept der graduellen Prominenz auf Satz- bzw. Phrasenebene mit der Variation des Tonhöhenumfangs von Pitchakzenten zu korrelieren scheint. Bei der Erstellung des experimentellen Materials wurden Modifikationen, die sich aus dem Kalkül der relativen Prominenz ergeben, nicht berücksichtigt, da die Berechnung von Prominenzrelationen in dem von uns verwendeten Resyntheseprozeß – automatische Konturgenerierung und PSOLA-Resythese – noch nicht implementiert ist. Dieser Mangel kann als entscheidender Faktor zur Erklärung des kontraintuitiven Ergebnisses des Perzeptionsexperimentes gewertet werden. Der Tonhöhenumfang der automatisch erzeugten Pitchakzente ergibt sich aus der unabhängig und für alle Items identisch definierten Top- und Baseline (s. Kapitel 4). Dies bedeutet, daß ein Pitchakzent, der mit einer bestimmten Silbe innerhalb einer gegebenen Phrase assoziiert ist, stets mit dem selben Tonhöhenumfang realisiert wird – vorausgesetzt die Top- und Baseline-Definition bleibt unverändert –, unabhängig davon, über welche Prominenz diese Silbe relativ zu den anderen Silben innerhalb der Phrase verfügt. Bezüglich des experimentellen Materials muß also davon ausgegangen werden, daß die oben angesprochene Verschiebung der Prominenzverhältnisse zugunsten der kontrastiv betonten Pronomina und zulasten des ›neutralen‹ Akzents nicht adäquat repräsentiert ist. Der Akzent an der ›neutralen Position‹, der in den Akzentmustern #6 und #7 realisiert wurde, unterscheidet sich nicht vom neutralen Akzent in Akzentmuster #1. Es ist naheliegend, daß eine in diesem Sinne unnatürliche Realisierung von Pitchakzenten und von Relationen zwischen Pitchakzenten zu einer von der Erwartung abweichenden Interpretation der prosodischen Eigenschaften einer Äußerung und letztendlich der Äußerung selbst führen kann. Wenn dies zutrifft, so läßt sich das Scheitern des Perzeptionsexperimentes im wesentlichen auf die unzureichende Modellierung der prosodischen Komponente zurückführen.

Die Modifikation der zugrundeliegenden prosodischen Komponente und die adäquate Realisierung von Effekten der relativen Prominenz in einem Perzeptionsexperiment ist jedoch kein triviales Unterfangen. Wie die Arbeiten zum Gussenhoven-Rietveld-Effekt zeigen, führen Experimente zur Wahrnehmung relativer Prominenz teils zu sehr unterschiedlichen Ergebnissen. Gussenhoven & Rietveld (1988) kommen zu dem überraschenden Ergebnis, daß in einer Intonationsphrase mit zwei Pitchakzenten der zweite Akzent als *weniger* prominent wahrgenommen wird, wenn – unter Beibehaltung des

Tonhöhenumfangs des zweiten Akzents – der Tonhöhenumfang des ersten Akzents *reduziert* wird. Ladd *et al.* (1994) zeigen dagegen, daß dieser Effekt nur dann auftritt, wenn der zweite Pitchakzent mit ›normalem‹, d.h. nicht-emphatischem Tonhöhenumfang realisiert wird. Bei größerem, emphatischem Tonhöhenumfang kommt es zu einer Umkehrung des Effekts, der zweite Akzent wird in diesem Fall mit *größerer* Prominenz wahrgenommen, wenn der Tonhöhenumfang des ersten Akzents reduziert wird. Außerdem konnte in der Studie von Ladd *et al.* (1994) auch ein Unterschied zwischen naiven und trainierten Hörern festgestellt werden. Während bei naiven Hörern wie oben beschrieben der Gussenhoven-Rietveld-Effekt bei nicht-emphatischem *pitch range* repliziert werden konnte und bei emphatischem *pitch range* die Umkehrung des Effekts zu beobachten war, konnte bei trainierten Hörern keiner der beiden Effekte nachgewiesen werden.

Ladd und Mitarbeiter schlagen die folgende Interpretation dieser Beobachtungen vor: Bei normalem, nicht-emphatischem *pitch range* wird die Prominenz eines Pitchakzentes nicht individuell erkannt. Stattdessen ›berechnet‹ der Hörer aus dem *pitch range* der gesamten Äußerung den Grad der Emphase und beurteilt die Prominenz der einzelnen Akzente relativ dazu. Dies hat zur Folge, daß die Prominenz eines nicht-emphatischen zweiten Akzents als geringer eingeschätzt wird, wenn der erste Akzent einen sehr kleinen Tonhöhenumfang aufweist, da in diesem Fall der *range* der gesamten Äußerung als gering wahrgenommen wird. Wenn dagegen beide Akzente mit größerem aber nach wie vor nicht-emphatischem Tonhöhenumfang realisiert werden, wird die gesamte Äußerung als emphatischer interpretiert und der zweite Akzent folglich mit größerer Prominenz wahrgenommen. Bei emphatischen Akzenten, die in der Studie eine Umkehrung des Gussenhoven-Rietveld-Effekts ausgelöst haben, wird, so Ladd und Mitarbeiter, die Prominenz von Pitchakzenten nicht relativ zur gesamten Äußerung beurteilt, sondern individuell:

> „When emphatic pitch range is used […] we speculate that this is a paralinguistic signal to override normal phonologically specified prominence rela-

> tions, and to interpret pitch range on the basis of 'every accent for itself'. Specifically, if P_2 [zweiter Pitchakzent] is emphatic, then a low P_1 [erster Pitchakzent] is interpreted as having been reduced to set the scene for the expanded pitch range on P_2. Therefore lowering P_1 *increases* the perceived prominence of the emphatic P_2. A higher P_1, on the other hand, is interpreted as not having been so reduced, and consequently it downplays the emphasis on P_2 and causes its perceived prominence to be lower." (Ladd *et al.*, 1994: 98; Hervorhebung im Original)

Ladd und Mitarbeiter betonen, daß es sich bei dieser Interpretation ihrer Daten um eine Spekulation handelt, die (noch) nicht experimentell überprüft ist (Ladd *et al.*, 1994: 98). Nichtsdestotrotz führt jedoch die Anwendung dieser Analyse auf unser Perzeptionsexperiment zu einem interessanten Ergebnis. Die automatisch erzeugten Pitchakzente in Multiakzentphrasen sind alle nicht-emphatisch und werden alle mit etwa dem selben Tonhöhenumfang realisiert, d.h. der erste Akzent, der stets mit einem Pronomen assoziiert ist, ist relativ hoch. Dies führt laut obiger Analyse dazu, daß der zweite Akzent mit größerer Prominenz wahrgenommen wird. Der zweite Akzent fällt bei allen Multiakzent-Items entweder auf das zweite Pronomen oder auf die ›neutrale Position‹, d.h. im einen Fall wird das zweite Pronomen als prominenter wahrgenommen als das erste und im anderen Fall übertrifft das ›neutrale Element‹ (Objekt bzw. Verb) das betonte Pronomen an Prominenz. Die folgende – ebenfalls spekulative – Interpretation bietet sich an: Im ersten Fall ist die prosodische Markierung der Pronomina aufgrund ihrer unterschiedlichen Prominenz nicht interpretierbar, die Hörer wählen die *Default*-Lesart. Im zweiten Fall schwächt die größere Prominenz des zweiten Akzents die ›Wirkung‹ des ersten Akzents, also die prosodische Markierung des Pronomens, was ebenfalls zur *Default*-Lesart führt. Diese Interpretation erhält Unterstützung aus den Ergebnissen der Produktionsstudie. Bei der natürlichen Produktion des hier untersuchten prosodischen Kontrasts führt die generelle Reduktion des Tonhöhenumfangs der Akzente in den Multiakzentphrasen (S2-

Bedingung) zur Wahrnehmung einer insgesamt geringeren Emphase und dadurch zu einer *geringeren* Prominenz des Akzents an der neutralen Position.

Die Diskussion hat gezeigt, daß das Konzept der relativen Prominenz für die Untersuchung der in diesem Kapitel behandelten und anderer, vergleichbarer Phänomene zu berücksichtigen ist und wichtige Einsichten in das Verständnis der Interpretation prosodischer Merkmale erlaubt. Gleichzeitig sollte jedoch auch deutlich geworden sein, daß zahlreiche Produktions- und Perzeptionsaspekte dieses Konzepts noch nicht annähernd vollständig verstanden werden und der zukünftigen Erforschung überlassen bleiben. Unser Fazit ist daher, daß eine Wiederholung des Experiments zur Anaphernresolution unter adäquater Berücksichtigung von prosodischen Prominenzrelationen und deren Perzeption und Interpretation beim derzeitigen Kenntnisstand in dieser Form kaum möglich ist.

7 Globale prosodische Parameter und Diskursstruktur

Längere Diskurse sind strukturiert, d.h. sie können als eine Aneinanderreihung einzelner, kleiner Segmente beschrieben werden, die durch spezifische Relationen verknüpft und zu einem kohärenten Ganzen zusammengefügt sind. Verschiedene Arbeiten der letzten Jahre konnten zeigen, daß ein Zusammenhang zwischen dieser Diskursstrukturierung und bestimmten prosodischen Eigenschaften der lautsprachlichen Realisierung von Diskursen besteht (z.B. Hirschberg & Pierrehumbert, 1986; Grosz & Hirschberg, 1992; Passenout & Litman, 1993; Sluijter & Terken, 1993; Swerts & Geluykens, 1993; Ayers, 1994; Selting, 1995; Couper-Kuhlen, 1996; Hirschberg, 1996). Einige Aspekte dieses Zusammenhangs sind das Thema dieses letzten experimentellen Kapitels. Bevor ich jedoch die Daten präsentiere und diskutiere, möchte ich zunächst die Grundzüge eines Modells der Diskursstrukturierung vorstellen, welches eine Erweiterung der in den vorangegangenen Kapiteln schon einige Male erwähnten Diskursrepräsentationstheorie darstellt. Dieses Modell bietet sehr mächtige Formalismen zur Kalkulation von Diskursstrukturen und soll die diskurssemantische Basis für die in Abschnitt 7.3 präsentierte phonologisch-phonetische Analyse bilden.

7.1 Segmentierte Diskursrepräsentationsstrukturen (SDRSen)

Eine Besonderheit der Theorie der Diskursstruktur von Asher (1993: Kap. 7) ist die enge Anknüpfung von diskursstrukturellen Prinzipien an ein Modell der formalen Semantik. Asher entwickelt seine Theorie im Sinne einer Erweiterung der Diskursrepräsentationstheorie (DRT) von Hans Kamp (Kamp, 1981; Kamp & Reyle, 1993). Die DRT beschreibt die semantische Interpretation von Diskursen als einen dynamischen Prozeß der Konstruktion von Diskursrepräsentationsstrukturen (DRSen). Basierend auf der syntaktischen und logischen Form, wird für jeden Satz eines Diskurses eine DRS aufgebaut. Eine DRS repräsentiert die semantischen Objekte, die durch einen Satz eingeführt werden, die semantischen Beschränkungen, denen diese Objekte unterliegen, und die semantischen Relationen, die zwischen diesen Objekten bestehen. Da die Prinzipien der Konstruktion von DRSen und deren interne Struktur nichts wesentliches zur Analyse und Repräsentation von globalen Diskursstrukturen im Sinne der Diskursgliederung beitragen, werde ich auf diese Aspekte der DRT hier nicht weiter eingehen und DRSen im weiteren Verlauf als ‚*black boxes*‘ behandeln[10]. Von zentralem Interesse ist dagegen der dynamische Aspekt der DRT, also die Möglichkeit, semantische Repräsentationen von Diskursen zu erweitern und zu aktualisieren, wenn im Verlauf eines Diskurses neue Informationen mitgeteilt werden. Dieser Aspekt ist in der klassischen DRT nur wenig ausgearbeitet und besteht im wesentlichen aus dem simplen Mechanismus, neue semantische Objekte in das bereits bestehende Geflecht semantischer Relationen zu integrieren. Das Ergebnis ist eine stetig wachsende DRS, die die globale Struktur des Diskurses – die Identifikation von Diskurssegmenten und ihrer gegenseitigen Beziehungen – kaum mehr widerspiegelt. Dieser Mechanismus wird durch das Diskursstrukturmodell von Asher ersetzt. Dieses Modell führt eine zweite Ebene der Diskursinterpretation ein, welche der zugrundeliegenden Ebene – den semantischen Regeln der DRS-Konstruktion – eine zusätzliche, übergeordnete Struktur zuweist (Asher, 1993: 257).

