

Intonationsforschungen

Herausgegeben von
Hans Altmann

Max Niemeyer Verlag
Tübingen 1988





8771170

CIP-Titelaufnahme der Deutschen Bibliothek

Intonationsforschungen / hrsg. von Hans Altmann. – Tübingen : Niemeyer, 1988.
(Linguistische Arbeiten ; 200)
NE: Altmann, Hans [Hrsg.]; GT

ISBN 3-484-30200-3 ISSN 0344-6727

© Max Niemeyer Verlag Tübingen 1988
Alle Rechte vorbehalten. Ohne Genehmigung des Verlages ist es nicht gestattet,
dieses Buch oder Teile daraus photomechanisch zu vervielfältigen.
Printed in Germany. Druck: Weihert-Druck GmbH, Darmstadt.

0 20113010 7 1 3 1 3

INHALTSVERZEICHNIS

VORWORT

Dieter Wunderlich (Düsseldorf): Der Ton macht die Melodie - Zur Phonologie der Intonation des Deutschen	1
Caroline Féry (Konstanz): Rhythmische und tonale Struktur der Intonationsphrase	41
Susanne Uhmann (Konstanz): Akzenttöne, Grenztöne und Fokussilben. Zum Aufbau eines phonologischen Intonationssystems für das Deutsche	65
Joachim Jacobs (München): Fokus-Hintergrund-Gliederung und Grammatik	89
Leila Iuukko-Vinchenzo (München): Zur Intonation finnischer Fragesätze	135
Wilhelm Oppenrieder (München): Intonation und Identifikation. Kategorisierungstests zur kontextfreien Identifikation von Satzmodi	153
Wilhelm Oppenrieder (München): Intonatorische Kennzeichnung von Satzmodi	169
Anton Batliner (München): Produktion und Prädiktion. Die Rolle intonatorischer und anderer Merkmale bei der Bestimmung des Satzmodus	207
Anton Batliner (München): Modus und Fokus als Dimensionen einer Non-metrischen Multidimensionalen Skalierung	223
Anton Batliner (München): Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des Fo-Gipfels	243
Lieselotte Schiefer und Anton Batliner (München): Intonation, Ordnungseffekt und das Paradigma der Kategorialen Wahrnehmung	273
Doris Rittinghaus (Bochum): Experimente zur Zeitstruktur von synthetischen Phrasierungseinheiten	293
Thomas P. Goldbeck, Walter F. Sendlmeier (Gießen/Nijmegen): Wechselbeziehungen zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen	305

DER TON MACHT DIE MELODIE- ZUR PHONOLOGIE DER INTONATION DES DEUTSCHEN

Dieter Wunderlich (Düsseldorf)

0. Zusammenfassung

Intonation manifestiert sich phonetisch in zeitlichen Änderungen der Grundfrequenz (F_0) bei lautsprachlichen Äußerungen. Charakteristischen Änderungen (F_0 -Konturen) können bestimmte Interpretationen gegeben werden. In manchen Fällen ist die Interpretation grammatisch verankert wie beim Satzakzent, Satzmodus, syntaktischem Kontrast oder bei der Disambiguierung solcher Sätze wie "Er zog und verlor den Springer". Die Intonation hat hier distinktive Eigenschaften, die durch entsprechend distinktive phonologische Werte beschreibbar sind. Dies setzt voraus, daß wir der phonologischen Intonationskomponente einen Ort in der Grammatik zuschreiben können und von einer phonetischen Intonationskomponente unterscheiden, die die distinktiven Werte im Hinblick auf die charakteristischen F_0 -Konturen verarbeitet. Im Einklang z.B. mit Pierrehumbert (1980), Ladd (1983) argumentiere ich dafür, daß eine F_0 -Melodie phonologisch aus einzelnen Tönen zusammengesetzt ist, die nur die Werte "Hoch" und "Tief" annehmen können. Ich argumentiere außerdem dafür, daß die Sprachen universal eine Tonkomponente besitzen, die lexikalisch in Anspruch genommen werden kann (wie bei Tonsprachen) oder nicht (besonders dann steht sie für Funktionen der Intonation zur Verfügung, wie im Deutschen). Daraus ergibt sich, daß die Phonologie der Intonation grundsätzlich mit denselben Kategorien und ähnlichen Prinzipien operieren sollte wie die Tonologie der Tonsprachen. Auf dieser Basis versuche ich, die beobachteten und experimentell überprüften Intonationsmuster des Deutschen phonologisch zu charakterisieren. Indem die Intonation grammatikalisierte Funktionen wie die oben genannten erfüllt, interagiert sie mit Syntax und Lexikon. Ich mache Annahmen über tonale Satzmodusmorpheme sowie darüber, wie Satzakkentsilben aufgrund der Fokus- und Kontraststruktur eines Satzes gefunden werden und welche kontextabhängigen Tonwerte sie erhalten. Mit diesen Annahmen sind alle

Intonationsmuster des Deutschen ableitbar, und gleichzeitig ist auch vorhersagbar, mit welchen Intonationen ein in bestimmter Weise fokussierter bzw. in Kontrastphrasen zerlegbarer Satz geäußert werden kann. Ein interessantes Nebenergebnis ist, daß die Intonation des Deutschen rechtseinstimmende Strukturen gegenüber anderen Strukturen bevorzugt.

1. Funktionen der Intonation

Die Intonation kann zur Steuerung des Diskurses dienen, eine besondere Emphase des Sprechers anzeigen, zu rhetorischen und rituellen Zwecken eingesetzt werden. All dies sind pragmatische Funktionen der Intonation, die vermutlich auf globalen Eigenschaften der Äußerung beruhen. Was uns interessiert, sind grammatikalisierte Funktionen der Intonation.

Zwei wichtige semantische Funktionen der Intonation sind (a) die Kennzeichnung des Satzmodus und (b) die Kennzeichnung der Fokusdomäne einer Äußerung. Beide Funktionen können auf die verschiedenen Komponenten einer Grammatik verteilt sein, teils von Sprache zu Sprache unterschiedlich, teils mit Variation innerhalb einer Sprache. Wir können lexikalische, syntaktische und intonatorische Realisierungen der jeweiligen Funktion unterscheiden.

Der Satzmodus läßt sich lexikalisch durch die Kategorie der Komplementierer oder Satzpartikeln anzeigen. Syntaktisch kann er durch spezielle Wortstellungen gekennzeichnet werden (z.B. Verberst versus Verbzweit im Deutschen). Intonatorisch läßt sich der Satzmodus insbesondere durch die Grenztöne (initial oder final) einer Äußerung festlegen.

Die Fokusdomäne einer Äußerung kann lexikalisch durch gewisse Fokuspartikeln (wie sogar, nur im Deutschen) festgelegt werden, die als Spezifikatoren einer Phrase verwendbar sind. Syntaktisch ist z.B. die Konstruktion des Cleft-Satzes möglich ("Es ist das Mensaessen, das mir nicht bekommt"), außerdem kann eine Sprache spezifische Fokuspositionen aufweisen (z.B. vor dem Verb). Intonatorisch läßt sich die Fokusdomäne durch einen Satzakzent, d.h. durch intonatorische Hervorhebung von ausgezeichneten Silben innerhalb der Domäne kennzeichnen. Welche Silben als ausgezeichnet gelten bzw. welche Domäne durch die hervorgehobenen Silben festgelegt wird, muß durch Prinzipien der Fokusabbildung bzw. Fokusprojektion erfaßt werden (siehe § 5).

Eine weitere Funktion hat die Intonation bei der Abbildung von syntaktischer Hierarchie auf die phonetische Kette. Ein Satz kann in mehreren Into-

... ..
 nationsphrasen realisiert werden, in der Regel wird die letzte Silbe einer Phrase gelängt. Ein satzinternes Phrasenende kann zusätzlich durch einen hohen Grenzton gekennzeichnet werden. Syntaktische Konstituenten können auf Abschnitte der Deklinationslinie einer Äußerung abgebildet werden (vgl. Cooper/Sorensen 1981). Indirekt gibt auch die Art und Verteilung von Akzenttönen eine Information über die syntaktische Gliederung, und zwar kraft der Prinzipien der Fokusabbildung (siehe § 6).

Schließlich dient die Intonation dazu, syntaktisch kontrastierende Elemente zu kennzeichnen. Dadurch ermöglicht sie u.a., die fehlenden Elemente einer koordinativen oder komparativen Konstruktion zu ergänzen, eine extrahierte Phrase auf die passende Domäne zu beziehen, eine Alternativfrage von einer Ja-Nein-Frage zu unterscheiden (siehe § 7).

2. Interaktion der Grammatikkomponenten

Die Sprachtheorie konzipiert die Grammatik einer Sprache als eine Konfiguration von Komponenten mit je spezifischen Domänen und zum großen Teil spezifischen Prinzipien. Abb. 1 veranschaulicht die Grammatikkonfiguration, die ich zugrunde legen möchte und die in groben Zügen wohl weithin akzeptiert ist.

Fett eingerahmt sind die Komponenten der Grammatik: Semantik, Syntax, Phonologie sowie das Lexikon (auf das alle drei einwirken, mit einfachem Pfeil dargestellt). Sie ermöglichen es, eine Äußerung auf verschiedenen Strukturebenen (einfacher Rahmen) zu repräsentieren bzw. zu beschreiben. Die Semantik ist im konzeptuellen System, die Phonologie in der Phonetik verankert (Doppelpfeil); diese Komponenten und die durch sie bestimmten Strukturebenen sind weitgehend außersprachlich und gehören daher nicht zur Grammatik (gestrichelter Rahmen).