Die Basiseinheiten dieser übergeordneten Ebene der Diskursinterpretation sind segmentierte DRSen (SDRSen). Damit entscheidet Asher die Frage nach dem relevanten Level für die Identifikation von Diskurssegmenten – Textlevel, syntaktische Struktur oder semantische Struktur – zugunsten der semantischen Struktur. Ein Argument (von insgesamt drei, s. Asher, 1993: 261) für diesen Standpunkt ist die Analyse der anaphori-

[10] Eine gute Einführung in die DRT bietet das zweite Kapitel von Asher (1993), eine Gesamtdarstellung bieten Kamp & Reyle (1993).

schen Referenz auf abstrakte (semantische) Objekte im Diskurs. Zwei Beispiele von Asher (1993: 259):

(1) After thirty-eight months, America is back in space. The shuttle Discovery roared off the pad from Cape Kennedy at 10:38 this morning. The craft and crew performed flawlessly. Later in the day the TDRS shuttle communications satellite was successfully deployed. **This** has given a much needed boost to NASA morale.

(2) The shuttle Discovery roared off the pad from Cape Kennedy at 10:38 this morning. Later in the day the TDRS shuttle communications satellite was successfully deployed. The craft and crew performed flawlessly. After thirty-eight months, America is back in space. **This** has given a much needed boost to NASA morale.

Für (1) kann in etwa die folgende Struktur angenommen werden: Der erste Satz führt das Diskurstopik ein und die nächsten drei Sätze elaborieren dieses Topik. Für das anaphorische Pronomen *this* im fünften Satz (im Beispiel fettgedruckt) stehen damit drei mögliche Antezedenten zur Verfügung: 1.) die komplette Sequenz der ersten vier Sätze, 2.) nur die Elaborationssequenz (Sätze 2-4) oder 3.) nur der letzte Satz der Elaborationssequenz, das erfolgreiche Freisetzen des Satelliten. In (2) dagegen werden die Details vor dem generellen Topik mitgeteilt, das Pronomen kann sich in diesem Fall nur auf den vierten Satz und die Inhalte, die dieser zusammenfaßt, beziehen.

Die Beispiele machen deutlich, daß für eine angemessene Analyse anaphorischer Referenz abstrakte semantische Objekte (z.B. Propositionen) als mögliche Antezedenten in der Diskursrepräsentation definiert sein müssen; dies ist durch die Verfügbarkeit von DRSen bzw. SDRSen als Antezedenten gewährleistet. Außerdem machen diese Beispiele deutlich, daß die Zugänglichkeit von Referenten im Diskurs von deren Organisation

innerhalb der Diskursstruktur abhängig ist: In (1) sind drei SDRSen[11] als Antezedenten verfügbar, in (2) – aufgrund der veränderten Diskursstruktur! – nur eine.

DRSen sind also geeignete Einheiten, aus denen Diskurssegmente aufgebaut sind. Eine Menge von (einer oder mehreren) DRSen bildet ein Diskurssegment – eine SDRS. Außerdem besteht eine SDRS aus einem Satz von Bedingungen, die festlegen, welche Diskursrelationen zwischen den Konstituenten der SDRS bestehen (Asher, 1993: 262f). Verschiedene Arbeiten zur Theorie der Diskursstruktur haben zahlreiche solcher Diskursrelationen identifiziert, z.B. ELABORATION, CONTINUATION, COMMENT, CONTRAST, CAUSE etc. (einen Überblick gibt Asher, 1993: 264ff), ohne jedoch eine klare Semantik dieser Relationen zu formulieren. Ashers Modell übernimmt einen großen Teil dieser Relationen, ausgestattet mit einer expliziten Semantik (Asher, 1993: Appendix zu Kap. 7). Als Beispiel mag hier die Definition von ELABORATION in (3) dienen, zu deren Verständnis jedoch einige Vorbemerkungen nötig sind: Das Symbol $\Downarrow$ bezeichnet Diskurs-Dominanz (d-Dominanz) und trägt in (3) der Tatsache Rechnung, daß ELABORATION nicht eine einfache, zweistellige Relation darstellt, sondern eine asymmetrische Relation zwischen einer Konstituente, die ein Topik einführt, und einer anderen Konstituente, die dieses Topik elaboriert. Die Konstituente mit der generelleren Semantik d-dominiert die Konstituente mit der spezifischeren Semantik. Ein Ausdruck ME(α) bezeichnet die Menge der *main events* (‚Ereignisse' im Sinne der DRT), die in α eingeführt werden. Die formalen Definitionen von d-Dominanz, ME(α) und ‚*is more complex than*' finden sich in Asher (1993: Appendix zu Kap. 7).

(3) ELABORATION(α, β) iff ($\alpha \Downarrow \beta \vee$ (for every $e \in$ ME(β) there is an $e' \in$ ME(α) such that e is a part of e') & β is more complex than α.
(Asher, 1993: 300)

---

[11] SDRSen fassen in der Regel mehrere DRSen zusammen; so bildet z.B. die Elaborationssequenz in (1) eine SDRS, die aus drei DRS-Konstituenten besteht (*default*: jeder Satz konstituiert eine DRS). Nach der Definition von Asher (1993: 262f) kann eine SDRS jedoch auch aus nur einer DRS-Konstituente bestehen, insofern ist es zulässig, hier generell von SDRSen als Antezedenten zu sprechen.

Die komplementäre Relation zu ELABORATION ist CONTINUATION. CONTINUATION involviert niemals d-Dominanz, es ist eine symmetrische Relation zwischen zwei Konstituenten. Außerdem muß für alle Konstituenten, die durch CONTINUATION verknüpft sind, die selbe Relation zu allem, was diese Konstituenten d-dominiert, gelten (Asher, 1993: 267). Konstituenten, die durch CONTINUATION verknüpft sind, teilen das selbe Topik.

Ein weiterer wichtiger Aspekt der Definition von SDRSen ist die Rekursivität (Asher, 1993: 262f). Es ist zugelassen, daß die Konstituenten von SDRSen selbst SDRSen sind. Dies wurde bereits kurz anhand von Beispiel (1) erwähnt und spiegelt sich auch in der aus der DRT bekannten Box-Notation wider, die auch zur Darstellung von SDRSen herangezogen werden kann[12]:

(4)

k1
⇓

κ :=

k2, k3, k4

CONTINUATION(k2, k3)
CONTINUATION(k3, k4)

ELABORATION(k1, κ)

Ich werde diese kurze Einführung in Ashers Theorie der Diskursstruktur hier beenden, nachdem die wichtigsten Prinzipien erwähnt wurden. Ich bin mir bewußt, daß dieser kurze Überblick der Komplexität von Ashers Theorie kaum gerecht wird, für unsere Zwecke ist dieses Fragment jedoch ausreichend. Für eine ausführliche Diskussion der Theorie verweise ich auf den Originaltext (Asher, 1993: Kap. 7) und möchte nun die Anwendung der Theorie an einem Beispiel aufzeigen.

[12] k1, k2,… sind SDRS-Konstituenten und entsprechen den DRSen zu Satz 1, Satz 2,…; κ ist eine komplexe SDRS.

7.2 Ein Beispiel: *›Das dicke Kind‹*

Die Eingangssequenz der Erzählung *›Das dicke Kind‹* von Marie Luise Kaschnitz erlangte unter DR-Theoretikern einige Berühmtheit: Es ist einer der wenigen ‚echten', d.h. nicht von Linguisten konstruierten Texte, für die eine sehr detaillierte DRT-Analyse verfügbar ist (Rossdeutscher, 1994; Kamp & Rossdeutscher, 1994a,b). Hier der Wortlaut (aus Kaschnitz, 1994):

(5) Es war Ende Januar, bald nach den Weihnachtsferien, als das dicke Kind zu mir kam. Ich hatte in diesem Winter angefangen, an die Kinder aus der Nachbarschaft Bücher auszuleihen, die sie an einem bestimmten Wochentag holen und zurückbringen sollten. Natürlich kannte ich die meisten dieser Kinder, aber es kamen auch manchmal Fremde, die nicht in unserer Straße wohnten. Und wenn auch die Mehrzahl von ihnen gerade nur so lange Zeit blieb, wie der Umtausch in Anspruch nahm, so gab es doch einige, die sich hinsetzten und gleich auf der Stelle zu lesen begannen. Dann saß ich an meinem Schreibtisch und arbeitete, und die Kinder saßen an dem kleinen Tisch bei der Bücherwand, und ihre Gegenwart war mir angenehm und störte mich nicht.
Das dicke Kind kam an einem Freitag oder Samstag, jedenfalls nicht an dem zum Ausleihen bestimmten Tag. Ich hatte vor auszugehen …

Eine der DRT-Analyse vergleichbare, detaillierte Analyse der Diskursstruktur dieses Textes liegt leider nicht vor. Da der Text jedoch eine relativ klare Diskursstruktur aufweist und sich aufgrund der geleisteten Vorarbeiten und seiner ‚dichten' Semantik als Paradigma für verschiedenste Untersuchungen zur Prosodie-Semantik-Schnittstelle anbietet, ist er auch für die vorliegende Studie als Material ausgewählt worden. Ich habe versucht, die im folgenden vorgeschlagene Diskursstruktur entlang der von Asher entwickelten Theorie auszuarbeiten, unterstützt durch einige sehr wertvolle Hinweise von Antje Rossdeutscher. Vor der formalen Beschreibung gebe ich in (6) zunächst einen Überblick über die Diskursstruktur, den Inhalt der einzelnen SDRS-Konstituenten und die Einbettungsebenen.

(6)
k1(*Es war Ende Januar, bald nach den Weihnachtsferien, als das dicke Kind zu mir kam*)
  k2(*Ich hatte in diesem Winter angefangen, an die Kinder aus der Nachbarschaft Bücher auszuleihen, die sie an einem bestimmten Wochentag holen und zurückbringen sollten*)
    k3(
    k3a(*Natürlich kannte ich die meisten dieser Kinder*)
    k3b(*aber es kamen auch manchmal Fremde, die nicht in unserer Straße wohnten*))
    k4(
    k4a(*Und wenn auch die Mehrzahl von ihnen gerade nur so lange Zeit blieb, wie der Umtausch in Anspruch nahm*)
    k4b(*so gab es doch einige, die sich hinsetzten und gleich auf der Stelle zu lesen begannen*))
      k5(*Dann saß ich an meinem Schreibtisch und arbeitete, und die Kinder saßen an dem kleinen Tisch bei der Bücherwand, und ihre Gegenwart war mir angenehm und störte mich nicht*)
k6(*Das dicke Kind kam an einem Freitag oder Samstag, jedenfalls nicht an dem zum Ausleihen bestimmten Tag*)

Das Topik dieses Diskurses ist offensichtlich das dicke Kind, nicht zuletzt signalisiert durch den Titel der Erzählung. Das Topik wird im ersten Satz (*Es war Ende Januar…zu mir kam.*) als *main event* eingeführt und im sechsten Satz (*Das dicke Kind…Ausleihen bestimmten Tag.*) wieder aufgenommen.

(7) CONTINUATION(k1, k6)

Dazwischen, in den Sätzen 2 bis 5 (k2 bis k5), werden Hintergrundinformationen zum ersten Satz mitgeteilt.

(8) BACKGROUND(k1, κ_1), κ_1:={k2, k3, k4, k5}

Das Subtopik (*main event*) dieses Hintergrundes ist das Ausleihen von Büchern, welches in k2 (*Ich hatte in diesem Winter…zurückbringen sollten.*) eingeführt wird. Das Hintergrundtopik wird durch k3, k4 und k5 elaboriert.

(9) ELABORATION(k2, κ_2), κ_2:={k3, k4, k5}

Zwischen dem ersten Teil von Satz 3 (k3a, *Natürlich kannte…dieser Kinder,*) und dem zweiten Teil von Satz 3 (k3b, *aber es kamen…Straße wohnten.*) besteht eine CONTRAST-Relation, ebenso wie zwischen dem ersten Teil des vierten Satzes (k4a, *Und wenn auch die mehrzahl…in Anspruch nahm,*) und dem zweiten Teil des vierten Satzes (k4b, *so gab es doch…zu lesen begannen.*). k3 und k4 sind jedoch ansonsten gleichgeordnet.

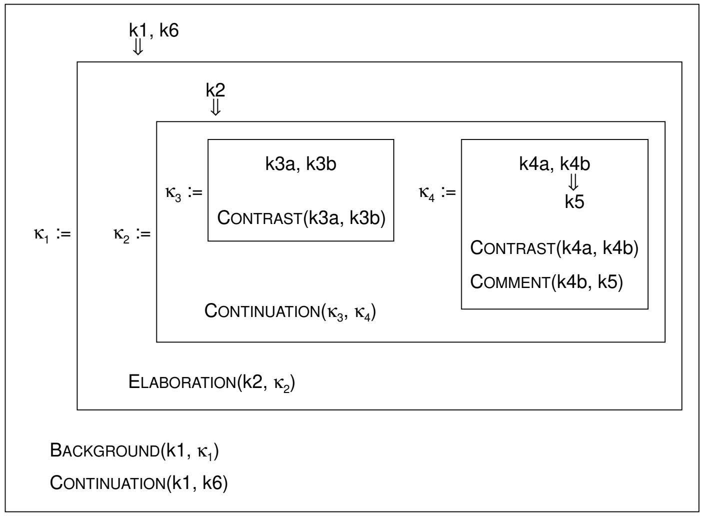
(10) CONTRAST(k3a, k3b)
CONTRAST(k4a, k4b)
CONTINUATION(k3, k4)

k5 schließlich ist ein Kommentar zu k4b.

(11) COMMENT(k4b, k5)

Die Box-Notation in (12) faßt diese SDRS-Analyse konsistent zusammen und kennzeichnet explizit die d-Dominanzbeziehungen (⇓) zwischen SDRSen.

(12)



7.3 SDRSen und Prosodie

Die in diesem Abschnitt vorgestellte Untersuchung hat, ebenso wie die Untersuchung aus Kapitel 5, den Charakter einer *in-depth* Produktionsstudie. Eine der häufigsten zitierten experimentellen Arbeiten zum Thema Diskursstruktur und Prosodie – Hirschberg & Pierrehumbert (1986) – verwendet ein *Analysis-by-Synthesis*-Paradigma. Eine der Voraussetzungen für ein solches Paradigma ist ein gewisses Maß an Vorwissen bezüglich der Relevanz bestimmter Parameter in Hinblick auf den Untersuchungsgegenstand. Da jedoch der Schwerpunkt der hier vorgelegten Studie die Beziehung zwischen Tonregistern und Diskursstruktur ist und dieser Zusammenhang bisher kaum systematisch untersucht wurde, habe ich ein anderes, dem Synthese-Paradigma gewissermaßen vorgeschaltetes Vorgehen gewählt: eine Einzelfallstudie, angelegt als Produktionsexperiment. Diese Form erlaubt es, die im Mittelpunkt stehenden Fragestellungen zunächst exemplarisch

zu untersuchen und Hypothesen herauszuarbeiten, welche dann die Basis für weitergehende und validere Untersuchungen mit anderen Paradigmen bilden können.

Der Text in (5) wurde von einer Sprecherin vorgelesen, die mit dem Inhalt und der DRT-Analyse dieser Sequenz bestens vertraut war – Antje Rossdeutscher –, nicht jedoch mit den prosodischen Fragestellungen, die hinter dieser Studie stehen. Die Äußerung wurde auf DAT-Band aufgenommen, auf eine SGI Indy Workstation überspielt (16 kHz, 16 Bit) und mit ESPS/*waves* und dem in Kapitel 3 vorgestellten Labellingsystem weiterverarbeitet. Die Grundfrequenz wurde zunächst mit ESPS *get_f0* berechnet und dann mit einem Medianfilter (Fensterbreite 5) geglättet, um grob abweichende F_0-Werte zu eliminieren.

7.3.1 Pausen

Bevor ich zu der Analyse von Tonregistern – einem genuin intonatorischen Parameter – komme, möchte ich zunächst kurz einen anderen prosodischen Parameter behandeln, der von verschiedenen Autoren als relativ stabiles Korrelat einer (sehr groben) Diskursgliederung beschrieben wird: Pausendauer. Lehiste (1979), Grosz & Hirschberg (1992), Passenout & Litman (1993) und andere konnten zeigen, daß Pausen nach Sätzen innerhalb von Absätzen kürzer sind als Pausen am Ende eines Absatzes. Die Studie von Ayers (1994), die sowohl spontane als auch gelesene Diskurse untersucht hat, konnte dieses Ergebnis dagegen in dieser Eindeutigkeit nicht bestätigen (Ayers, 1994: 14f).

Unsere Ergebnisse unterstützen jedoch sehr deutlich die ursprüngliche Hypothese: Auf der Suche nach prosodischen Korrelaten der Diskursstruktur sollte der Parameter Pausendauer berücksichtigt werden. Es scheint sogar so, daß die Abstufung der Pausendauer noch feinere Aspekte der Diskursstruktur widerspiegeln kann, als nur die Unterscheidung zwischen Satzenden und Absatzenden. Abbildung 7-1 zeigt die Pausendauern in der oben beschriebenen Aufnahme von *›Das dicke Kind‹*. Alle im Sprachsignal erkennbaren Pausen wurden gemessen und in das Diagramm eingetragen, d.h. außer den im Diagramm abgetragenen wurden von der Sprecherin keine Pausen produziert. Jeder Balken repräsentiert eine Pause, die Ausprägung auf der y-Achse repräsentiert jeweils die Pausendauer. Die Kürzel zwischen den Balken stehen für die Textabschnitte, die vor bzw. nach einer Pause produziert wurden:

> s1: *Es war Ende Januar, bald nach den Weihnachtsferien, als das dicke Kind zu mir kam.*
> s2a: *Ich hatte in diesem Winter angefangen, an die Kinder aus der Nachbarschaft Bücher auszuleihen,*
> s2b: *die sie an einem bestimmten Wochentag holen und zurückbringen sollten.*
> s3a: *Natürlich kannte ich die meisten dieser Kinder,*
> s3b: *aber es kamen auch manchmal Fremde, die nicht in unserer Straße wohnten.*
> s4a: *Und wenn auch die Mehrzahl von ihnen gerade nur so lange Zeit blieb,*
> s4b: *wie der Umtausch in Anspruch nahm,*
> s4c: *so gab es doch einige, die sich hinsetzten und gleich auf der Stelle zu lesen begannen.*
> s5a: *Dann saß ich an meinem Schreibtisch und arbeitete,*
> s5b: *und die Kinder saßen an dem kleinen Tisch bei der Bücherwand,*
> s5c: *und ihre Gegenwart war mir angenehm und störte mich nicht.*
> s6: *Das dicke Kind kam an einem Freitag oder Samstag, jedenfalls nicht an dem zum Ausleihen bestimmten Tag.*

Der Schwärzungsgrad der Balken deutet eine mögliche Kategorisierung der Pausendauern an. Die längste Pause (schwarzer Balken) wird zwischen dem Ende des fünften Satzes und dem Beginn des sechsten Satzes realisiert, also dort, wo nach einer langen, eingebetteten BACKGROUND-Sequenz das Haupttopik des Diskurses wieder aufgenommen und weitergeführt wird (CONTINUATION(k1, k6))[13]. Eine zweite Klasse von Pausen (dunkelgraue Balken) wird jeweils dort realisiert, wo in der SDRS-Analyse eine d-Dominazbeziehung markiert wurde: zwischen dem ersten und dem zweiten Satz ($k1 \Downarrow \kappa_1$), zwischen

[13] An dieser Stelle befindet sich auch der einzige gekennzeichnete Absatz in der Lesevorlage (vgl. (5)).

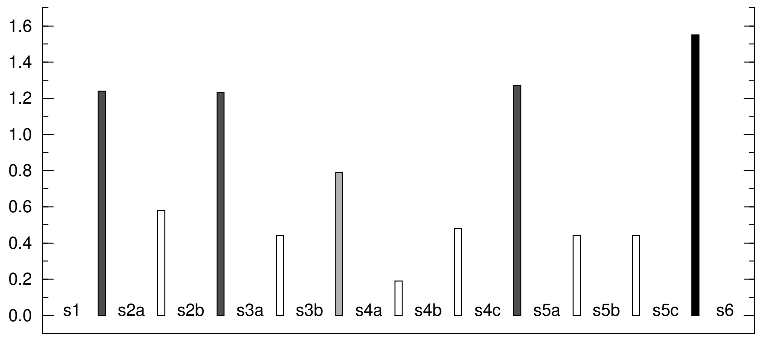


Abbildung 7-1: *Pausendauer zwischen (Teil-) Sätzen in Sekunden (y-Achse). Der Schwärzungsgrad deutet eine mögliche Kategorisierung an.*

dem zweiten und dem dritten Satz ($k2 \Downarrow \kappa_2$) sowie zwischen dem vierten und dem fünften Satz ($k4b \Downarrow k5$). Eine dritte Pausenkategorie (hellgrauer Balken) findet sich zwischen dem Ende des dritten und dem Beginn des vierten Satzes, also zwischen zwei gleichgeordneten SDRSen (κ_3, κ_4). Die restlichen Pausen (weiße Balken) bilden zwar eine relativ inhomogene Gruppe, sind jedoch alle kürzer als 0,6 Sekunden; ihre Funktion scheint die diskursstrukturell irrelevante, interne Gliederung von langen Sätzen zu sein, sie treten nicht an Konstituentengrenzen von SDRSen auf.

Selbstverständlich lassen sich diese Beobachtungen nicht ohne weiteres generalisieren. Die Ergebnisse dieser Untersuchung sind jedoch so klar, daß sich die folgenden Hypothesen formulieren lassen:

1. ‚Sehr lange' Pausen treten nach einer längeren Einbettungssequenz (im Sinne der SDRS-Theorie) mit mindestens einer Instanz von d-Dominanz auf (‚*topic reset*' bzw. Absatzende).

2. ‚Lange' Pausen treten an Konstituentengrenzen auf, wenn die vorangehende Konstituente die nachfolgende Konstituente d-dominiert.

3. ‚Kurze' Pausen sind zwischen gleichgestellten, koordinierten Konstituenten zu erwarten.

4. Die Verwendung von ‚sehr kurzen' Pausen ist beschränkt auf die interne prosodische Gliederung von Sätzen; hinsichtlich der Diskursstruktur sind sie irrelevant.

Die Untersuchung hat außerdem gezeigt, daß eine detaillierte, theoriebasierte Analyse der Diskursstruktur zu einem tieferen Verständnis globaler prosodischer Parameter entscheidend beitragen kann. Nur auf der Basis wohldefinierter diskurssemantischer Objekte und Relationen läßt sich die Interaktion zwischen diskursstruktureller Repräsentation und phonologisch-prosodischer Repräsentation angemessen beschreiben und analysieren. Derartige Analysen müssen jedoch auch – und dies ist ein weiteres wichtiges Resultat dieser Studie – zu einer Erweiterung und Überarbeitung der vorhandenen phonologisch-prosodischen Modelle führen. Mit dem klassischen Tonsequenzmodell in seiner derzeitigen Form können die hier beobachteten Regularitäten der Pausendauer (sollten sich die Hypothesen zumindest teilweise bestätigen) nicht angemessen modelliert werden. Auch die Ergebnisse der Registeranalyse, die ich im folgenden Abschnitt vorstellen werde, lassen eine Modifikation des klassischen Modells notwendig erscheinen. Für den ToBI-Ansatz zur prosodischen Annotation gesprochener Sprache stellt hingegen die Integration von ‚Pausenkategorien' kein besonderes Problem dar. Das *break index tier* mit seinem z.T. sehr eng an die tonale Struktur gekoppelten Inventar bietet durchaus das Potential, theoretisch basierte, diskursgesteuerte Pausenindizes zu definieren und den herkömmlichen Kategorien an die Seite zu stellen bzw. diese zu ersetzen.