Von der Syntax her interessiert hier nur die Oberflächen- bzw. S-Struktur, die den Bedingungen der lexikalischen Projektion und den syntaktischen Konfigurationsmöglichkeiten genügen soll. Die Kontrast-Fokus-Struktur (KF-Struktur) ergibt sich aus der S-Struktur, indem zueinander kontrastierende Konstituenten sowie Fokuskonstituenten ausgezeichnet werden. Die Notwendigkeit einer Kontrastauszeichnung ergibt sich vor allem bei Koordinationen und Komparativkonstruktionen, wo fehlende Elemente semantisch rekonstruiert werden müssen, möglicherweise auch bei Extraktionen. Ein solcher Kontrast wird im allgemeinen auch metrisch und intonatorisch hervorgehoben. (Mehr

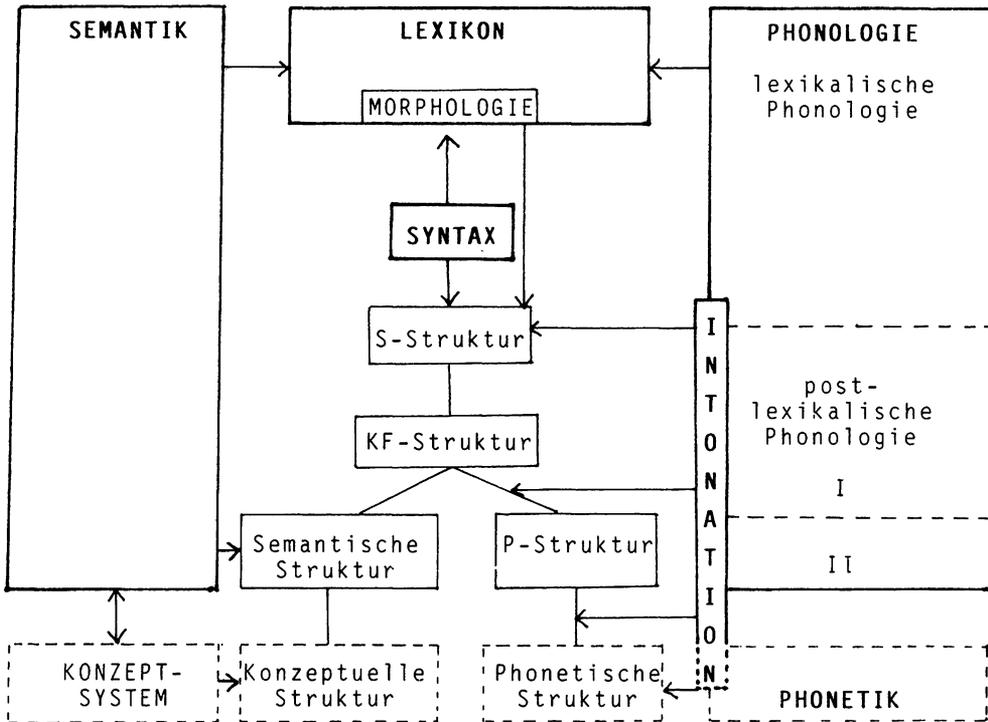


Abb. 1 Konfiguration der Grammatik

dazu in § 7.) Auch unabhängig von möglichen Kontrasten sind in jeder Äußerung bestimmte Informationen semantisch hervorgehoben; sie werden durch syntaktische Fokusdomänen gekennzeichnet, die sich dann intonatorisch auf die Silben des Satzakkentes (= σ^*) abbilden lassen (vgl. dazu Jacobs 1982, v. Stechow/Uhmann 1985).

Die P-Struktur (für 'Prosodisch') umfaßt die metrische und intonatorische Gliederung einer Äußerung. Die metrische Struktur kann durch Kategorien wie Intonationsphrase, phonologische Phrase, phonologisches Wort, Fuß und Silbe (vgl. Nespor/Vogel 1982) und durch deren relative Gewichte zueinander in einem metrischen Baum (vgl. Culicover/Rochemont 1983) dargestellt werden. Die intonatorische Gliederung ergibt sich durch die Verteilung von Tonmerkmalen auf solche Einheiten.

Die Unterscheidung von semantischer und konzeptueller Struktur ist für das hier zu behandelnde Thema nicht wesentlich. Zu bemerken ist aber zweierlei. Erstens ist die Konzeption über den Zusammenhang der Strukturebenen repräsentational und nicht derivational. Das bedeutet, daß die verschiedenen Strukturebenen als simultane Ebenen angenommen werden, zwischen denen keine gerichteten Ableitungsbeziehungen bestehen sollen. (Nur aus Darstellungsgründen wird es manchmal so erscheinen, daß höhere Ebenen auf niedrigere Ebenen abgebildet werden.) Zweitens besteht nach der hier vertretenen Konzeption keine direkte Beziehung zwischen der semantischen Struktur und der P-Struktur. Alle semantischen Funktionen der Intonation müssen daher schon in der KF-Struktur angelegt sein. Für die Kontrast- und Fokusphänomene ist das evident, weil die KF-Struktur gerade zur abstrakten Repräsentation dieser Phänomene dient. Probleme ergeben sich bei der intonatorischen Realisierung des Satzmodus (siehe unten). Pragmatische Funktionen der Intonation sind davon natürlich nicht betroffen, weil sie wie alle pragmatischen Funktionen auf globale (bzw. beliebig instantiierte) Eigenschaften einer Äußerung zurückgehen können.

Man unterscheidet heute zwischen lexikalischer und postlexikalischer Phonologie (vgl. Kiparsky 1982, Wiese 1986b). Lexikalisch sind z.B. die Prinzipien bzw. Regeln, die die Segmentstruktur der Wörter festlegen, ihre Silbifizierung und den Wortakzent. Die Silben eines Wortes haben unterschiedliche Prominenz (ebenfalls durch einen metrischen Baum oder durch die Anzahl von Schlägen in einem metrischen Baum darstellbar, vgl. Liberman/Prince 1977, Hayes 1984, Selkirk 1984, Féry 1986, Wiese 1986a); die jeweils prominenteste von ihnen kann als die Silbe mit dem Wortakzent (= σ^W) angesprochen werden. In der Regel wird der Satzakzent auf einer Silbe mit Wortakzent realisiert. Postlexikalisch sind alle die Phänomene einer Äußerung, die keine unmittelbar semantische oder wortbezogene Funktion haben, wie die rhythmische Gliederung, Klitisierungen beim Schnellsprechen oder die Auslautverhärtung im Deutschen. Soweit diese Phänomene von der syntaktischen Struktur der Äußerung abhängen (wie z.B. die Klitisierung des bestimmten Artikels im Deutschen), rechnet man sie zur postlexikalischen Phonologie I, sonst zur postlexikalischen Phonologie II. Regeln bzw. Prinzipien der lexikalischen versus postlexikalischen Phonologie sind nicht grundsätzlich distinkt (z.B. Silbifizierung oder Schwa-Epenthese im Deutschen, vgl. Wiese 1986b, Pulleyblank 1986), distinkt sind sie primär hinsichtlich ihrer Domäne.

Die Intonation kann nun weitgehend zur postlexikalischen Phonologie

(teils I, teils II) gerechnet werden, aber wie alle phonologischen Phänomene hat sie ein phonetisches Substrat und ist daher in der Phonetik verankert. Mithin sind einerseits natürlich phonetische Untersuchungen notwendig, andererseits ist erwartbar, daß es phonetische Differenzierungen in der Intonation gibt, die kein phonologisches Korrelat haben (oft nicht einmal vom Hörer perzipiert werden).

Der Status des Satzmodus als grammatische Kategorie ist nicht unbestritten. Altmann (1987) nimmt einen modularen Status an, der sich aus der Kombinatorik ganz verschiedenartiger Merkmale ergibt. Damit ist aber das Problem verbunden, daß für die einzelne Komponente (Strukturebene) nicht gesagt werden kann, ob dort die Kategorie anwesend ist oder nicht. Man würde eine eigene Komponente benötigen, in der die Kombinatorik der formalen Parameter zu prüfen ist. Ich möchte stattdessen annehmen, daß der Satzmodus ein syntaktisches Merkmal ist, das immer dem Satz als Ganzen zukommt und eine einheitliche semantische Interpretation erhält; diese Merkmale müssen daher auf der Ebene der KF-Struktur zur Verfügung stehen. Die Merkmale können positiv oder negativ spezifiziert sein. Ich benutze die folgenden Konventionen: [+ASS] für Assertiv bzw. Aussagesatz, [+INT] für Interrogativ bzw. Fragesatz, [+IMP] für Imperativ, [+ECHO] für die Echofrage. Es gelten mindestens die folgenden Implikationen

- (1) a. [+IMP] → [-ASS, -INT]
 b. [+ECHO] → [+INT]

Inwieweit Exclamativ oder Desiderativ eigene Satzmodi sind oder sich lediglich als Instanzen der Merkmalkombination [-ASS, -INT] ergeben, will ich offenlassen.

Im Sinne der Vorstellungen von Travis (1984), Chomsky (1986) läßt sich die zugrundeliegende Struktur des deutschen Satzes wie in (2) formulieren.

- (2) [_{CP} ... C [_{IP} ... [_{VP} ... V] I]]

C für Komplementierer, I für INFL, V für Verb

Generell wird angenommen, daß bei der Erst- oder Zweitstellung des Verbs ein finites Verb (als Instanz von INFL) die Position C besetzt. Die davorstehende Topik-Position kann durch eine XP (mit X = N, V, A, P) aus der IP- oder VP-Domäne besetzt werden. Man kann weiter annehmen, daß ein finites Verb [V, +FIN] in sich die Eigenschaften eines V und eines I vereinigt, also eine lexikalisch fusionierte Kategorie ist (vgl. Wunderlich 1987).

Wenn nun Satzmodus eine Kategorie ist, die dem Satz als Ganzen zukommt,

dann muß sie innerhalb der Konfiguration $[_{CP} \dots C]$ determiniert sein. Es gibt Randfälle im Deutschen, wo bereits ein lexikalisches C den Satzmodus festlegt: ob [+INT] oder daß [-INT] in Beispielen wie (3).

- (3) a. ob du wohl daran denkst
b. und daß du dich bitte beeilst

Wenn V einen morphologischen Imperativ aufweist, dann nimmt V wegen [+IMP] → [+FIN] auch die C-Position ein und legt den Satzmodus fest. Sonst ergeben sich die Satzmodi aus den Besetzungsmöglichkeiten der Topik-Position.

- (4) a. [+FIN, +TOP [+WH]] → [+INT]
b. [+FIN, +TOP [-WH]] → [+ASS]
c. [+FIN, -TOP] → [-ASS]
[+WH] kennzeichnet dabei die WH-Phrasen mit einem lexikalischen Fragewort.

Nach dem bisherigen Vorschlag müßte der Satzmodus in Fällen wie den folgenden auf der Ebene der S-Struktur unbestimmt bleiben.

- (5) a. Geht er wieder langsam. (Interrogativ oder Exclamativ)
b. Gehen Sie nach Hause. (Interrogativ oder Imperativ)

Es ist nicht anzunehmen, daß die Fokusausszeichnung (weder hier noch sonst) irgendetwas mit dem Satzmodus zu tun hat, also kann das Problem auch nicht auf der Ebene der KF-Struktur geregelt werden. Evident ist vielmehr, daß die Sätze durch einen hohen bzw. nicht-hohen finalen Grenzton disambiguiert werden. Ich betrachte dies als den einzig entscheidenden Parameter, egal was sonst noch (z.T. in Folge davon) die Äußerungen intonatorisch zu unterscheiden vermag. Ich möchte nun vermeiden, einfach arbiträr ein Satzmodusmerkmal anzunehmen, das dann intonatorisch zu realisieren ist. Stattdessen nehme ich an, daß es tatsächlich ein Morphem [+INT] gibt, das aber keinerlei segmentale, sondern nur tonale Information (Hochton für $[_{CP}]$) bereitstellt. Für Tonsprachen ist die Existenz solcher Morpheme nichts Ungewöhnliches (vgl. Goldsmith 1976). Man kann wohl auch zeigen, daß die Sprachlerner solche tonalen Morpheme, wenn sie an der Äußerungsgrenze auftreten, sogar früher erwerben als die segmentalen; sie gehören damit zum Lexikon.