7.3.2 Register

Auf der Suche nach prosodischen Korrelaten der Diskursstruktur hat sich die einschlägige Forschung neben den Dauer- bzw. Timingmerkmalen – Pausendauer und Sprechgeschwindigkeit[14] – vor allem auf *pitch range*-Effekte konzentriert (ein Überblick über die Arbeiten der letzten Jahre findet sich in Ayers, 1994). Der Beginn eines neuen Topiks wird in der Regel durch einen expandierten *pitch range* signalisiert (Lehiste, 1975; Brazil, Coulthard & Johns, 1980; Brown, 1983), während der Abschluß eines Topiks bzw. ‚Finalität' mit komprimiertem *range* korreliert (Hirschberg & Pierrehumbert, 1986; Silverman, 1987). Sowohl *pitch range*-Expansion als auch -Kompression sind graduelle Prozesse. Einen Ansatz, diese *range*-Variationen durch eine phonetische Struktur zu repräsentieren und mit einer verfeinerten Topikstruktur zu korrelieren, beschreibt Ayers (1994). In ihrem Ansatz werden intermediäre Phrasen (ip) in einer hierarchischen Baumstruktur gruppiert, wobei der F_0-Wert des höchsten Gipfelakzents innerhalb einer ip das zugrundeliegende Merkmal für die Strukturzuweisung ist. Ayers hat verschiedene Experimente durchgeführt, um den Zusammenhang zwischen dieser phonetischen Struktur und einer unabhängig motivierten, ebenfalls hierarchischen Topikstruktur zu untersuchen. Ihre Ergebnisse sind nicht einheitlich, die Übereinstimmung zwischen phonetischer Struktur und Topikstruktur unterscheidet sich sowohl zwischen den beiden von ihr untersuchten Sprechern als auch zwischen gelesener Sprache und Spontansprache. Die beste Übereinstimmung zwischen phonetischen Segmenten und Diskurssegmenten findet sich in der gelesenen Sprache eines Sprechers. Für die gelesene Sprache des anderen Sprechers und noch mehr für die spontansprachlichen Abschnitte beider Sprecher konnte nur noch eine Korrelation zwischen der höchsten Ebene der Topikstruktur (Einführung neuer Haupttopiks) und der phonetischen Segmentierung gefunden werden; die Segmen-

[14] Die Untersuchungen zur Korrelation zwischen Sprechgeschwindigkeit und Diskursstruktur kommen z.T. zu sehr widersprüchlichen Ergebnissen. Ein Überblick über diese Arbeiten findet sich in Ayers (1994).

tierung in Subtopiks stimmte nur noch selten mit der phonetischen Segmentierung der hierarchischen Baumstruktur überein.

Der Ansatz, den ich im folgenden vorschlagen möchte, unterscheidet sich in einigen wesentlichen Punkten von Ayers‘ Modell und vergleichbaren *range*-Ansätzen. Der wichtigste Unterschied ist die von mir vorgeschlagene kategorielle Beschreibungsebene. Ich werde versuchen, von den graduellen Variationsmöglichkeiten der phonetischen Realisierungsebene zu abstrahieren und die beobachtbaren *range*-Effekte mithilfe von kategoriellen Registermerkmalen zu analysieren. Insofern läßt sich mein Ansatz als eher phonologisch begreifen. Der zweite Unterschied, insbesondere bezüglich Ayers‘ Modell, ist die Berücksichtigung nicht nur hoher, sondern auch tiefer Targets (L-Töne) zur Bestimmung des *range* bzw. des Registers. Drittens verstehe ich die von mir vorgeschlagenen Registermerkmale als globale Eigenschaften von Phrasen und nicht als lokales, nur den Tonhöhenumfang des nuklearen Pitchakzents betreffendes Merkmal. Der vierte Unterschied schließlich betrifft das zugrundeliegende Diskursmodell. Statt auf eine hierarchische Topikstruktur, werde ich mich auf die oben vorgestellte SDRS-Analyse stützen.

Die Transkription in (13) zeigt die Akzentstruktur, die Phrasierung, sowie die Grenztöne der analysierten Äußerung. Der letzte Teilsatz (*Ich hatte vor auszugehen…*; siehe (5)) wurde nicht in die Analyse aufgenommen; er diente nur dazu, eventuelle prosodische Einflüsse, die mit dem Äußerungsende zusammenhängen, auf sich zu ziehen, damit das zu analysierende Teilstück nicht durch diese Einflüsse beeinträchtigt wird.

(13)

1)	*Es war Ende Januar*	H*L H*L H%	k1
2)	*bald nach den Weihnachtsferien*	H*L L*H %	
3)	*als das dicke Kind zu mir kam*	H*L L%	
4)	*Ich hatte in diesem Winter angefangen*	H* ..L H*L !H*L H%	k2
5)	*an die Kinder aus der Nachbarschaft Bücher auszuleihen*	H*L L*H H*L H%	
6)	*die sie an einem bestimmten Wochentag*	H* ..L H*L L*!H %	
7)	*holen und zurückbringen sollten*	L*H !H*L L%	
8)	*Natürlich kannte ich die meisten dieser Kinder*	H* H*L L*H H%	k3a
9)	*aber es kamen auch manchmal Fremde die nicht in unserer Straße wohnten*	H* !H*L H*L L%	k3b
10)	*Und wenn auch die Mehrzahl von ihnen gerade nur so lange Zeit blieb*	L*H L*HL H%	k4a
11)	*wie der Umtausch in Anspruch nahm*	H*L L*!H %	
12)	*so gab es doch einige*	L*H %	k4b
13)	*die sich hinsetzten und gleich auf der Stelle zu lesen begannen*	H*L L*H L*H HH*L HH*L L%	
14)	*Dann saß ich an meinem Schreibtisch und arbeitete*	H* ..L H*L L*!H %	k5
15)	*und die Kinder saßen an dem kleinen Tisch bei der Bücherwand*	H*L L*!H H* L*H %	
16)	*und ihre Gegenwart war mir angenehm und störte mich nicht*	L*H L*H L*H H*L L%	
17)	*Das dicke Kind kam an einem Freitag oder Samstag*	H*L H*L L*H !H*L %	k6
18)	*jedenfalls nicht an dem zum Ausleihen bestimmten Tag*	L*H L*!H H*L L%	

Am Beispiel der ersten drei Intonationsphrasen (entspricht dem ersten Satz der Äußerung) möchte ich meine Analyse demonstrieren. Die F_0-Kontur dieser Intonationsphrasen ist in Abbildung 7-2 dargestellt; Quadrate kennzeichnen H-Targets, Kreise kennzeichnen L-Targets. Das Register einer Intonationsphrase – dargestellt durch eine obere und eine untere horizontale Linie – wurde zunächst definiert als das Intervall zwischen dem höchsten H-Target und dem tiefsten L-Target. Da jedoch der tatsächliche Tonhöhenumfang eines individuellen Pitchakzents außer vom gewählten, zugrundeliegenden Regi-

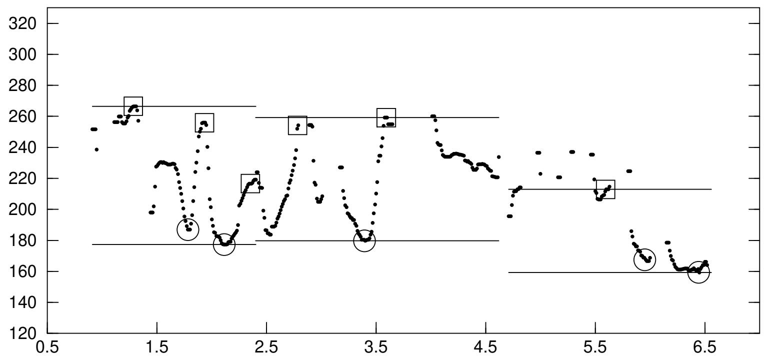


Abbildung 7-2: *F_0-Kontur der ersten 3 Intonationsphrasen (x-Achse: Zeit in Sekunden; y-Achse: Frequenz in Herz). Quadrate kennzeichnen H-Targets, Kreise L-Targets, die horizontalen Linien beschreiben den* range *jeweils bezogen auf eine Intonationsphrase.*

ster auch von lokalen F_0-Zuweisungsregeln determiniert werden kann (Clements, 1990: 65ff), wurde dieses Vorgehen in zwei Fällen revidiert. Das erste H-Target in den Intonationsphrasen 14 und 15 (*Dann saß*(H*) *ich…* und *und die Kin*(H*)*der saßen…*) wurde jeweils unverhältnismäßig hoch realisiert im Vergleich mit den übrigen H-Targets innerhalb der selben Phrase. Gleichzeitig waren die beiden betroffenen betonten Silben auch sehr stark gelängt. In Clements‘ Modell (Clements, 1990: 66ff) sind lokale, individuelle Pitchakzente betreffende F_0-Zuweisungsregeln sehr stark restringiert. Ein Fall, der die Anwendung einer solchen Regel lizenziert, ist *expressive raising*, also die Erhöhung eines H-Targets z.B. bei besonderer Emphase. Da dieser individuelle F_0-Faktor unabhängig vom Register zugewiesen wird, muß eine Registeranalyse, die auf Produktionsdaten basiert, solche Faktoren zunächst neutralisieren, bevor das Register einer Intonationsphrase kalkuliert wird. Basierend auf diesem Modell wurden die beiden erwähnten Akzente als eine Instanz von *expressive raising* eingestuft und bei der Registerkalkulation nicht berücksichtigt. Die Motivation für diese Interpretation war – um es noch ein-

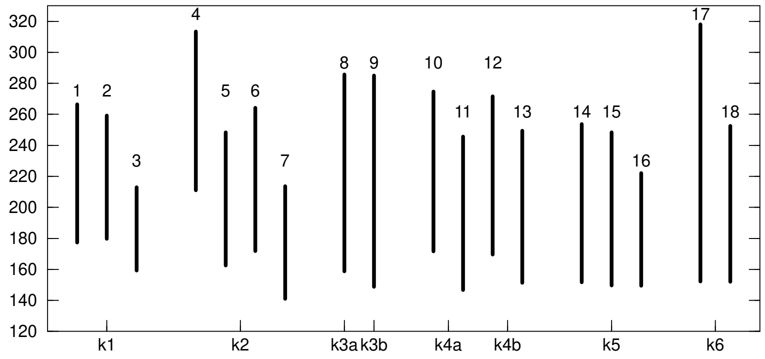


Abbildung 7-3: *Registerkandidaten für alle 18 Intonationsphrasen, geordnet nach* SDRS*-Konstituenten (k1-k6).*

mal zu wiederholen – die unverhältnismäßige Höhe der H-Targets, kombiniert mit der starken Längung der betonten Silben. Eine Erklärung für diese Phänomene ist möglicherweise der sehr ausgeprägte ‚Vorlesestil‘ der Sprecherin.

Die Registerkandidaten, die auf diese Weise gewonnen wurden, sind in Abbildung 7-3 für alle 18 Intonationsphrasen abgetragen. Jeder Balken repräsentiert für eine Intonationsphrase das Intervall zwischen tiefstem L-Target und höchstem H-Target. Auf der x-Achse ist die Zugehörigkeit der Intonationsphrasen zu SDRS-Konstituenten (vgl. Abschnitt 7.2) gekennzeichnet, d.h. die Intonationsphrasen 1 bis 3 entsprechen k1, die Phrasen 4 bis 7 entsprechen k2 usw.