3. Vergleich mit Tonsprachen

Eine Tonsprache ist dadurch ausgezeichnet, daß die auf den Silben eines Wortes realisierte Tonhöhe lexikalisch distinktiv ist, also Wortbedeutungen

oder grammatische Morpheme zu unterscheiden vermag. Im Rahmen der nicht-linearen, insbesondere autosegmentalen Phonologie sind die tonalen Eigenschaften der Tonsprachen sehr intensiv studiert worden (vgl. u.a. Goldsmith 1976, Clements/Ford 1979, Halle/Vergnaud 1982, Pulleyblank 1986). Es hat sich gezeigt, daß als Kategorien die Tonmerkmale H (für 'high = hoch') und L (für 'low = tief') ausreichen. Sie bilden eine eigene Schicht von Segmenten, deren Elemente nach generellen Prinzipien entweder mit den V-Positionen (für [+silbisch]) eines CV-Skeletts oder den Elementen einer Silben-Schicht assoziiert werden. Pulleyblank (1986:11) nimmt dafür diese Prinzipien an:

(6) Assoziationskonventionen:

Bilde eine Sequenz von Tönen auf eine Sequenz von tontragenden Einheiten ab:

- (a) von links nach rechts,
- (b) in einer 1:1-Beziehung.

(7) Wohlgeformtheitsbedingung:

Assoziationslinien dürfen sich nicht überkreuzen.

Töne, die aufgrund der lexikalischen Information eines Morphems bereits fest mit einer tontragenden Einheit verbunden sind, können der Konvention (6a) nicht mehr unterliegen, weil sonst (6b) verletzt würde. Von den Konventionen (6) werden also nur die freischwebenden ('floating') Töne erfaßt.

Die Assoziationskonventionen werden automatisch auf jeder Derivationsstufe im Lexikon angewendet (ähnlich wie z.B. Silbifizierungsprinzipien). Daneben gibt es Regeln des Tonsandhi, die postlexikalisch wirksam sind und bestimmte Tonwerte in der Nachbarschaft anderer Töne innerhalb einer Intonationsphrase modifizieren. Diese Regeln können von zweierlei Art sein. Entweder gehört das Ergebnis wieder zum phonologischen Repertoire der Töne {H, L}, oder das Ergebnis legt bestimmte relative Tonwerte für die phonetische Realisierung fest (z.B. auch mittlere Tonhöhen). Die Regeln der ersten Art gehören zur postlexikalischen Phonologie, ein Beispiel dafür sind die Regeln des Tonsandhi im Chinesischen (in der Analyse von Wiese 1986b:122f).

- (8) a. $L \rightarrow H / \begin{array}{c} L \quad L \\ \vee \quad \vee \\ \sigma \quad \sigma \end{array}$
- b. $\rightarrow H / \begin{array}{c} L \quad L \\ \vee \quad \vee \\ \sigma \quad \sigma \end{array} \text{] Intonationsphrase}$

(Die Morpheme des Chinesischen sind hier fest mit einer Sequenz zweier

Töne verbunden und daher von den Konventionen (6) nicht mehr betroffen.) Wenn man die Annahme macht, daß die Phonologie nur über distinktive binäre Werte verfügt, während graduelle Werte in der Phonetik-Komponente bereitgestellt werden (vgl. Pulleyblank 1986:8), dann gehören Tonsandhi-Regeln der zweiten Art zur Phonetik. Prominente Beispiele sind die besonders aus westafrikanischen Sprachen bekannten Downstep (Herabstufung eines H im Kontext von L_) und Upstep (Heraufstufung eines H im Kontext von H_).

In der Intonation nun hat man lange Zeit ein völlig anderes Phänomen als das des lexikalischen Tons gesehen, das für Akzentsprachen im Gegensatz zu Tonsprachen charakteristisch sei. Als Kategorien wurden ganzheitliche Melodien (z.B. Bolinger 1986), tonale Gitter (z.B. Gårding 1981), Akzentmuster (z.B. Thorsen 1980) oder einfach Konturtöne wie Fallend, Steigend (z.B. Kohler 1977) angenommen. Das Problem, wie diese Intonationsfiguren in prinzipieller Weise mit einer Kette von phonetischen Einheiten zu verbinden sind, blieb mehr oder weniger ungelöst. Andererseits ist aus der Musik bekannt, daß man sich eine Melodie als Abfolge einzelner (abstrakter) Tonstufen merken kann. Prinzipiell sollte also auch eine Intonationskontur als Abfolge von Tonstufen darstellbar sein, und diese lassen sich dann mit phonologischen Einheiten wie den Silben oder V-Positionen assoziieren.

Ton wie auch Intonation werden in der Tonhöhe bzw. im F_0 -Verlauf einer Äußerung manifest. Methodologisch sollte man davon ausgehen, daß gleichartige Phänomene auch mit gleichartigen Kategorien und Prinzipien erfaßt werden. Ich unterstelle deshalb, daß die Phonologie der Intonation mit denselben Kategorien und zumindest ähnlichen Prinzipien arbeiten sollte wie die Phonologie der Tonsprachen (vgl. dazu Leben 1976, Pierrehumbert 1980, Ladd 1983). Insbesondere werde ich annehmen, daß es auch bei der Intonation Floating-Töne gibt, die den Assoziationsprinzipien in (6) unterliegen, und daß es auch im Deutschen Tonsandhi-Regeln gibt. Der Unterschied zwischen lexikalischem Ton und Intonation besteht dann allein darin, daß der Ton in unterschiedlichen Domänen festgelegt wird: lexikalisch in der Wortdomäne; postlexikalisch in der Domäne des Satzes oder der Intonationsphrase.

Die Annahme der Komplementarität von Tonsprachen versus Akzentsprachen geht von den Bedingungen des Lexikons aus: entweder gibt es dort eine phonologische Tonkomponente oder nicht. Ein Kind, das eine Sprache lernt, hat also zunächst nur festzustellen, ob die Sprache, mit der es konfrontiert ist, den Ton lexikalisch ausnutzt. Dadurch steht der Ton für weitere Funktionen nicht mehr so ohne weiteres zur Verfügung, denn es gibt die universale

Beschränkung, daß lexikalisch zugewiesene Eigenschaften nicht postlexikalisch 'überschrieben' werden können; sie können nur mit einem tonalen Kontext interagieren (Tonsandhi). Jedenfalls kann Fokus in den Tonsprachen nicht dadurch festgelegt werden, daß bestimmten Silben ein weiterer Ton zugewiesen wird; allenfalls könnte sich eine Fokusdomäne indirekt, z.B. als Domäne für Tonsandhi manifestieren. Auch syntaktische Konstituenten könnten u.U. 'intonatorisch', nämlich als prosodische Domänen für Tonsandhi in Erscheinung treten.

Etwas anders ist die Situation beim Satzmodus. Hier steht einer Tonsprache prinzipiell dieselbe Option zur Verfügung, die ich eben für das Deutsche angedeutet habe: nämlich als terminaler Ton (bzw. Grenzton), der keinen silbischen Positionen, sondern den absolut terminalen Segmenten einer Intonationsphrase zuzuordnen ist und dann natürlich mit den benachbarten silbischen Tönen interagieren kann. Wenn ich oben vorgeschlagen habe, daß das Deutsche ein tonales Morphem für [+INT] besitzt, so widerspricht das also nicht der Feststellung, daß das Deutsche keine lexikalische Tonsprache ist.

Insgesamt ergibt sich also folgendes vergleichendes Bild (Abb. 2).

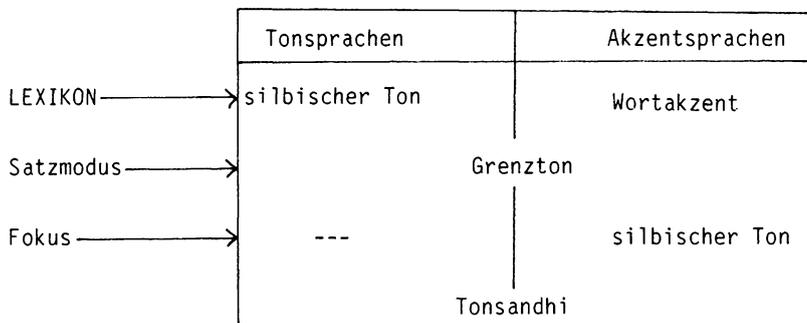
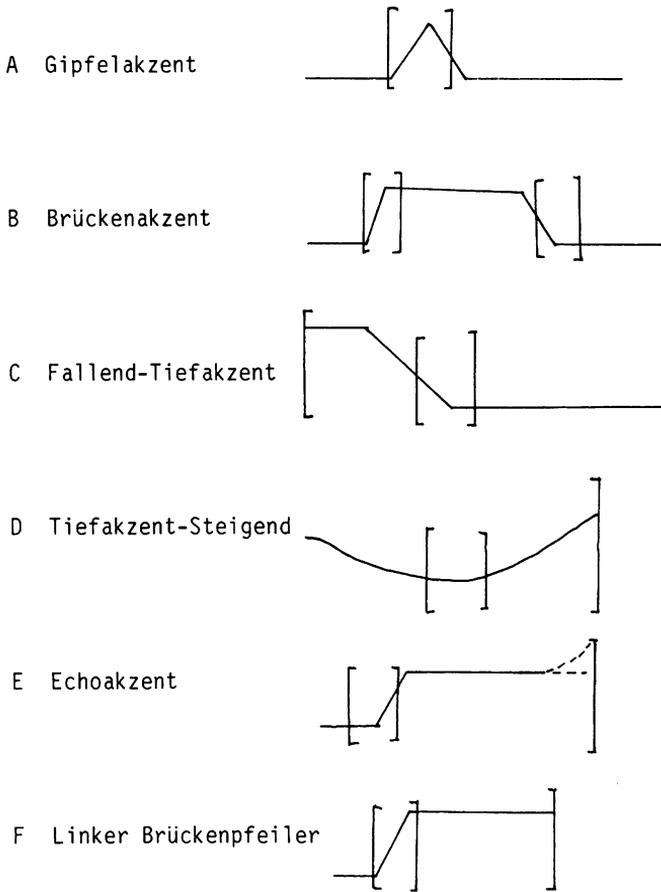


Abb. 2 Funktionspotential von Tonsprachen versus Akzentsprachen

4. Intonationsmuster des Deutschen

Unsere empirischen Beobachtungen anhand von F_0 -Sonagrammen gesprochener deutscher Sätze haben zur Unterscheidung folgender Intonations- bzw. Akzentmuster geführt. Silben, die einen Satzakkzent tragen (und somit zur Festlegung einer Fokusdomäne dienen), sind mit beidseitigen Klammern [] gekennzeichnet. Die Grenzen der Intonationsphrase sind mit einseitigen Klammern gekennzeichnet.

Abb.3 Akzentmuster



Diese Intonationsmuster sind Abstraktionen; jedes Muster kann vielfältige inter- und intraindividuelle Variation aufweisen. Entscheidend für die Identifizierung eines Musters sind der F_0 -Verlauf auf der Akzentsilbe und ihrer Umgebung sowie (gegebenenfalls) der initiale oder finale F_0 -Verlauf der Intonationsphrase. Phonetisch ist auch die Segmentstruktur der Akzentsilbe relevant: die einzelnen Segmente führen zu unterschiedlichen intrinsischen Anhebungen oder Absenkungen der F_0 ; das Segment mit der größten Sonorität (also der maximalen Intensität, in der Regel auf dem Ende des Vokals oder einem nachfolgenden sonoren Konsonanten) bildet bei einem Hoch-Akzent auch den größten F_0 -Wert der Silbe aus.