Abbildung 7-4 stellt den Versuch dar, diese Registerkandidaten zu klassifizieren und verschiedenen Registerkategorien zuzuordnen. Diese Klassifikation beruht zum einen auf Ähnlichkeiten der phonetischen Realisierung von gruppierten Registerkandidaten, zum anderen aber auch auf Rückschlüssen aus der im folgenden entwickelten Analyse. Die Bezeichnungen der Registerkategorien (x-Achse) sind beliebig gewählt. In der hier entwickelten linguistischen, d.h. phonologisch-phonetischen Repräsentation werden die

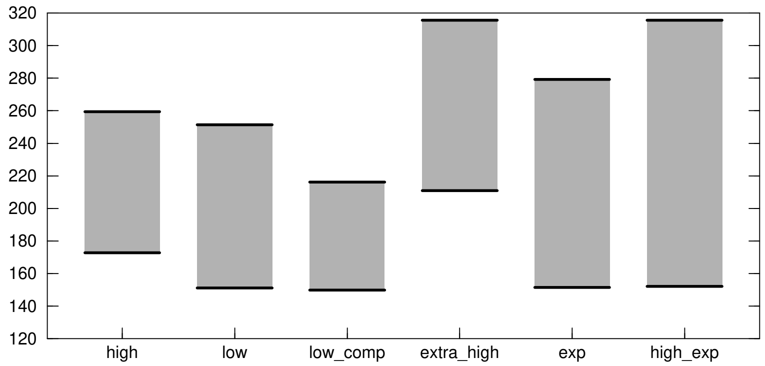


Abbildung 7-4: *Registerkategorien, basierend auf phonetisch ähnlich realisierten Kandidaten aus Abbildung 7-3.*

Register dann durch Kombinationen der Merkmale h und l charakterisiert. Das *default*-Register für nicht-finale Intonationsphrasen (nicht-final bezüglich der übergeordneten SDRS-Konstituente) wird in Abbildung 7-4 als high-Register (h) bezeichnet. In ihm sind die Phrasen 1, 2, 5 und 6 zusammengefaßt.[15] Das high-Register erreicht weder die untere noch die obere Grenze des Äußerungsstimmumfangs der Sprecherin (der Äußerungsstimmumfang reicht hier von ca. 150 Hz bis ca. 320 Hz). Das *default*-Register für finale Phrasen, welches die Phrasen 11, 13 und 18 repräsentiert, wird mit low (l) bezeichnet. Es reicht bis an die untere Grenze des Äußerungsstimmumfangs, der obere Grenzwert liegt geringfügig unter dem des high-Registers.

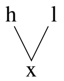
Die restlichen vier Register stellen Modifikationen der beiden *default*-Register dar (Clements, 1981). Ich werde weiter unten versuchen zu zeigen, daß diese Modifikationen durch die spezifische Diskursstruktur ausgelöst werden. low_comp (*low compressed*) ist

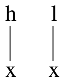
[15] Die in Abbildung 7-4 abgetragenen Intervalle basieren jeweils auf den Mittelwerten der oberen bzw. unteren Grenzwerte der gruppierten Registerkandidaten.

eine Modifikation von low; low_comp reicht ebenfalls bis an die untere Grenze des Äußerungsstimmumfangs, die obere Grenze ist jedoch im Vergleich zu low deutlich abgesenkt; das Ergebnis ist ein komprimiertes Register im unteren Bereich des Stimmumfangs. Beispiele für low_comp sind die Phrasen 3, 7 und 16. Eine Modifikation des high-Registers stellt extra_high dar (Phrase 4); extra_high reicht bis an die obere Grenze des Stimmumfangs, während der untere Grenzwert deutlich angehoben ist, so daß zwar die Lage, nicht jedoch der Umfang des Registers verändert wird. low_comp und extra_high werden durch eine einfach Modifikation des zugrundeliegenden Registers wie in (14) repräsentiert.

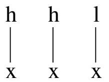
(14)	low_comp	extra_high	
	l	h	Modifikation
	|	|	
	l	h	zugrundeliegendes Register
	|	|	
	x	x	Intonationsphrasen

Das in Abbildung 7-4 mit exp (*expanded*) bezeichnete Register ist eine komplexe Kategorie, die zunächst die Phrasen 8 und 9 repräsentiert. Die Kombination der Phrasen 10/11 bzw. 12/13 weist jedoch einen sehr ähnlichen Umfang auf, d.h. die oberen Grenzen der Phrasen 8, 9, 10 und 12 liegt auf ähnlichem Niveau (nämlich zwischen der oberen Grenze von high und der oberen Grenze von extra_high) und die unteren Grenzen der Phrasen 8, 9, 11 und 13 sind ebenfalls vergleichbar; sie reichen bis an die untere Grenze des Äußerungsstimmumfangs. Aufgrund der identischen Diskursrelation zwischen k3a und k3b bzw. zwischen k4a und k4b (CONTRAST), möchte ich eine gemeinsame Analyse der assoziierten Intonationsphrasen 8, 9, 10/11 und 12/13 vorschlagen. Diese Analyse basiert auf der Annahme, daß jede SDRS-Konstituente mit mindestens einem h-Register und genau einem l-Register realisiert wird, wobei alle nicht-finalen Phrasen mit h und die finale Phrase mit l assoziiert wird. Daraus ergibt sich für k3a bzw. k3b die Repräsentation in (15) und für k4a bzw. k4b die Repräsentation in (16).

(15)


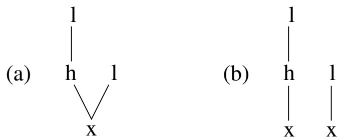
(16)


Wird eine Konstituente durch mehrere Intonationsphrasen realisiert (z.B. k1), ergibt sich die Repräsentation in (17).

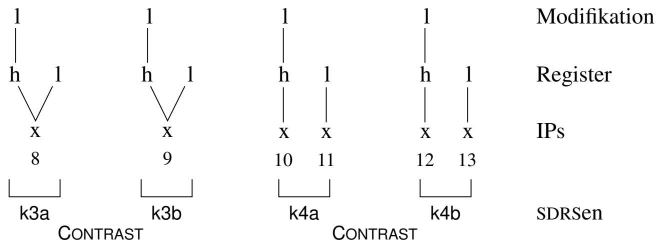
(17)


Das in Abbildung 7-4 postulierte exp-Register wird nun auf der Basis dieser zugrundeliegenden Struktur reanalysiert als l-Modifikation des zugrundeliegenden obligatorischen h-Registers, gefolgt vom – ebenfalls obligatorischen – finalen l-Register.

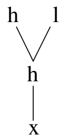
(18) Reanalyse exp-Register



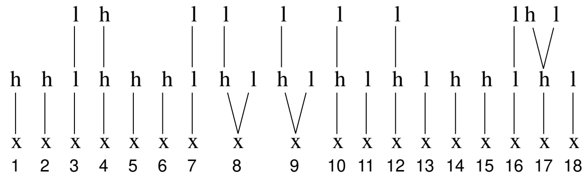
Die Phrase 8 und die Phrase 9 werden jeweils durch die Analyse in (18)(a) repräsentiert, die Phrasen 10/11 und die Phrasen 12/13 durch die Analyse in (18)(b). Das heißt, die Gemeinsamkeit der SDRS-Konstituenten k3a, k3b, k4a und k4b liegt darin, daß aufgrund der identischen Diskursrelation, nämlich CONTRAST, jeweils das initiale h-Register mit l modifiziert wird.

(19)


Das letzte der in Abbildung 7-4 postulierten Register ist high_exp (*high expanded*), welches die Phrase 17 repräsentiert. Es umfaßt den gesamten Äußerungsstimmumfang. Ich möchte hierfür die Repräsentation in (20) vorschlagen, also die hl-Modifikation des zugrundeliegenden initialen h-Registers.

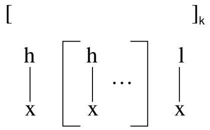
(20)


So ergibt sich für die 18 Intonationsphrasen des gesamten Diskurses die Registeranalyse in (21).

(21)


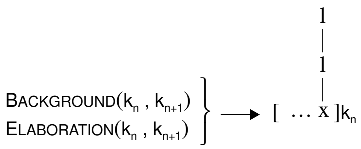
Zusammenfassung: Jede Intonationsphrase ist obligatorisch mit einem zugrundeliegenden Registermerkmal (Basismerkmal) h oder l assoziiert. Alle nicht-finalen Intonationsphrasen innerhalb der Domäne einer SDRS-Konstituente werden mit dem Registermerkmal h assoziiert. Die finale Intonationsphrase innerhalb der Domäne einer SDRS-Konstituente wird mit dem Registermerkmal l assoziiert:

(22)

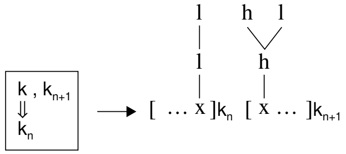


Die Basismerkmale können mit modifizierenden Registermerkmalen (h, l) assoziiert werden. Modifizierende Merkmale werden von der Diskursstruktur kontrolliert und gesteuert. Aus dem hier untersuchten Beispiel können die folgenden, diese Steuerung betreffenden Hypothesen abgeleitet werden:

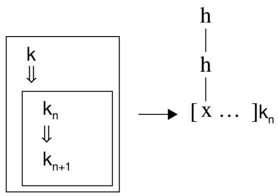
1. BACKGROUND und ELABORATION lösen eine l-Modifikation der letzten Phrase vor der Konstituentengrenze aus; evtl. gilt auch die Generalisierung: d-dominierende Segmente werden mit einer l-Modifikation der finalen Intonationsphrase abgeschlossen (Phrasen 3 und 7).



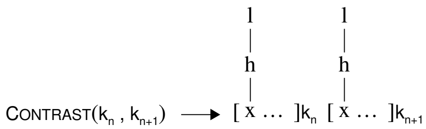
2. Werden zwei Segmente k_n und k_{n+1} nicht von dem selben Segment d-dominiert, so wird die finale Phrase in k_n l-modifiziert und die initiale Phrase in k_{n+1} hl-modifiziert (d.h. nach einer rekursiven Einbettungssequenz (‚*topic reset*') wird eine l-Modifikation der letzten Phrase vor der Konstituentengrenze ausgelöst (Phrase 16) und eine hl-Modifikation der ersten Phrase nach der Konstituentengrenze (Phrase 17)).



3. Die erste Phrase in einem d-dominierenden Segment, welches selbst d-dominiert wird (‚*sub-topic*'), wird h-modifiziert (Phrase 4).



4. CONTRAST(k_n , k_{n+1}) löst eine l-Modifikation der initialen Phrasen in k_n und k_{n+1} aus (Phrasen 8, 9, 10 und 12).



7.4 Diskussion

Die hier vorgeschlagene Registeranalyse stellt den Versuch dar, die linguistische Intuition des Zusammenhangs zwischen Diskursstruktur und globaler tonaler Struktur von Äußerungen, die durch zahlreiche ‚phonetische' Studien unterstützt wird, auf einer kategorialen, phonologischen Ebene zu modellieren. Hierzu wurde das von Clements (1981) zur Analyse von Tonsprachen entwickelte Registermodell adaptiert und zu der expliziten Diskurstheorie von Asher (1993) in Beziehung gesetzt. Die Ergebnisse der *in-depth*-Analyse einer einzelnen Äußerung sind zwar zweifellos vorläufig, aber vielversprechend. Die für die tonale Strukturierung (und auch für die Pausenstruktur) einer Äußerung wichtigste diskursstrukturelle Beziehung scheint die d-Dominanz zu sein. Die d-Dominanz – als eine Eigenschaft bestimmter Diskursrelationen (z.B. BACKGROUND und ELABORATION) zwischen SDRSen – führt zu einer hierarchischen Gliederung von Diskursen. Die Hervorhebung dieser Gliederung bei der lautsprachlichen Realisierung scheint eine der Hauptaufgaben von Registermerkmalen zu sein.

Aber auch Diskursrelationen, die nicht durch d-Dominanz charakterisiert sind, können Registermodifikationen auslösen, z.B. CONTRAST. Darüber hinaus würde man bei einer Erweiterung des Monolog-basierten Diskursmodells zu einem Dialogmodell zusätzliche Abhängigkeiten der Registerkontrolle erwarten. Eine mögliche Konfiguration, die dem postulierten Registermodell immanent ist, in der vorliegenden Analyse jedoch nicht demonstriert werden konnte, ist die h-Modifikation eines zugrundeliegenden, finalen l-Registers:

(23) h
|
l
|
x

In Anlehnung an die ‚Semantik' eines H-Tons auf der Akzent- und Grenztonebene, nämlich Unvollständigkeit zu signalisieren, wäre die Konfiguration in (23) dann zu erwarten, wenn in einer Dialogsituation die Weiterführung des Gesprächs dem Kommunikationspartner überlassen werden soll (*turn taking*).

Zur Weiterentwicklung des hier vorgeschlagenen Ansatzes muß neben der Ausweitung auf Dialoganalysen auch die Unterscheidung zwischen gelesener und Spontansprache berücksichtigt werden. Ayers (1994) konnte zeigen, daß – zumindest in ihrem Untersuchungsparadigma – ein erheblicher Unterschied zwischen spontaner und gelesener Sprache bestand, und zwar in dem Sinne, daß die Korrelation zwischen phonetischer Struktur und Diskursstruktur bei gelesenen Äußerungen sehr viel konsistenter ist als bei spontansprachlichen Äußerungen. Insofern ist es zwar möglich (und evtl. sogar ratsam), anhand exemplarischer gelesener Äußerungen ein prototypisches Modell zu entwickeln, da sich die Grundprinzipien der Beziehung zwischen globalen prosodischen Parametern und globaler Diskursstruktur an diesen Beispielen deutlicher herausarbeiten lassen; eine Überprüfung und gegebenenfalls eine Modifikation des Modells anhand spontansprachlicher Äußerungen ist jedoch unerläßlich.

Die Integration meines Ansatzes in das TSM und ToBI stellt meines Erachtens keine besondere Herausforderung dar. Diejenigen Aspekte, die durch die Registeranalyse abgedeckt werden, werden im klassischen TSM auf der Ebene der phonetischen Realisierungsregeln beschrieben (vgl. z.B. Liberman & Pierrehumbert, 1984), d.h. durch die Addition einer zusätzlichen, unabhängigen Ebene der autosegmentalen Repräsentation – die Ebene der Registerphonologie – und eine Straffung bzw. Umformulierung der Phonetikregeln[16] wäre eine Erweiterung des TSM denkbar, ohne prinzipielle Aspekte des klassischen Modells zu verfälschen. Entsprechend den ToBI-Prinzipien ist eine solche Erweiterung als zusätzliche ToBI-Spur zur Annotation von Registermerkmalen (‚*register*

[16] Die Bezugnahme auf den Äußerungsstimmumfang bei der Beschreibung der Register in Abschnitt 7.3.2 deutet einen phonetischen Parameter an, mit dessen Hilfe Realisierungsregeln formuliert werden könnten. Selbstverständlich sind aber weitere Analysen verschiedener Sprecher und Sprecherinnen nötig, um solche Regeln endgültig zu formulieren.

tier') zu realisieren. Dies ist in einer neueren Version von GToBI(S) (s. Abschnitt 3.2) bereits versuchsweise verwirklicht.

Zusammenfassung und Ausblick

Die Verknüpfung von Modellen der formalen Semantik und Ansätzen aus dem Bereich der suprasegmentalen Phonologie führt zu einem besseren Verständnis und zu einer fundierten Analyse der Prosodie-Semantik-Schnittstelle. Mit der ersten Studie konnte gezeigt werden, daß die Berücksichtigung der Akzentstruktur einer Äußerung zur Disambiguierung von DRSen beitragen kann, und daß umgekehrt die semantische Repräsentation einer Äußerung deren Akzentmuster zumindest teilweise determiniert. Es wurde jedoch auch deutlich, daß nicht alle Aspekte der semantischen Repräsentation phonologisch interpretiert werden, und daß nicht alle Aspekte der phonologischen Form semantisch determiniert sind. Es besteht aber ein regelhafter Zusammenhang zwischen bestimmten Elementen der semantischen Repräsentation (Präsuppositionen, Fokusstruktur) und obligatorischen intonatorischen Merkmalen einer Äußerung. Zukünftige Arbeiten, die durch diese Studie angeregt werden, betreffen zum einen die formale, DR-theoretisch Charakterisierung derjenigen semantischen Elemente, die obligatorisch mit der phonologischen Form korrelieren, und zum anderen die Analyse von optionalen intonatorischen Merkmalen, sowohl vom Standpunkt der Produktion (Synthese) als auch vom Standpunkt der Interpretation (Spracherkennung und -verarbeitung).

Die zweite Studie zeigt, daß die Akzentuierung eines Pronomens alleine nicht ausreicht, um die referentielle Präferenz umzukehren. Die Analyse von Produktionsdaten weist darauf hin, daß Akzentuierung zwar eine notwendige, aber keine hinreichende Bedingung ist. Neben der Akzentposition scheint auch die relative Prominenz eines Akzents ein relevanter Faktor für die Interpretation intonatorischer Strukturen zu sein.

Sowohl das Phänomen der relativen Prominenz selbst als auch sein Einfluß auf die Interpretation einer Äußerung bedürfen weiterer Analysen.

Mit der dritten und letzten Studie schließlich konnte gezeigt werden, daß auch die globale Diskursstruktur mit prosodischen Merkmalen korreliert werden kann. Sowohl interphrasale Pausen als auch das Tonregister von Intonationsphrasen sind von der Konstituentenstruktur von Diskursen, wie sie in Ashers SDRS-Modell beschrieben werden, abhängig. Für die Analyse diskursrelevanter Registermerkmale wurde eine 2-stufige autosegmentale Repräsentation vorgeschlagen. Der Zusammenhang zwischen Diskursstruktur und Prosodie wurde in dieser Studie anhand eines exemplarischen Einzelfalls untersucht. Die Evaluation der postulierten Analyse soll Gegenstand zukünftiger Arbeiten sein.

Mit diesen drei Studien wurden einige Fragen beantwortet, aber mindestens ebenso viele neue Fragen aufgeworfen. Meine Hoffnung ist, daß mit dieser Arbeit zumindest ein kleiner Schritt auf dem Weg zu einem umfassenden phonologisch-semantischen Modell der Prosodie-Semantik-Schnittstelle geschafft ist, und daß vielleicht ein wenig mehr Klarheit darüber besteht, wie dieser Weg weitergegangen werden kann.

Anhang A

Materialien zu Kapitel 6 (1)

Ambige Diskurse und Akzentmuster

Diskurs 1

a. *Die Enkelin verreist nicht mit ihrer Oma.*

b. *Die Oma verreist nicht mit ihrer Enkelin.*

Akzentmuster #	*Sie*	*ist*	*krank.*	
1			H*L	L%
2	H*L			L%
6	L*H		H*L	L%

Diskurs 2

a. *Der Ober verhandelt mit dem Gast.*

b. *Der Gast verhandelt mit dem Ober.*

Akzentmuster #	*Er*	*hat*	*kein*	*Kleingeld.*	
1				H*L	L%
2	H*L				L%
6	L*H			H*L	L%

Diskurs 3

a. *Die Nichte besucht ihre Tante.*

b. *Die Tante besucht ihre Nichte.*

Akzentmuster #	*Sie*	*zeigt*	*ihr*	*neue*	*Fotos.*	
1					H*L	L%
2	H*L					L%
3			H*L			L%
4	L*H		H*L			L%
5	H*L		H*L			L%
6	L*H				H*L	L%
7			L*H		H*L	L%

Diskurs 4

a. *Der Motorradfahrer und der Radfahrer haben einen Unfall.*

b. *Der Radfahrer und der Motorradfahrer haben einen Unfall.*

Akzentmuster #	*Er*	*hat*	*ihm*	*die*	*Vorfahrt*	*genommen.*	
1					H*L		L%
2	H*L						L%
3			H*L				L%
4	L*H		H*L				L%
5	H*L		H*L				L%
6	L*H				H*L		L%
7			L*H		H*L		L%

Diskurs 5

a. *Die Enkelin wäscht bei der Oma.*

b. *Die Oma wäscht bei der Enkelin.*

Akzentmuster #	*Sie*	*muß*	*ihr*	*die*	*Maschine*	*erklären.*	
1					H*L		L%
2	H*L						L%
3			H*L				L%
4	L*H		H*L				L%
5	H*L		H*L				L%
6	L*H				H*L		L%
7			L*H		H*L		L%

Diskurs 6

Die Verkäuferin berät die Kundin.

Akzentmuster #	*Sie*	*wird*	*ungeduldig.*	
1			H*L	L%
2	H*L			L%
6	L*H		H*L	L%

Diskurs 7

Der Maurer schimpft mit dem Jungen.

Akzentmuster #	*Er*	*zeigt*	*ihm*	*den*	*Vogel.*	
1					H*L	L%
2	H*L					L%
3			H*L			L%
4	L*H		H*L			L%
5	H*L		H*L			L%
6	L*H				H*L	L%
7			L*H		H*L	L%

Diskurs 8

Der Maler malt den Sänger.

Akzentmuster #	*Er*	*ärgert*	*sich*	*über*	*ihn.*	
1		H*L				L%
2	H*L					L%
3					H*L	L%
4	L*H				H*L	L%
5	H*L				H*L	L%

Diskurs 9

Die Masseurin behandelt die Patientin im Wohnzimmer.

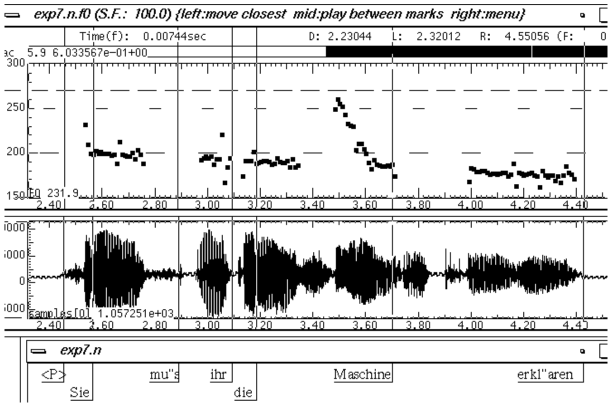
Akzentmuster #	*Sie*	*ist*	*zu*	*ihr*	*nach*	*Hause*	*gekommen*	
1						H*L		L%
2	H*L							L%
3				H*L				L%
4	L*H			H*L				L%
5	H*L			H*L				L%
6	L*H					H*L		L%
7				L*H		H*L		L%

Anhang B

Materialien zu Kapitel 6 (2)

Grundfrequenzkonturen der Akzentmuster (Beispiele)

Akzentmuster #1 (neutraler Akzent)



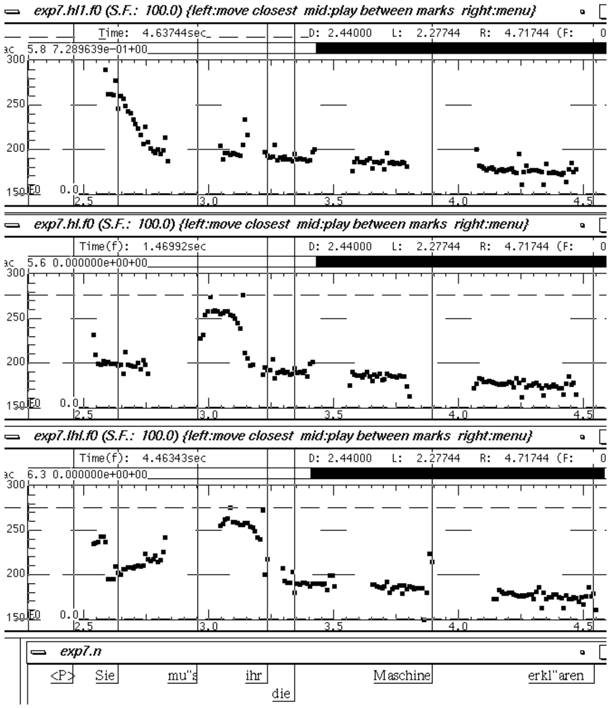
H*L

Akzentmuster #2, #3, #4

#2

#3

#4



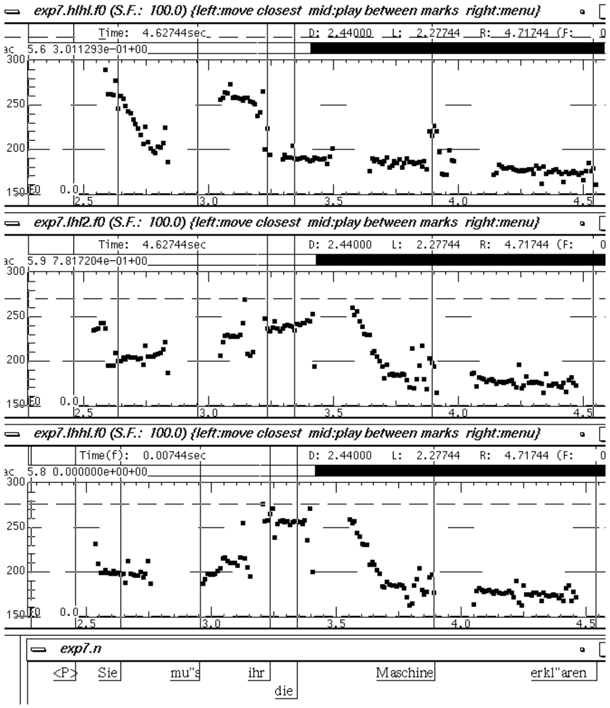
#2	H*L	
#3		H*L
#4	L*H	H*L

Akzentmuster #5, #6, #7

#5

#6

#7



#5	H*L	H*L	
#6	L*H		H*L
#7		L*H	H*L

Bibliographie

ALTMANN H (Hg., 1988) *Intonationsforschungen* (Tübingen: Niemeyer).