Die von uns abstrahierten Intonationsmuster sind in einer größeren Zahl von Experimenten weiter geprüft worden. Durch Resynthese haben wir mehrere Varianten eines F_0 -Verlaufs über die natürlich gegebenen Segmente einer

Äußerung gelegt. In Perzeptionsexperimenten wurde dann bestimmt, welche Silbe als die betonte wahrgenommen wird. Insbesondere konnte so zwischen Echoakzent und linkem Brückenpfeiler unterschieden werden. Beim Echoakzent wird der maximale F_0 -Wert grundsätzlich erst in der folgenden Silbe erreicht, beim linken Brückenpfeiler auf der Akzentsilbe selbst. Da die Akzentsilben gewisse Fokusdomänen anzeigen sollten, wurde außerdem geprüft, in welchen Äußerungskontexten, die zu einer bestimmten Verteilung von gegebener und nichtgegebener Information führen sollten, die resynthetisierten F_0 -Verläufe akzeptiert werden. Über diese Experimente soll hier aber nicht berichtet werden.

Die einzelnen Intonationsmuster sind in einer Intonationsphrase kombinierbar. Beschränkungen ergeben sich aus zwei Quellen: (a) Bezugnahme auf die Grenzen einer Intonationsphrase in der Charakteristik eines Musters; (b) Syntaktische Bedingungen, unter denen der Brückenakzent nur möglich ist, so daß dessen Kombinierbarkeit begrenzt ist. Eine weitere mögliche Beschränkung könnte durch das phonologische Filter $*L^*L^*$ (= Verbot zweier aufeinander folgender Tiefakzente) beschrieben werden. Dies schließt die Kombinationen CD, CE, CF, BD, BE, BF aus. Die Kombinationen mit B sind z.T. schon aus syntaktischen Gründen ausgeschlossen. Die Kombination CD scheint allerdings, wenn auch markiert (als 'ungewöhnlich klingend'), vorkommen zu können. Der Gipfelakzent A ist der einzige, bei dem ein Teil der Konfiguration, nämlich der Abfall, durch keinerlei Bedingungen an die Akzentsilbe oder die Grenze festgelegt ist. Der abfallende Zweig spielt auch für die Wahrnehmung des Akzents keine Rolle. Es könnte sein, daß es im wesentlichen der metrische Fuß (beginnend mit der Akzentsilbe) ist, der die Dauer des Abfalls festlegt. Verlängert man den Abfall im Sinne von Abb. 4 weit über den Fuß hinaus, so wird an der mit Pfeil gekennzeichneten Position ein weiterer Akzent wahrgenommen; dies ist je nach Grenzton der rechte Brückenpfeiler oder der Akzent D.

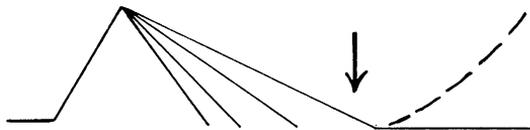


Abb. 4 Ausdehnung des Gipfelakzents

Der Brückenakzent B ist offensichtlich eine Kombination zweier auch unabhängig, und zwar in Nachbarschaft einer Grenze, vorkommender Akzente.

Wenn man die Brücke im Sinne der Abb. 5 durch eine Intonationsphrasengrenze auftrennt, so ergeben sich links der Akzent F und rechts der Akzent C.

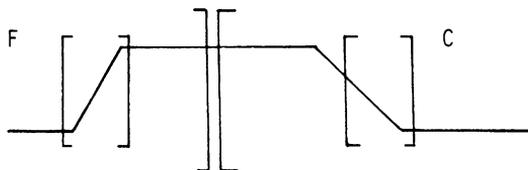


Abb. 5 Auftrennung des Brückenakzents

In phonologischer Hinsicht sind, bis auf den Gipfelakzent A, alle Intonationsmuster als Kombination aus zwei oder drei Elementen beschreibbar. Die Intonationsmuster sind also nicht die letzten Einheiten der Intonation. Daß nur diese und keine anderen Intonationsmuster zustandekommen, muß sich aus den Eigenschaften der Einheiten immanent ergeben. Die Intonationsmuster haben auch nicht zwingend eine einheitliche Funktion: ein Muster kann mehrere Funktionen haben. Es ist Aufgabe der Theorie der Intonation, dies zu klären.

Im phonologischen Teil sollte die Theorie der Intonation mit folgenden Kategorien operieren:

- Grenztöne versus Nichtgrenztöne
- Hoch (H) versus Tief (L)

Grenztöne werden im folgenden als %T (initial) oder T% (final) wiedergegeben. Sie sind mit den Grenzen einer Intonationsphrase assoziiert. Die genaue phonetische Position des initialen Grenztons ist wegen des Einschwingvorgangs der Glottis nicht beobachtbar; das Maximum eines %H wird oft erst in der zweiten Silbe der Äußerung erreicht. Der finale Grenztöne H% ist dagegen oft am absoluten stimmhaften Ende der Äußerung maximal, er ist also nicht an Sonoritätsbedingungen gebunden. Nichtgrenztöne sind entweder Töne, die fest mit den Akzentsilben assoziiert sind (T*), oder freischwebende ('floating') Töne (T), die mit allen Silben einer durch Grenze oder Akzentsilbe begrenzten Domäne zu assoziieren sind.

Die Intonationsmuster aus Abb. 3 sollen nun folgende phonologische Charakterisierung erhalten.

- A: H*
- B: H* H L*
- C: %H L*
- D: L* H%
- E: L* H (H%)
- F: H* H

Zu zeigen ist, wie diese Muster phonologisch ableitbar sind.

5. Fokuzuweisung und Fokusabbildung

Im Hinblick auf die zu realisierenden Satzakkente enthält die S-Struktur zunächst nur die folgenden Informationen.

a) Syntaktisch induzierte Fokuszeichnungen des Cleftsatzes:

(9) a. $XP \rightarrow [+F]/[{}_{VP} \underline{es} \text{ ___ } \underline{sein} [CP, +REL, AGR XP]]$

b. $XP \rightarrow [+F]/[{}_{VP} \underline{es} \text{ ___ } \underline{sein} [CP, COMP]/XP]]$

wie in den folgenden Beispielen

(10) a. weil es Peter sein mußte, den man treffen wollte.

b. obwohl es gestern war, daß/als wir ihn trafen.

AGR steht für das Kongruenzmerkmal; der Schrägstrich A/B in der Kategorienangabe bedeutet, daß die Kategorie A eine Lücke der Sorte B aufweist. F ist syntaktisches Fokusmerkmal.

b) Lexikalisch induzierte Fokuszeichnungen durch Fokuspartikeln (FP)

wie sogar, nur:

(11) $XP \rightarrow [+F]/[{}_{XP} FP \text{ ___ }]$

c) Kontrastiv induzierte Fokuszeichnungen:

(12) $[XP, +K] \rightarrow [+F]$

wie in den Beispielen

(13) a. Peter soll kommen und auch er soll kommen.

b. Peter ißt die Kartoffeln und das Fleisch Anna.

[+K] ist syntaktisches Kontrastmerkmal. Mehr dazu in § 7.

d) Wörter mit einer ausgezeichneten Wortakzent-Silbe (σ^W), die als mögliche Targets von Satzakkenten in Betracht kommen.

Offen sind nun also zwei Fragen. Erstens, welche Konstituenten sonst noch fokussiert werden können, und zweitens, wie eine fokussierte Konstituente in Beziehung zu setzen ist zu einer Akzentsilbe.

Die erste Frage ist vom isolierten Satz her gesehen sehr leicht zu beantworten: Auf den nicht schon durch (9), (11) oder (12) erfaßten Konstituenten eines Satzes kann die Fokuzuweisung frei erfolgen. Dies bedeutet dann natürlich, daß der jeweilige Fokus auch semantisch relevant wird und der Satz nur noch in bestimmten Kontexten akzeptiert werden kann. Die Konvention für die freie Fokuzuweisung lautet so:

(14) a. Indiziere jeden beliebigen Knoten der S-Struktur mit [+F].

b. Kein F-indizierter Knoten darf von einem anderen F-indizierten Knoten dominiert werden.

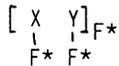
c. Mindestens ein Knoten der S-Struktur muß F-indiziert sein.

eines Teils ab. Dies soll 'Fokusabbildung' genannt werden. Sie erfolgt rekursiv, indem die syntaktische Struktur schrittweise 'von oben nach unten' bis zu den terminalen Worteinheiten abgearbeitet wird. Alle dabei gekennzeichneten Einheiten, die den Akzentträger enthalten, sollen 'F*-Phrase' genannt werden.

Welche Regeln der Fokusabbildung die adäquaten sind, kann ich noch nicht mit Gewißheit sagen. Versuchsweise formuliere ich drei solcher Regeln. Die erste (Fokuszerlegung) soll zum Brückenakzent führen, wenn eine hierarchiehöhere Satzkonstituente (vorzugsweise der ganze Satz) fokussiert ist; diese Regel wird in § 7 anhand von Kontrastdaten modifiziert bzw. verallgemeinert. Die zweite Regel (Kopfregel) soll unbeschränkt anwendbar sein, während die dritte Regel (traditionell als 'Nuclear Stress Rule' bekannt) nur als Default-Regel eintreten soll, wenn die Kopfregel kein Ergebnis liefert.

(17) Fokuszerlegung:

Verteile an die maximalen Konstituenten einer F*-Phrase, die normalerweise betonbar sind, je ein F*-Merkmal:



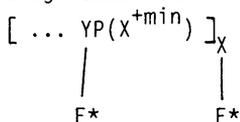
(18) Kopfregel:

Wenn der einzige Kopf einer F*-Phrase lexikalisch (d.h. minimal) ist, so vergib F* an den Nichtkopf, sonst vergib F* an den Kopf.



(19) NSR (Nuclear Stress Rule):

Vergib F* an das am weitesten rechts stehende Element einer F*-Phrase, ausgenommen den einzigen lexikalischen Kopf.



Die Kopfregel (18) kann als syntaktisches Pendant zu einem semantischen Prinzip von Selkirk (1974) angesehen werden, wonach in einer Funktor-Argument-Struktur das jeweilige Argument F* erhält; dabei ergeben sich im wesentlichen dieselben Ergebnisse wie bei Anwendung von (18). Ich sehe keine

Möglichkeit, wie ein solches semantisches Prinzip auf der Ebene der F-Struktur formuliert werden kann. Unklar ist, ob die Verteilung der O-Rollen einen Einfluß hat (vgl. v. Stechow/Uhmann 1985), dann müßte u.U. auf den Inhalt der lexikalischen Einheiten Rücksicht genommen werden.

Die Kopfregel liefert in zwei Fällen kein Ergebnis:

- a) wenn die Phrase zwei nichtlexikalische Köpfe hat (das ist bei einer Koordination der Fall);
- b) wenn die Phrase neben einem lexikalischen Kopf mehrere Nichtköpfe aufweist.