ALTMANN H, BATLINER A, OPPENRIEDER W (Hg., 1989) *Zur Intonation von Modus und Fokus im Deutschen* (Tübingen: Niemeyer).

ASHBY M (1978) A study of two English nuclear tones. *Language and Speech* **21,** 326-336.

ASHER N (1993) *Reference to Abstract Objects in Discourse* (Dordrecht: Kluwer Academic Publishers).

AYERS GM (1994) Discourse functions of pitch range in spontaneous and read speech. *Ohio State University Working Papers in Linguistics 44,* 1-49.

BARTELS C, KINGSTON J (1994) Salient pitch cues in the perception of contrastive focus. In: Bosch P, van der Sandt R (Hg.), *Focus and Natural Language processing, Vol.1: Intonation and Syntax* (Heidelberg: IBM Working Papers of the Institute for Logic and Linguistics). Seiten 11-28.

BATTISTELLA EL (1990) *Markedness: The Evaluative Superstructure of Language* (Albany: State University of New York Press).

BECKMAN ME (1986) *Stress and Non-Stress Accent* (Dordrecht: Foris).

BECKMAN ME, AYERS G (1994) *Guidelines for ToBI labelling, version 2.0* (Ohio State University).

BECKMAN ME, EDWARDS J (1994) Articulatory evidence for differentiating stress categories. In: Keating PA (Hg.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form* (Cambridge: Cambridge University Press). Seiten 7-33.

BECKMAN ME, HIRSCHBERG J (1994) *The TOBI annotation conventions* (Ohio State University: Technischer Bericht).

BECKMAN ME, PIERREHUMBERT JB (1986) Intonational structure in Japanese and English. *Phonology Yearbook* **3,** 255-309.

BENDER A (1997) *Aphasie und Kernbereiche der formalen Grammatiktheorie* (Dissertation, Universität Stuttgart).

BOLINGER D (1958) A theory of pitch accent in English. *Word* **14,** 109-149.

BOLINGER D (1982) Intonation and its parts. *Language* **58,** 505-533.

BRAZIL D, COULTHARD M, JOHNS C (1980) *Discourse Intonation and Language Teaching* (Longman).

BROWN G (1983) Prosodic structure and the given/new distinction. In: Cutler A, Ladd DR (Hg.), *Prosody: Models and Measurements* (Berlin: Springer Verlag). Seiten 67-78.

BRUCE G (1977) *Swedish Word Accents in Sentence Perspective* (Lund: Gleerup).

BÜRING D (1995) *The 59th Street Bridge Accent. On the Meaning of Topic and Focus* (Dissertation, Universität Tübingen).

CAHN J (1990) *The effect of intonation on pronoun referent resolution.* (MIT Media Laboratory: Technical Report 94-06).

CAHN J (1995) The effect of pitch accenting on pronoun referent resolution. *Proceedings, 33rd Annual Meeting of the Association of Computational Linguistics.*

CHOMSKY N, HALLE M (1968) *The Sound Pattern of English* (New York: Harper and Row).

CINQUE G (1993) A null theory of phrase and compound stress. *Linguistic Inquiry* **24,** 239-297.

CLARK HH, MARSHALL CR (1981) Definite reference and mutual knowledge. In: Joshi AK, Webber B, Sag IA (Hg.), *Elements of Discourse understanding* (Cambridge: Cambridge University Press).

CLARK J, YALLOP C (1995^2) *An Introduction to Phonetics and Phonology* (Oxford, Cambridge: Blackwell).

CLASSEN K (1996) *Phonetische Korrelate der Wortbetonung im Deutschen* (Magisterarbeit, Universität Stuttgart).

CLEMENTS GN (1981) The hierarchical representation of tone features. In: Dihoff IR (Hg.), *Current Approaches to African Linguistics, Vol. 1* (Dordrecht: Foris). Seiten 145-176.

CLEMENTS GN (1990) The status of register in intonation theory. In: Kingston J, Beckman ME (Hg.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press). Seiten 58-71.

COLLIER R, 'T HART J (1981) *Cursus Nederlandse Intonatie* (Louvain: Acco/De Horstink).

COUPER-KUHLEN E (1996) *Prosody in Conversation: Interactional Studies* (Cambridge: Cambridge University Press).

CRUTTENDEN A (1986) *Intonation* (Cambridge: Cambridge University Press).

DOGIL G (1979) *Autosegmental Account of Phonological Emphasis* (Edmonton: Linguistic Research, Inc.).

DOGIL G (1995) The phonetic manifestation of word stress. *Phonetik-AIMS (Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, Universität Stuttgart)* **2(2),** 1-59. (erscheint in: van der Hulst HG (Hg.), *Word Prosodic Systems of European Languages* (Berlin, New York: De Gruyter)).

DOGIL G, KUHN J, MAYER J, RAPP S (1997) Entwicklung eines prosodischen Moduls für die Diskursrepräsentationstheorie (Zwischenbericht C4). In: *Sonderforschungsbereich 340, Sprachtheoretische Grundlagen für die Computerlinguistik, Arbeits- und Ergebnisbericht 1995-1996-1997* (Universität Stuttgart, Universität Tübingen). Seiten 377-415.

EBERLE K (1996) *Disambiguation by information structure in DRT* (Manuskript, Universität Stuttgart).

FÉRY C (1993) *German Intonational Patterns* (Tübingen: Niemeyer).

FUJISAKI H (1983) Dynamic characteristics of voice fundamental frequency in speech and singing. In: MacNeilage PF (Hg.), *The Production of Speech* (Heidelberg: Springer). Seiten 39-55.

GARTENBERG R, PANZLAFF-REUTER C (1991) Production and perception of F_0 peak patterns in German. In: Kohler KJ (Hg.), *Studies in German Intonation* (Universität Kiel: Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Nr. 25). Seiten 29-113.

GIBBON D (1976) *Perspectives of Intonation Analysis* (Frankfurt: Peter Lang; Forum Linguisticum 9).

GOLDSMITH J (1976) *Autosegmental Phonology* (PhD Thesis, MIT) (veröffentlicht 1979, New York: Garland Press).

GOLDSMITH J (1990) *Autosegmental and Metrical Phonology* (Cambridge, Oxford: Basil Blackwell).

GÓSY M, TERKEN J (1994) Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* **22,** 269-281.

GRICE M (1995) *The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory* (Tübingen: Niemeyer).

GRICE M, BENZMÜLLER R (1995) *Transcription of German using TOBI-tones: the Saarbrücken System* (Universität des Saarlandes, Saarbrücken: Technischer Bericht).

GRICE M, REYELT M, BENZMÜLLER R, MAYER J, BATLINER A (1996) Consistency in transcription and labelling of German intonation with GTOBI. *Proceedings, 4th International Conference on Spoken Language Processing,* Philadelphia, 1716-1719.

GROSZ BJ, HIRSCHBERG J (1992) Some intonational characteristics of discourse structure. *Proceedings, 2nd International Conference on Spoken Language Processing,* Banff/Canada, 429-432.

GROSZ BJ, JOSHI AK, WEINSTEIN S (1983) Providing a unified account of definite noun phrases in discourse. *Proceedings, 21st Annual Meeting of the Association of Computational Linguistics,* Cambridge MA, 44-50.

GROSZ BJ, JOSHI AK, WEINSTEIN S (1995) Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics* **21,** 203-225.

GROSZ BJ, SIDNER CL (1986) Attention, intentions, and the structure of discourse. *Computational Linguistics* **12,** 175-198.

GUSSENHOVEN C (1984) *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

GUSSENHOVEN C, RIETVELD ACM (1988) Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics* **16,** 355-369.

HALLE M, IDSARDI W (1995) General properties of stress and metrical structure. In: Goldsmith J (Hg.), *The Handbook of Phonological Theory* (Oxford, Cambridge: Basil Blackwell). Seiten 403-443.

HALLE M, VERGNAUD J-R (1987) *An Essay on Stress* (Cambridge: MIT Press).

HAYES B (1989) The prosodic hierarchy in meter. In: Kiparsky P, Youmans G (Hg.), *Rhythm and Meter* (Orlando: Academic Press). Seiten 201-260.

HAYES B, LAHIRI A (1991) Bengali intonational phonology. *Natural Language and Linguistic Theory* **9,** 47-96.

HERMES D, VAN GESTEL J (1991) The frequncy scale of speech intonation. *Journal of the Acoustical Society of America* **90,** 97-102.

HIRSCHBERG J (1996) The pragmatics of prosodic variation. *Vortrag, Tagung des SFB 340, „Interfaces of Grammar: Theoretical and Computational Aspects“.* Universität Tübingen, 30.9. - 3.10.1996.

HIRSCHBERG J, PIERREHUMBERT JB (1986) Intonational structuring of discourse. *Proceedings, 24th Annual Meeting of the Association for Computational Linguistics,* New York, 136-144.

HIRST D (1993) Detaching intonational phrases from syntactic structure. *Linguistic Inquiry* **24,** 781-788.

HOBBS JR (1990) The Pierrehumbert-Hirschberg theory of intonational meaning made simple: comments on Pierrehumbert and Hirschberg. In: Cohen P, Morgan J, Pollack M (Hg.), *Intentions in Communications* (Cambridge: MIT Press). Seiten 313-323.

HOMBERT JM (1978) Consonant types, vowel quality, and tone. In: Fromkin VA (Hg.), *Tone: A Linguistic Survey.* (New York: Academic Press). Seiten 77-111.

HYMAN L, KATAMBA F, WALUSIMBI L (1987) Luganda and the Strict Layer Hypothesis. *Phonology Yearbook* **4,** 87-108.

JAKOBSON R (1990) The concept of mark / with Krystyna Pomorska. In: Waugh LR, Monville-Burston M (Hg.), *On Language / Roman Jakobson* (Cambridge: Harvard University Press). Seiten 134-140.

JESSEN M (1994) *A Survey of German Word Stress* (Manuskript, Universität Stuttgart).

JESSEN M, MARASEK K, SCHNEIDER K, CLASSEN K (1995) Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. *Proceedings, 13th International Conference of Phonetic Science,* Stockholm, 428-431.

JILKA M (1996) *Regelbasierte Generierung natürlich klingender Intonation des Amerikanischen Englisch* (Magisterarbeit, Universität Stuttgart).

JOSHI AK (1982) The role of mutual beliefs in question-answer systems. In: Smith N (Hg.), *Mutual Knowledge* (New Nork: Academic Press).

KAMEYAMA M (1994a) *Indefeasible semantics and defeasible pragmatics* (Menlo Park CA: SRI International, Technical Note 544).

KAMEYAMA M (1994b) Stressed and unstressed pronouns: complementary preferences. In: Bosch P, van der Sandt R (Hg.), *Focus and Natural Language Processing, Vol. 3: Discourse* (Heidelberg: IBM Working Papers of the Institute for Logic & Linguistics Nr. 8). Seiten 475-484.