In beiden Fällen tritt die NSR (19) als Defaultregel in Kraft. Im Unterschied zur Kopfregel ist die NSR eigentlich weder syntaktisch noch semantisch begründbar, sondern nur durch die Tatsache, daß eine Intonationsphrase ihren relativen Schwerpunkt eher rechts aufweist (ein prominentes Beispiel dafür ist das Französische). Eine Variante der NSR tritt dann auch bei prosodischen Konflikten in Kraft: Wenn zwei intern als stark ausgewiesene prosodische Einheiten oberhalb des Fußes zusammentreten, so hat die rechte von ihnen Priorität. Die Regel der Fokuszerlegung ist hauptsächlich durch die Existenz des Brückenakzents auf Sätzen, die 'out of the blue' geäußert werden, motiviert. Bei Anwendung von (17) ist es nicht mehr möglich, daß eine einzelne Akzentsilbe den Fokus auf eine syntaktische Phrase projiziert. Das Projektionsprinzip kann nur bewahrt bleiben, wenn die Folge F^*F^* in dieser Konfiguration eine einheitliche intonatorische Gestalt bildet, die als Ganzes den Fokus projiziert. Der Brückenakzent stellt in der Tat eine solche Gestalt dar; sie steht aber in der Syntax nicht zur Verfügung, sondern erst in der Intonationskomponente bzw. auf der P-Struktur. Versteht man eine F^* -indizierte Einheit als F-Kopf ihrer Domäne, so ist die Domäne mit Brückenakzent als Domäne mit zwei F-Köpfen anzusehen.

Aufgrund der Regeln (17) bis (19) wird die Fokuskennzeichnung F^* bis auf die lexikalischen Einheiten einer F-Struktur abgebildet. Terminiert wird die Fokusabbildung durch die Zuordnung zu einer Wortakzentsilbe. Das wird durch (20) geleistet.

(20) Vergib F^* an die Silbe mit dem Wortakzent.

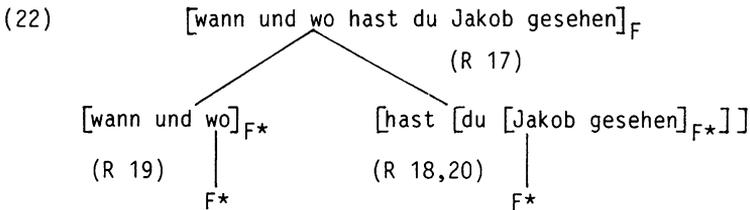
$$\left[\dots \underset{\substack{\sigma^W \\ | \\ F^*}}{\sigma^W} \dots \right]_{\chi^0} \underset{\substack{| \\ F^*}}{\sigma^W}$$

(Abkürzend soll σ^W auch als σ^* bzw. Fokussilbe erwähnt werden.)

Die Wirkung der Fokusabbildung soll an einem Beispiel illustriert werden.

(21) Wann und wo hast du Jakob gesehen?

Wenn (21) 'out of the blue' geäußert wird, wird ein Brückenakzent auf den Fokussilben wo und Ja realisiert. Dies läßt sich wie folgt ableiten:



Dabei spielen die speziellen Annahmen über die interne Syntax des rechten Astes (ob man eine mehr oder weniger 'flache' Struktur annimmt) kaum eine Rolle. Insofern sind die Regeln (17) bis (19) wünschenswert robust. Ob sie das grundsätzlich in allen denkbaren Fällen sind, muß dahingestellt bleiben.

6. Intonationskomponente des Deutschen

Die Domäne der Intonation ist nicht der Satz, sondern die Intonationsphrase. Die Intonationsphrase ergibt sich durch Abbildung einer F-Struktur auf eine prosodische Struktur (P-Struktur). Über die Einzelheiten der P-Struktur soll hier nichts ausgeführt werden; ich stelle mir eine metrische Baumstruktur, etwa im Sinne von Culicover/Rochemont (1983) vor. Jeder maximale Baum entspricht genau einer Intonationsphrase. Ein Satz ist in mehrere Intonationsphrasen zerlegbar. Die wesentliche Bedingung dabei ist, daß jede Intonationsphrase mindestens eine Fokussilbe enthalten muß.

Zu unterscheiden sind zwei Kategorien von Tönen: Grenztöne, die mit den Grenzen einer Intonationsphrase zu assoziieren sind und silbische Töne, die mit einzelnen Silben zu assoziieren sind. Eine Unterklasse der silbischen Töne sind die Akzenttöne, die den Fokussilben zugeordnet werden. In der phonologischen Intonationskomponente wird jeder Silbe ein Ton zugeordnet, gegebenenfalls durch Default.

Die erforderlichen Defaultprinzipien sind durch eine einfache Markiertheitsbetrachtung zu motivieren. Ich nehme an, daß der natürliche F_0 -Verlauf tief ist, demgegenüber lassen sich prominente Silben am einfachsten durch einen hohen F_0 -Verlauf abheben. Dies ergibt das folgende Markiertheitschema.

(23)	unmarkiert	markiert
Grenztton	L	H
Nichtakzentton	L	H
Akzentton	H	L

Die unmarkierten Werte brauchen nicht spezifiziert zu werden; sie ergeben sich durch Default. (Ein Grenztton wird dem ersten bzw. letzten Segment der Phrase zugeordnet.)

$$(24) \text{ a. } \left[\begin{array}{c} \rightarrow \\ \left[\begin{array}{c} ; \\ \downarrow \\ L \end{array} \right] \rightarrow \left[\begin{array}{c} \\ \downarrow \\ L \end{array} \right] \end{array} \right. \quad \text{b. } \sigma \rightarrow \begin{array}{c} \sigma \\ \downarrow \\ L \end{array} \quad \text{c. } \sigma^* \rightarrow \begin{array}{c} \sigma^* \\ \downarrow \\ H \end{array}$$

Wie erwähnt, nehme ich an, daß es tonale Satzmodus-Morpheme gibt, die bereits auf der S-Struktur zur Verfügung stehen. Sie spezifizieren einen markierten Grenztton und legen dadurch die Satzgrenze bereits als Grenze einer Intonationsphrase fest. Bei einer modularen Betrachtung im Rahmen von Markiertheitsüberlegungen ist es nicht erforderlich (nicht einmal wünschenswert), für alle Satzmodi ihre u.U. leeren tonalen Bedingungen zu spezifizieren. So endet z.B. der Aussagesatz mit einem Tiefton, der durch Default gegeben ist, aber auch andere Satzarten (z.B. der Imperativsatz) enden tief. Deshalb wäre es ganz verfehlt, z.B. L% als [+ASS]-Morphem anzusehen.

Genauer nehme ich an, daß es im Deutschen jedenfalls zwei tonale Satzmodus-Morpheme gibt, ein Interrogativmorphem und ein Echomorphem.

$$(25) \text{ a. } \left[\begin{array}{c} \left[\text{CP, +INT} \right] \\ \downarrow \\ H \end{array} \right. \quad \text{b. } \left[\begin{array}{c} \left[\text{CP, +ECHO} \right] \\ \downarrow \\ H \end{array} \right.$$

Das Echomorphem besteht in einem Floating-Ton, der gemäß den Assoziationsprinzipien (6, 7) mit allen Silben in einer durch * und] begrenzten Domäne zu assoziieren ist, optional auch mit der Grenze, was zu einem weiteren finalen F₀-Anstieg führt (s. unten (32)). Für die oben unter (1b) erwähnte Implikation [+ECHO] → [+INT] gibt es also ein erkennbares tonales Gegenstück: die Echo-Intonation impliziert die Interrogativ-Intonation (Diese Analyse der Echofrage ersetzt die in Wunderlich 1986).

Für die Echofrage ist (25b) das einzige Kennzeichen, beim Fragesatz ergeben sich aufgrund von (25a) evtl. Überspezifizierungen. Zwar wird ein Verberstsatz erst durch (25a) eindeutig als Fragesatz (Ja-Nein-Frage) festgelegt, für die W-Sätze besteht aber schon eine syntaktische Spezifizierung, nämlich durch (4a) von oben. Deshalb kann hier der überspezifizierende

finale H-Ton für den Ausdruck spezieller Kontextbedingungen dienen (das Fragewort bezieht sich auf Elemente, die außerhalb des kontextuell hervorgehobenen Fragebereichs liegen, oder der Fragende geht nicht davon aus, daß die Existenzpräsupposition gilt, vgl. Tropf/Wunderlich 1985). Ähnlich kann ein Verbzweitsatz mit [+ASS] (aufgrund von 4b)) durch (25a) auf den speziellen Kontext einer 'assertiven Frage' (vgl. Altmann 1987) festgelegt werden. Man beachte übrigens auch, daß (25a) für einen Satz wie (26) Willst du Kaffee oder Tee.

die Ja-Nein-Interpretation festlegt. Bei einer Alternativfrage müssen Kaffee und Tee einen syntaktischen Kontrast bilden und unterliegen deshalb anderen Interpretationsbedingungen sowie einer Kontrastgliederung, die intonatorisch zum Brückenakzent führt (vgl. § 7).

Unentschieden bin ich in der Frage, ob es neben (25) auch ein Satzmodusmorphem mit initialem H-Ton gibt. Die Fälle, die ich im Auge habe, gehören zur Kategorie [-ASS, -INT], nur ist das keine Kategorie, für die man eine formale Spezifikation erwartet. Am ehesten dient der initiale H-Grenztone als Morphem für Exclamative mit Verberst wie in Beispiel (27).

(27) Hat der wieder gelogen.

Ein phonologisch gegebener initialer H-Grenztone prägt sich phonetisch, wegen des Einschwingvorgangs der Glottis, nicht unbedingt schon am absoluten Beginn der Äußerung aus; deshalb ist hier die Interpretation der Beobachtungsdaten generell viel problematischer als am Äußerungsende.

Erst wenn satzinterne Intonationsphrasen gebildet sind (besonders um Nebensatzgrenzen zu kennzeichnen), also beim Übergang zur P-Struktur und noch nicht auf der Ebene der F-Struktur, kann eine Regel wie (28) wirksam werden.

$$(28) \] \rightarrow \] / [U \ _ \ V]_S$$

|
H

mit U, V als Variablen über nichtleeren Ketten.

Nebensätze können aber auch schon vorher durch Fokuszerlegung oder eine Kontraststruktur (siehe § 7) erfaßt werden, insofern stellt (28) nur eine Default-Regel für satzinterne Intonationsphrasengrenzen dar.

Die Spezifikation der Akzenttöne nimmt z.T. Bezug auf ihre tonale Umgebung. Daher müssen die Grenztöne zuerst zugewiesen werden. In der Tat ist es plausibel, daß eine intonatorische Gestalt wie die Intonationsphrase von ihren Rändern her festgelegt wird. Hier ergibt sich das automatisch, insoweit finale Hoch-Töne z.T. schon auf der Ebene der S-Struktur vorliegen.

Für die Tonzuweisung an Fokussilben (soweit nicht die Defaultregel (24c) in Kraft tritt) gelten nun zwei ganz unterschiedlich geartete Regeln.

$$(29) \sigma^* \rightarrow \begin{array}{c} \sigma^* \\ | \\ H \quad H \end{array} / [_ \sigma^*]_{F^*}$$

$$(30) \sigma^* \rightarrow \begin{array}{c} \sigma^* \\ | \\ L \end{array} / \left\{ \begin{array}{c} _ H \\ _ H _ \end{array} \right\} \text{ (wobei H nicht mit } \sigma^* \text{ verbunden ist)}$$

Regel (29) kennzeichnet den linken Brückenpfeiler. Der zusätzliche Floating-Ton dient zum Aufbau des Brückenplateaus. Er ist notwendigerweise an den linken Brückenpfeiler gebunden (und nicht den rechten), schon weil dieser isoliert vorkommen kann, aber auch aus unabhängigen Gründen der Links-nach-Rechts-Prozessierung. Es würde auch keinen Sinn machen, ihn von einer mittleren pivot-Position aus einzuführen (in diese Richtung gehen wohl Vorstellungen von Garding 1981), weil sich diese Position nicht morphologisch lokalisieren läßt und von ihr aus auch nicht der Ort der beiden Fokussilben bestimmt werden kann. U.U. kann man in (29) statt H*H nur einfach ein Floating-H vorsehen, falls gesichert ist, daß H mit der Fokussilbe und nach rechts, aber nicht nach links assoziiert werden kann. Die Kontextangabe in Regel (29) muß auf das Verfahren der Fokuserlegung (siehe (17)) bezugnehmen. In § 7 werde ich ein alternatives Verfahren diskutieren, dann wird auch die Kontextangabe von (29) zu modifizieren sein.

Regel (30) ist als eine spezifische Regel der Defaultregel (24c) vorgeordnet. Der in (30) verlangte H-Kontrast kann von daher auch kein Akzentton sein, sondern ist entweder ein Floating-H oder ein Grenzton. Wenn es der finale Grenzton ist, muß er sich innerhalb eines gewissen Planungsbereichs (in der Größenordnung von bis zu mehreren Silben) befinden, der sich allerdings phonologisch nicht beschreiben läßt.

Ich will nun zeigen, wie die in § 4 aufgeführten Intonationsmuster des Deutschen abgeleitet werden. Der Gipfelakzent (A) ergibt sich einfach durch Anwendung der Default-Regel (24c). Der Brückenakzent (B) entsteht so:

$$(31) \text{ a. } [\text{--- } \sigma^* \text{ --- } \sigma^* \text{ ---}]_F$$

$$\text{ b. } \begin{array}{c} \text{--- } \sigma^* \text{ --- } \sigma^* \text{ ---} \\ | \quad | \\ H \quad H \end{array} \quad (29)$$

$$\text{ c. } \begin{array}{c} \text{--- } \sigma^* \text{ --- } \sigma^* \text{ ---} \\ | \quad | \quad | \\ H \quad H \quad L \end{array} \quad (30)$$

$$\text{ d. } \begin{array}{c} \text{--- } \sigma^* \text{ --- } \sigma^* \text{ ---} \\ | \quad | \quad | \\ H \quad H \quad L \end{array} \quad \text{Assoziation (6, 7)}$$

$$\text{ e. } \begin{array}{c} \text{--- } \sigma^* \text{ --- } \sigma^* \text{ ---} \\ | \quad | \quad | \\ LLL \quad H \quad HHH \quad L \quad LLL \end{array} \quad \text{Default (24b)}$$

Die Akzente C und D beruhen auf Regel (30) unter der Voraussetzung, daß vorher ein H-Grenztön zugewiesen wurde. Der Echoakzent (E) entsteht so:

- (32) a. --- σ^* ---]_{+ECHO} (25b)
 H
 b. --- σ^* ---] (30)
 L H
 c. --- σ^* ---] Assoziation (6, 7)
 L H
 d. $\begin{matrix} \text{|||} & \sigma^* & \text{|||} \\ \text{LLL} & \text{L} & \text{HHH} \end{matrix}$] (H) Default (24b)

Die Ableitung des linken Brückenpfeilers (F) setzt die Existenz einer Intonationsphrasengrenze innerhalb einer $[F^*F^*]_{F^*}$ -Domäne voraus. Dafür kann es mehrere Gründe geben. Erstens könnte es ein, daß die Zerlegung in Intonationsphrasen keine Rücksicht zu nehmen braucht auf die Existenz von $[F^*F^*]$ -Domänen, in diesem Fall müßte sie der Anwendung von Regel (29) nachgeordnet sein. Dabei muß dann der Floating H-Ton auch mit den entstehenden Grenzen assoziiert werden.

- (33) $\begin{matrix} \text{---} \sigma^* \text{---} \text{IP} \\ | \quad | \\ \text{H} \quad \text{H} \end{matrix} \text{---} \text{IP} \text{---} \sigma^* \text{---}]_F$

Fokuserlegung erfolgt auf der S-Struktur, Intonationsphrasenzerlegung durch Abbildung auf eine P-Struktur. Die betreffenden Domänen brauchen nicht zu kongruieren, wenngleich Fall (33) wohl als der markierte Fall anzusehen ist, weil metrische und intonatorische Gliederung divergieren (s. unten die Diskussion zu (42a)). Dafür spricht auch, daß Versuchspersonen in der Lage sind, eine Intonationskontur wie in (33) zu erzeugen, aber in der Regel nur, wenn sie vorher angewiesen sind, eine bestimmte Gliederung durch Pausen kenntlich zu machen.

Zweitens gibt es Kontrastverknüpfungen, auch über Satzgrenzen hinweg (so daß eine Trennung in zwei Intonationsphrasen geradezu erwartbar ist), bei denen eine $[F^*F^*]_{F^*}$ -Domäne der angenommenen Art durch den Kontrast hergestellt wird. (34) gibt ein Beispiel.

- (34) $\begin{matrix} \text{---} & / & \text{---} \\ & & \backslash \\ \text{Peter} & & \text{Anna} \end{matrix}$ kommt, Anna geht.

Auf die Repräsentation solcher Kontraste werde ich im folgenden § 7 eingehen.

Vor allem kommt der linke Brückenpfeiler bei iterativer Fokuszerlegung vor. Man denke sich eine zweifache Anwendung der Fokuszerlegung.

- (35)
- a. $[[[F^*F^*]_{F^*} F^*]_{F^*}]_{F^*}$
 - b. $[F^* [F^*F^*]_{F^*}]_{F^*}$

Der Typ der Linkszerlegung in (35a) würde zu einer Tonzuweisung wie in (36a) führen; sie stellt keine einheitliche Gestalt mehr dar, die auf die Syntax 'zurückzuprojizieren' wäre; so etwas kommt daher nicht vor.

- (36)
- a. $\begin{array}{ccc} \sigma^* & \sigma^* & \sigma^* \\ | & | & | \\ H & H & L & H \end{array}$
 - b. $\begin{array}{ccc} \sigma^* & \sigma^* & \sigma^* \\ | & | & | \\ H & H & H & H & L \end{array}$

Der Typ der Rechtszerlegung in (35b) führt zu (36b), der zweite Akzentton würde sich aber in einer linken und rechten H-Umgebung nicht mehr hervorheben; daher muß durch dissimilatorisches Tonsandhi eine geeignete Umgebung hergestellt werden. Diese Tonsandhi-Regel ist in (37) angeführt.

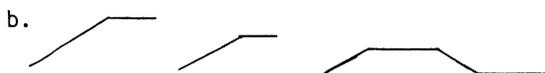
- (37) $\emptyset \rightarrow L/H \text{ — } \begin{array}{c} H \\ | \\ \sigma^* \end{array}$

Im Ergebnis von (37) sind die Floating-Töne H und L benachbart und werden dann wohl auf zwei verschiedene prosodische Domänen verteilt. Insbesondere ist es möglich, sie auf zwei Intonationsphrasen zu verteilen. Dadurch wird (36b) in (38) übergeführt.

- (38)
- $$\begin{array}{ccc} \sigma^* &] [& \sigma^* & \sigma^* \\ | & & | & | \\ H & H & L & H & H & L \end{array}$$

Ein typisches Beispiel für iterative Anwendung der Rechtszerlegung ist die Listenintonation.

- (39) a. dreizehn, vierzehn, fünfzehn, sechzehn. (Ende)



(39b) illustriert einen charakteristischen F_0 -Verlauf für die Listenintonation. Man erkennt hier auch, daß in der phonetischen Komponente des Deutschen in solchen Fällen Downstep die Regel ist.

Unsere Theorie macht eine spezifische Vorhersage über die unterschiedliche intonatorische Gliederung von Klammerungen wie in (40a, b) bzw. über die intonatorische Disambiguierung von Sätzen wie in (41).

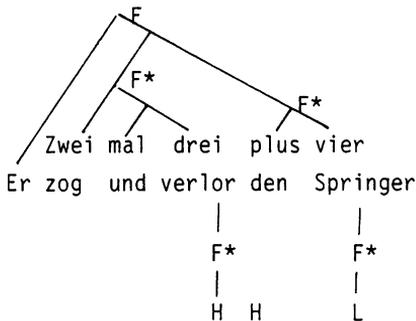
(40) a. $(2 \times 3) + 4$

b. $2 \times (3 + 4)$

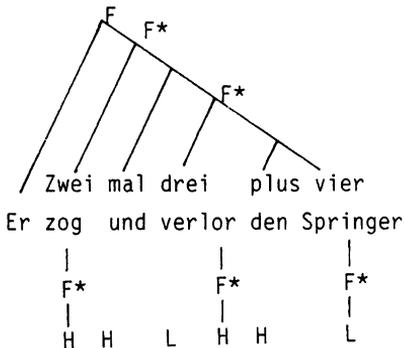
(41) Er zog und verlor den Springer.

Angenommen wird, daß die Äußerungen maximalen Fokus erhalten. Nach den Regeln (17)-(19), (29), (30), (37) ergeben sich dann die folgenden Strukturen.

(42)



(43)



(42) sollte intonatorisch mit einer Brücke realisiert werden. Dabei kann die Brücke eine eigene metrische Einheit bilden, die dann mit der syntaktischen Gliederung divergiert. Oder aber die metrische Gliederung entspricht der syntaktischen und divergiert mit der intonatorischen, wobei die Brücke im Sinne von (33) auf zwei verschiedene metrische Einheiten (z.B. Intonationsphrasen) verteilt wird. Einzig abweichend von der Vorhersage realisieren die Sprecher in der Regel außerdem einen normalen Gipfelakzent auf zwei bzw. zog. Hierbei handelt es sich möglicherweise um einen engen Fokus innerhalb des weiten Fokus, was unsere Theorie aber gemäß (14b) verbietet. Vielleicht ist aber die Konzeption der Fokuszerlegung so modifizierbar, daß in diesem Punkt eine korrekte Vorhersage möglich ist (siehe § 7).