KAMP H (1981) A theory of truth and semantic representation. In: Groenendijk JAG, Janssen TMV, Stokhof MBJ (Hg.), *Formal Methods in the Study of Language* (Amsterdam: Mathematical Centre Tract 135). Seiten 277-322. (Außerdem in: Groenendijk JAG, Janssen TMV, Stokhof MBJ (Hg.), *Truth, Interpretation and Information* (Dordrecht: Foris). Seiten 1-41.)

KAMP H, REYLE U (1993) *From Discourse to Logic* (Dordrecht: Kluwer Academic Publishers).

KAMP H, ROSSDEUTSCHER A (1994a) Remarks on lexical structure and DRS-construction. *Theoretical Linguistics* **20,** 97-164.

KAMP H, ROSSDEUTSCHER A (1994b) DRS-construction and lexically driven inference. *Theoretical Linguistics* **20,** 165-235.

KASCHNITZ ML (1994) *Lange Schatten* (München: dtv).

KLATT DH (1976) Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* **59,** 1208-1221.

KOHLER KJ (1991a) A model of German intonation. In: Kohler KJ (Hg.), *Studies in German Intonation* (Universität Kiel: Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Nr. 25). Seiten 295-360.

KOHLER KJ (1991b) Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. In: Kohler KJ (Hg.), *Studies in German Intonation* (Universität Kiel: Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Nr. 25). Seiten 115-185.

KOHLER KJ (Hg., 1991c) *Studies in German Intonation* (Universität Kiel: Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Nr. 25).

KUHN J (1996) *On Intonation and Interpretation in Context – is there a Unitary Explanation for Focus and Deaccenting?* (Diplomarbeit, Universität Stuttgart).

LADD DR (1978) Stylized intonation. *Language* **54,** 517-539.

LADD DR (1983) Phonological features and intonational peaks. *Language* **59,** 721-759.

LADD DR (1984) Declination: a review and some hypotheses. *Phonology Yearbook* **1,** 53-74.

LADD DR (1990) Metrical representation of pitch register. In: Kingston J, Beckman ME (Hg.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press). Seiten 35-57.

LADD DR (1994) Constraints on the gradient variability of pitch range, or, pitch level 4 lives! In: Keating PA (Hg.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form* (Cambridge: Cambridge University Press). Seiten 43-63.

LADD DR (1996) *Intonational Phonology* (Cambridge: Cambridge Univerity Press).

LADD DR, SILVERMAN KEA (1984) Vowel intrinsic pitch in connected speech. *Phonetica* **41,** 31-40.

LADD DR, VERHOEVEN J, JACOBS K (1994) Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *Journal of Phonetics* **22,** 87-99.

LAKOFF G (1971) Presupposition and relative wellformedness. In: Steinberg DD, Jakobovits LA (Hg.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (Cambridge: Cambridge University Press).

LAVER J (1980) *The Phonetic Description of Voice Quality* (Cambridge: Cambridge Univerity Press).

LEHISTE I (1975) The phonetic structure of paragraphs. In: Cohen A, Nooteboom SG (Hg.), *Structure and Process in Speech Perception* (Heidelberg: Springer). Seiten 195-203.

LEHISTE I (1979) Perception of sentence and paragraph boundaries. In: Lindblom B, Öhman S (Hg.), *Frontiers of Speech Research* (London: Academic Press). Seiten 191-201.

LIBERMAN M (1975) *The Intonational System of English* (PhD Thesis, MIT).

LIBERMAN M, PIERREHUMBERT J (1984) Intonational invariance under changes in pitch range and length. In: Aronoff M, Oehrle R (Hg.), *Language Sound Structure* (Cambridge: MIT Press). Seiten 157-233.

LIBERMAN M, PRINCE A (1977) On stress and linguistic rhythm. *Linguistic Inquiry* **8,** 249-336.

MARASEK K (1997) *EGG and Voice Quality* (Habilitationsschrift, Universität Stuttgart).

MAYER J (1995a) *Transcription of German intonation: the Stuttgart System* (Universität Stuttgart: Technischer Bericht).

MAYER J (1995b) Questionnaire for prosodic labelling systems/approaches in German. http://www.ims.uni-stuttgart.de/phonetik/joerg/stockholm/questionnaire.html.

MENNEN I, DEN OS E (1993) Intonation of modern Greek sentences. *Proceedings of the Institute of Phonetic Science, University of Amsterdam* **17,** 111-128.

MÖHLER G (1996) Rule based generation of fundamental frequency contours for German utterances. In: Bateman JA (Hg.), *Speech Generation in Multimodal Information Systems and its Practical Applications (Proceedings of the 2nd SPEAK! Workshop, Darmstadt 1995)* (Sankt Augustin: GMD-Studien Nr. 302). Seiten 56-59.

MÖHLER G (i.V.) *Eine theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese* (Dissertation, Universität Stuttgart).

MÖHLER G, DOGIL G (1995) Test environment for the two level model of germanic prominence. *Proceedings, 4th EUROSPEECH,* Madrid, 1019-1022.

MOULINES E, CHARPENTIER F (1990) Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* **9,** 435-467.

NAKATANI CH (1993) Accenting on pronouns and proper names in spontaneous narrative. *Proceedings, ESCA Workshop on Prosody* (Dept. of Linguistics and Phonetics, Lund, Sweden: Working Papers 41), 164-167.

NAKATANI CH (1994) Discourse structural constraints on accent in narrative. *Proceedings, 2nd ESCA/IEEE Workshop on Speech Synthesis,* New Paltz NY, 228-231.

NAKATANI CH (1997) Discourse structural constraints on accent in narrative. In: van Santen JPH, Sproat RW, Olive JP, Hirschberg J (Hg.), *Progress in Speech Synthesis* (New York: Springer). Seiten 139-156.

NEPPERT J, PÉTURSSON M (1986²) *Elemente einer akustischen Phonetik* (Hamburg: H. Buske Verlag).

NESPOR M, VOGEL I (1986) *Prosodic Phonology* (Dordrecht: Foris).

OHALA JJ (1978) The production of tone. In: Fromkin VA (Hg.), *Tone: A Linguistic Survey* (New York: Academic Press). Seiten 5-39.

PASSENOUT RJ, LITMAN DJ (1993) Feasibility of automated discourse segmentation. *Proceedings, 31st Annual Meeting of the Association for Computational Linguistics,* Ohio State University, 148-155.

PIERREHUMBERT J (1980) *The Phonology and Phonetics of English Intonation* (PhD Thesis, MIT).

PIERREHUMBERT J, BECKMAN ME (1988) *Japanese Tone Structure* (Cambridge: MIT Press).

PIERREHUMBERT J, HIRSCHBERG J (1990) The meaning of intonational contours in the interpretation of discourse. In: Cohen P, Morgan J, Pollack M (Hg.), *Intentions in Communications* (Cambridge: MIT Press). Seiten 271-311.

PITRELLI J, BECKMAN ME, HIRSCHBERG J (1994) Evaluation of prosodic transcription labeling reliability in the TOBI framework. *Proceedings, 3rd International Conference on Spoken Language Processing,* Yokohama, 123-126.

PRICE PJ, OSTENDORF M, SHATTUCK-HUFNAGEL S, FONG C (1991) The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* **90,** 2956-2970.

RAPP S (1994) *Maschinelles Lernen von einigen Aspekten des deutschen Wortakzents* (Diplomarbeit, Universität Stuttgart).

RAPP S (1995) Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models / An aligner for German. *Proceedings, ELSNET goes east and IMACS Workshop „Integration of Language and Speech in Academia and Industry“,* Moskau, 152-163.

RAPP S (i.V.) *Automatische Erstellung von Korpora für die Prosodieforschung* (Dissertation, Universität Stuttgart).

REYELT M, BATLINER A (1994) *Ein Inventar prosodischer Etiketten für VERBMOBIL* (TU Braunschweig, LMU München: Verbmobil-Memo 34).

REYELT M, GRICE M, BENZMÜLLER R, MAYER J, BATLINER A (1996) Prosodische Etikettierung des Deutschen mit TOBI. In: Gibbon D (Hg.), *Natural Language Processing and Speech Technology* (Berlin, New York: Mouton de Gruyter). Seiten 144-155.

RIETVELD ACM, GUSSENHOVEN C (1985) On the relation between pitch excursion size and prominence. *Journal of Phonetics* **13,** 299-308.

ROCA I (1994) *Generative Phonology* (London, New York: Routledge).

ROOTH M (1985) *Association with Focus* (PhD Thesis, University of Massachusetts, Amherst).

ROOTH M (1992) A theory of focus interpretation. *Natural Language Semantics* **1,** 75-116.

ROOTH M (1997) *On the Interface Principles for Intonational focus* (Manuskript, Universität Stuttgart).

ROSSDEUTSCHER A (1994) Fat child meets DRT. A semantic representation for the opening lines of Kaschnitz' ›Das dicke Kind‹. *Theoretical Linguistics* **20,** 237-305.

SCHARF G (i.V.) *Timingstörungen bei Dysarthrophonie. Zur akustischen und artikulatorischen Realisierung von Vokalquantität bei Patienten mit Morbus Parkinson und Cerebellärer Ataxie* (Dissertation, Universität Stuttgart).

SELKIRK EO (1984) *Phonology and Syntax: The Relationship between Sound and Structure* (Cambridge: MIT Press).

SELKIRK EO (1990) On the nature of prosodic constituency. In: Kingston J, Beckman ME (Hg.), *Papers in Laboratory Phonology, Vol. 1* (Cambridge: Cambridge University Press). Seiten 179-200.

SELTING M (1995) *Prosodie im Gespräch: Aspekte einer interaktionalen Phonologie der Konversation* (Tübingen: Niemeyer).

SILVERMAN KEA (1987) *The Structure and Processing of Fundamental Frequency Contours* (Dissertation, University of Cambridge).

SILVERMAN KEA, BECKMAN ME, PITRELLI J, OSTENDORF M, WIGHTMAN C, PRICE P, PIERREHUMBERT J, HIRSCHBERG J (1992) TOBI: a standard for labelling english prosody. *Proceedings, 2nd International Conference on Spoken Language Processing,* Banff/Canada, 867-870.

SLUIJTER AMC (1995) *Phonetic Correlates of Stress and Accent* (PhD Thesis, Leiden University) (veröffentlicht 1995, The Hague: Holland Academic Graphics).

SLUIJTER AMC, TERKEN JMB (1993) Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica* **50,** 180-188.

SLUIJTER AMC, VAN HEUVEN VJ (1994) *Spectral Balance as an Acoustic Correlate of Linguistic Stress* (Manuskript, Leiden University).

STEEDMAN M (1987) Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory* **5,** 403-439.

STEEDMAN M (1996) *Surface Structure and Interpretation* (Cambridge: MIT Press).

STEEDMAN M (1997) *Information Structure and the Syntax-Phonology Interface* (Manuskript, University of Pennsylvania).

SWERTS M, GELUYKENS R (1993) The prosody of information units in spontaneous monologue. *Phonetica* **50,** 189-196.

'T HART J (1981) Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* **69,** 811-821.

'T HART J, COLLIER R, COHEN A (1990) *A Perceptual Study of Intonation: an Experimental-Phonetic Approach.* (Cambridge: Cambridge University Press).

TRUBETZKOY NS (1958) *Grundzüge der Phonologie* (Göttingen: Vandenhoeck & Ruprecht).

UHMANN S (1991a) *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie* (Tübingen: Niemeyer).

UHMANN S (1991b) On the tonal disambiguation of focus structures. *Journal of Semantics* **8,** 219-238.

UMEDA N (1982) F_0 declination is situation-dependent. *Journal of Phonetics* **10,** 279-290.

VAISSIÈRE J (1983) Language-independent prosodic features. In: Cutler A, Ladd DR (Hg.), *Prosody: Models and measurements* (Heidelberg: Springer). Seiten 53-66.

VÄLIMAA-BLUM R (1993) Intonation: a distinctive parameter in grammatical constructions. *Phonetica* **50,** 124-137.

WARD G, HIRSCHBERG J (1985) Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* **61,** 747-776.

WILLIAMS CE, STEVENS KN (1972) Emotion and speech: some acoustical correlates. *Journal of the Acoustical Society of America* **52,** 1238-1250.

YIP M (1988) The Obligatory Contour Principle and phonological rules: a loss of identity. *Linguistic Inquiry* **19,** 65-100.