Bei einer syntaktischen Rechtseinbettung wie in (43) sollte eine Variante der Listenintonation entstehen, die mit einem linken Brückenpfeiler beginnt und dann mit einer Brücke fortsetzt. In diesem Fall gibt es keine Divergenzen zwischen syntaktischer, metrischer und intonatorischer Gliederung. (Dies scheint mir ein schöner Beleg dafür, daß das Deutsche 'natürlicherweise' als eine rechtseinbettende Sprache anzusehen ist.) Es gibt auch keine Divergenzen zwischen Vorhersage und intonatorischer Realisierung durch Versuchspersonen. Hinzuzufügen ist allerdings, daß sich sowohl (40a, b) wie auch (41) noch mit anderer Intonation realisieren lassen. Z.B. läßt sich bei der syntaktischen Struktur von (43) auch eine Brücke zwischen zwei/zog und vier/ Springer aufbauen (nicht aber bei der Struktur von (42)), und ebenso ist ein enger Fokus möglich. Alle hierbei entstehenden Varianten sind aber von unserer Theorie auch vorhersagbar.

Die Diskussion anhand der Beispiele (40)-(43) dürfte hinreichend verdeutlicht haben, daß die Art und Verteilung der Akzenttöne Auskunft über die syntaktische Gliederung eines Satzes geben kann. Insbesondere lassen sich rechtseinbettende Strukturen von anderen Strukturen unterscheiden. Insofern besteht eine wichtige Funktion der Intonation darin, syntaktische Gliederungen perzeptuell zugänglich zu machen.

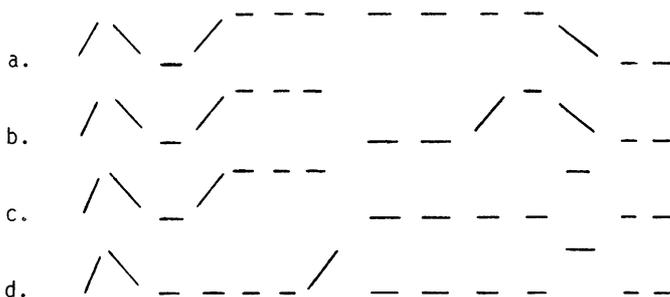
7. Syntaktische und prosodische Kontraste

Im Zusammenhang mit Kontrastphänomenen ergeben sich Daten, die die bisherige Theorie nicht zu erklären vermag. Diese Phänomene gehen z.T. über die Domäne des Satzes hinaus. Intonatorisch ergeben sich oft Brückenakzente, die sich nur mit problematischen Annahmen über die Regel der Fokuszerlegung (Regel (17)) begründen lassen. Deshalb soll nach einem alternativen Prinzip für die Ableitung des Brückenakzents gesucht werden, das sich dann evtl. auch auf die bisher über Fokuszerlegung behandelten Fälle ausdehnen läßt. Dazu werden zunächst einige Beispiele betrachtet.

Die Figuren in (45) geben in idealisierter Weise typische Intonationsverläufe einer Äußerung von (44) wieder. Jeder Strich steht für eine Silbe; man beachte, daß der Intonationsverlauf des ersten Teilsatzes in a bis c derselbe ist.

(44) Peter hat Anna geküßt und dann hat er sie geküßt.

(45)



(Hier wie im folgenden ist die Anzahl der möglichen Intonationsverläufe nicht unbedingt erschöpfend aufgeführt. Für die Wahl einer der Intonationen sind teils der globale Kontext, teils intern prosodische Präferenzen ausschlaggebend; sie sollen hier nicht erörtert werden.)

Interessant ist nun, daß in (44) der Kontrast erst durch den zweiten Teilsatz ausgelöst wird. Eine koordinative Verknüpfung ist nur dann wohlgeformt, wenn die Glieder der Verknüpfung semantisch differieren, in (44) muß also mindestens eines der Pronomina referentiell mit dem vorhergehenden Eigennamen kontrastieren, im Beispiel soll das sie sein. Diese Kontrastforderung beruht also nicht auf einem syntaktischen Prinzip (syntaktisch ist (44) wohlgeformt), aber sie hat Konsequenzen für die semantische Deutung und die intonatorische Realisierung. Im Rahmen der Grammatikkonzeption von § 2 ist es sinnvoll anzunehmen, daß die Kontrastforderung auf der S-Struktur operiert, bevor dort die Regel der freien Fokuzuweisung Anwendung findet. Sie gehört zu jenen Verfahren, die den bei koordinativen Verknüpfungen notwendigen Parallelismus regulieren. Lang (1977, 1987) hat überzeugend ausgeführt, daß es sich um Prinzipien einer sekundären Strukturbildung handelt, die man also einer unabhängigen Sprachkomponente zuweisen sollte. Obwohl ich im folgenden einen etwas anderen Phänomenumfang im Auge habe als Lang, wenn er Parallelismus diskutiert, denke ich, daß zumindest ein Teil dieser Komponente durch die Abbildung von einer S-Struktur auf eine K-Struktur erfaßt werden kann (mit K für 'Kontrast'). Hierbei interessieren natürlich vor allem die prosodischen Realisierungen eines Kontrastes. Hervorzuheben ist, daß sich Kontraste nicht auf die Domäne des Satzes beschränken, sondern mindestens auch Paare von Sätzen erfassen. Auf der Ebene der K-Struktur werden die Sätze zu einer prosodisch relevanten Einheit zusammengeführt.

Für das Beispiel (44) genügt es zunächst anzunehmen, daß sie das Merkmal [+K] für Kontrast erhält, das im Sinne von § 5 ein fokusinduzierendes Merkmal ist (s. Regel (12)). Das syntaktische Gegenstück zu sie (mit dem sie referentiell kontrastieren soll) ist Anna. Wie (45d) zeigt, wird das syntaktische Gegenstück nicht notwendigerweise auch fokussiert; deshalb soll es nur das semantisch relevante Merkmal [+C] erhalten. Im folgenden müssen nun die prosodisch kontrastierenden Konstituenten K_1 und K_2 gefunden werden, wobei $K_1 < K_2$ (im Sinne linearer Vorgängerschaft). Dabei gelten die folgenden Bedingungen:

(46) a. [+C] ist in K_1 eingeschlossen,
 [+K] ist in K_2 eingeschlossen

b. [+K] ist eine K_i -Konstituente mit $i = 1, 2$.

(46a) stellt sicher, daß sich die syntaktischen Gegenstücke in verschiedenen Kontrastphrasen befinden. (46b) stellt sicher, daß das [+K]-Element auf jeden Fall kontrastiven Fokus erhält.

Für die prosodische Kontrastgliederung gibt es dann die folgenden Optionen (die zu den Intonationsverläufen von (45) führen).

(a) [+C] und [+K] kontrastieren direkt miteinander, d.h. $K_1 = [\text{Anna}]$, $K_2 = [\text{sie}]$.

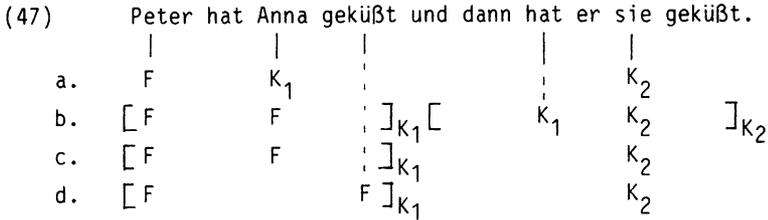
(b) [+C] und [+K] kontrastieren indirekt miteinander, d.h. $K_1 = [\text{Peter hat Anna geküßt}]$, $K_2 = [\text{und dann hat er sie geküßt}]$. Wegen Bedingung (46b) muß K_2 in einen weiteren Kontrast zerlegt werden. Diese Zerlegung entspricht im wesentlichen der Fokuszerlegung nach (17), bei der aber schon eines der F*-Elemente festgelegt ist, nämlich [sie]. Das Pronomen er ist (bei vorausgesetzter Referenzidentität mit Peter) nicht fokussierbar, aber das Finitum ist als Träger der Assertion fokussierbar (vgl. Höhle 1986 über Verum-Fokus). Ich betrachte daher $K_1 = [\text{hat}]$ und $K_2 = [\text{sie}]$ wieder als mögliche Kontrastkonstituenten. (Das erste Element eines jeden Kontrastes wird als " K_1 ", das zweite als " K_2 " angeführt.)

Innerhalb des ersten Teilsatzes (= K_1) ist die Fokuzuweisung frei, aber es muß mindesten einen Fokus geben.

(c)(d) [+K] kontrastiert mit dem ersten Teilsatz, d.h.

$K_1 = [\text{Peter hat Anna geküßt}]$, $K_2 = [\text{sie}]$. Die Fokuzuweisung in K_1 ist wieder frei, (c) und (d) repräsentieren verschiedene Optionen dafür.

In (47) sind diese verschiedenen Möglichkeiten der prosodischen Kontrastauszeichnung zusammengefaßt.



Alle minimalen K_1 -Konstituenten sind nun Fokusträger.

Im Vergleich zu der in § 5 diskutierten Fokuzuweisung sind nun zwei Deutungen unseres Vorgehens möglich.

a) Das Vorkommen eines [+K]-Elements induziert eine Zerlegung in Kontrastphrasen. Danach ist die Fokuzuweisung nicht mehr auf Sätze, sondern auf die entstandenen Kontrastphrasen anzuwenden. An die Stelle von (14) tritt (48).

(48) a. Indiziere jeden beliebigen Knoten einer K_1 -Konstituente mit F.

b. Kein F-indizierter Knoten darf von einem anderen F-indizierten Knoten dominiert werden.

c. Mindestens ein Knoten einer K_1 -Konstituente muß mit F indiziert werden.

Im Ergebnis führt das dazu, daß in (47) genau alle minimalen K_1 -Konstituenten auch F-indiziert werden. Mir ist allerdings nicht klar, ob es wirklich adäquat ist, hier überall engen Fokus anzunehmen.

b) Man betrachtet entweder die ganze Satzverknüpfung oder nur den ersten Teilsatz in einem fokusinduzierenden globalen Kontext. Ersteres würde für (45a), letzteres für (45b-d) anzunehmen sein. Dies könnte z.B. plausibel machen, warum nur im ersten Teilsatz freie Fokusindizierung stattfindet; (45d) verlangt z.B. einen erkennbar anderen Kontext als (45c). Bei (45b-d) würde sodann der erste Teilsatz als fokusinduzierender Kontext für den zweiten Teilsatz anzusehen sein, was im ersten Teilsatz intonatorisch durch den rechten Brückenpfeiler ausgedrückt wird. Im übrigen gelten aber die Vorstellungen des § 5, so daß insbesondere die Zerlegung des zweiten Teilsatzes in (45b) auf Fokuszerlegung zurückgeht.

Nach der Deutung in a) geht die Kontrastphrasenzerlegung der Fokuzuweisung voraus, nach b) gibt es eine Interaktion zwischen beiden. Eine Entscheidung zwischen diesen Deutungen muß ich hier offenlassen. Denkbar ist,

die Fokuszerlegung generell als eine Art von prosodischer Kontrastbildung anzusehen, dann wäre Deutung b) wohl vorzuziehen.

Um auf der Basis der Konfigurationen in (47) die Intonationskonturen in (45) abzuleiten, braucht man die Regel (29) aus § 6 nur durch Regel (49) zu ersetzen.

$$(29) \quad \sigma^* \rightarrow \begin{array}{c} \sigma^* \\ | \\ H \quad H \end{array} \quad / \quad [_ \sigma^*]_{F^*}$$

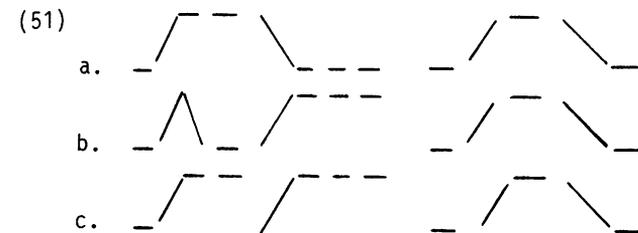
$$(49) \quad \sigma^* \rightarrow \begin{array}{c} \sigma^* \\ | \\ H \quad H \end{array} \quad / \quad _]_{K_1}$$

Die Regeln unterscheiden sich nur in der Kontextangabe: in (49) ist die letzte F*-Silbe einer Konstituente betroffen, in (29) die erste aus einer Folge von F*-Silben. (29) würde also weitaus mehr Voraussicht bei der Tonzuweisung erfordern als (49). Aus Lokalitätsgründen ist deshalb (49) die vorzuziehende Formulierung für den Brückenakzent.

Für die intonatorische Realisierung der F*-Silbe von K_2 braucht keine besondere Festlegung getroffen zu werden. Entweder steht die Silbe im Kontext eines Floating-H und erhält L oder sie ist durch eine Intonationsphrasengrenze mit %L gegenüber diesem H geschützt und erhält H*. Letzteres wäre der Fall bei sie in (45c, d). Dabei muß nur vorausgesetzt werden, daß maximale K_1 -Konstituenten Kandidaten für eigene Intonationsphrasen sind. Festzuhalten ist, daß Intonationskonturen wie in (45c, d) von der in den §§ 5 und 6 dargestellten Theorie nicht zugelassen wurden, während sie jetzt ableitbar sind.

Als weiteres Beispiel soll (50) betrachtet werden mit den idealisierten Intonationskonturen in (51) Hier sind jeweils die zweiten Teile der Kontur identisch.

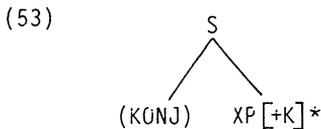
(50) Maria hat Anna besucht und Peter Heinrich.



Ich nehme an, daß Gapping-Sätze basisgeneriert sind und erst in der Kontrastkomponente auf Wohlgeformtheit geprüft werden. Eine Kette wie und Peter Heinrich kann nicht von sich aus auf syntaktische Wohlgeformtheit beurteilt werden, sondern nur in einem Kontrast-Kontext, der dann auch zur Rekonstruktion bei der semantischen Deutung herangezogen wird. In Kontexten wie (52) würde sich die Kette und Peter Heinrich als ungrammatisch erweisen.

- (52) a. Maria hat an Anna geschrieben - - -
- b. Maria hat geschlafen - - -

Als Wohlgeformtheitsbedingung für Gappingsätze (die im übrigen auch als Antwort auftreten, z.B. auf die Frage "Wer hat wen besucht?") soll (53) gelten. Der Kleene-Stern * steht hier für 1 bis n-faches Vorkommen.



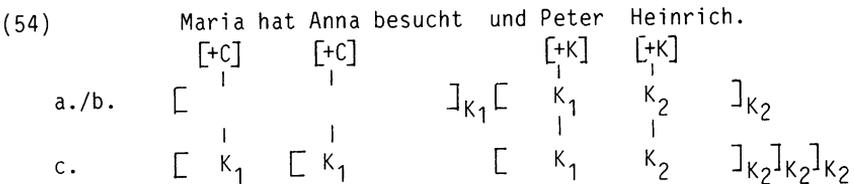
Wenn unter X nur lexikalische Hauptkategorien verstanden werden, dann ist mit (53) auch die bekannte Tatsache ausgedrückt, daß Gapping-Konstruktionen kein Finitum (= INFL) enthalten.

Für die Ableitung möglicher Kontraststrukturen von (50) betrachten wir den Satz mit seinen syntaktischen Kontrastmerkmalen.

(54) Maria hat Anna besucht und Peter Heinrich.



Der syntaktische Kontrast ist durch konfigurationellen Parallelismus definiert und führt dadurch zu einer Überschneidung der Kontrastpaare. Das ist eine Kontraststruktur, für die es keine Projektion auf die lineare Kette gibt. Die prosodische Kontrastorganisation muß nun dazu dienen, eine solche Projektionsmöglichkeit herzustellen. Dafür sind die oben erwähnten Bedingungen (46) maßgeblich. Sie erlauben die folgenden Möglichkeiten.



In (54a/b) kann eine weitere Option für den ersten Teilsatz getroffen werden. Entweder wird der syntaktische Parallelismus wiedergegeben: dann werden die [+C]-Glieder in K₁ genau so miteinander kontrastiert wie das für die

(58) Peter schenkt ihm häufiger ein Buch als ihr.

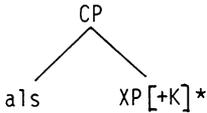


(59) Ich sehe ihn häufiger als Peter Heinrich.



Die als-Phrase ist legitimiert, wenn die Kontrastbedingung (60) erfüllt wird - ein offensichtliches Analogon zu (53).

(60)



Im Unterschied zur gleichgeordneten koordinativen Verknüpfung ist die als-Phrase Teil einer untergeordneten AP (und daher weniger leicht in einer eigenen Intonationsphrase abtrennbar; eine parallele Kontur wie (51a) ist für (59) kaum zu erwarten). Andererseits gehört das komparative Adjektiv immer zu K_1 (also zu dem Teil des Satzes, mit dem die als-Phrase kontrastiert wird); formal zeigt sich das an der möglichen Extraktion der als-Phrase wie in (58), und semantisch gehört das komparative Adjektiv zu dem Prädikat, das den Rekonstruktionshintergrund für die Bedeutung der als-Phrase bildet. (vgl. von Stechow 1985, Bierwisch 1987). Daher ergeben sich beim Aufbau der Kontraststruktur prinzipiell keine anderen Umstände als bei der koordinativen Verknüpfung. Zu bemerken ist, daß in den Beispielkonturen (58, 59) die freie Fokuszueweisung in K_1 auf das Adjektiv fällt, schon alleine deswegen, weil Pronomina (selbst wenn sie mit [+C] ausgezeichnet sind) normalerweise nicht Fokusträger sind.

Ich bin davon ausgegangen, daß auf der lexikalisch belegten S-Struktur eine Kontrastkomponente wirksam wird, die u.a. folgende Aufgaben erfüllt:

A. Überprüfung der Wohlgeformtheit von Gappingkonstruktionen durch Identifizierung passender syntaktischer Gegenstücke; dadurch Bereitstellung der Eingangsbedingungen für die semantische Rekonstruktion von Gapping.

B. Ableitung von prosodischen Kontraststrukturen, die der folgenden Fokuszueweisung dienen; dadurch Bereitstellung der Eingangsbedingungen für die metrische Komponente (u.a. Zerlegung in Intonationsphrasen) und die Intonationskomponente (Tonzuweisung).

Unter syntaktisch-semantischem Gesichtspunkt steht der Aufgabenkomplex A natürlich im Vordergrund. Es hat sich nun gezeigt, daß eine enge Beziehung zum Aufgabenkomplex B besteht. Offen ist die Frage, ob der Aufgabenkomplex B auch dann eine Rolle spielt, wenn Aufgaben von der Art A gar nicht anfal-

len. Im folgenden will ich in dieser Richtung argumentieren, und zwar im Hinblick auf drei Arten von Phänomenen:

- a) 'out of the blue' - Äußerungen,
- b) syntaktische Extraktionen,
- c) syntaktische Klammerungen.

'Out of the blue' - Äußerungen wurden bisher mit maximalem Fokus gekennzeichnet, der dann der Fokuserlegung unterliegt (Regel (17)). Wie schon verschiedentlich hervorgehoben, sind Fokuserlegung und Kontrastphrasenzerlegung nahezu identische Verfahren, die intonatorisch im Brückenakzent münden. Die Kontrastphrasenzerlegung ist aber überlegen, indem sie Fälle erfaßt, die der Fokuserlegung nicht zugänglich waren, und eine einfachere Regel für den linken-Brückenpfeiler-Akzent erlaubt (Regel (49)). Deshalb ist zu prüfen, ob man die Fokuserlegung nicht einfach als Spezialfall der Kontrastphrasenzerlegung ansehen kann.

Bei einem Satz mit engem Fokus besteht ein impliziter Kontrast zwischen dem fokusetablierenden Kontext und diesem Satz. Bei einer völlig neuartigen Äußerung wie z.B. (61) besteht aber keine solche Thema-Fokus-Relation.

(61) (Weißt du schon das Neueste:)

Der Papst fährt Ski.

Hier werden lediglich Subjekt und Prädikat in Beziehung gesetzt. Es ist nicht unplausibel, dies als eine Möglichkeit interner Kontrastierung anzusehen, die dann auch ihren intonatorischen Ausdruck in Gestalt des Brückenakzentes findet. Nicht von ungefähr bedienen sich typischerweise selbständige Sätze wie Sprichwörter oft auch eines internen (lexikalischen oder konstruktionellen) Parallelismus.

(62) Erlaubt ist, was gefällt.

Schenken heißt angeln.

Wie du mir, so ich dir.

Vorrede spart Nachrede.

Ich will deshalb die Regel der Fokuserlegung (Regel (17)) durch eine Regel der Kontrastphrasenzerlegung ersetzen.

(63) Zerlege einen selbständigen F-indizierten Satz in seine maximalen normalerweise betonbaren K-Konstituenten.

$$[\quad]_{S[+F]} \rightarrow [K_1 K_2]$$

Bei untergeordneten Konstituenten sollte die Kontrastphrasenzerlegung immer durch andere Umstände bewirkt werden: z.B. untergeordnete Koordination oder induzierter Parallelismus. Die formalen Bedingungen dafür will ich nicht versuchen zu formulieren.

Wir können hier aber erneut das Beispiel (21) aus § 5 betrachten, wo sich die topikalisierte W-Phrase mit dem Rest des Satzes kontrastieren läßt. Als koordinierte Phrase erlaubt [wann und wo] eine erneute Kontrastzerlegung, und dies induziert dann in der Regel einen parallelen Kontrast im Rest des Satzes. Es ergeben sich also mindestens die folgenden Möglichkeiten.

- (64) Wann und wo hast du Jakob gesehen
- a. [| |]_{K₁} [| |]_{K₂}
- b. [| |]_{K₁} [| |]_{K₁} [| |]_{K₂} [| |]_{K₂}

Bei syntaktischen Extraktionen entsteht eine semantisch relevante Beziehung zwischen dem Zielort und dem Ausgangsort. In der Rektions- und Bindungstheorie der Syntax ('Government and Binding') nimmt man an, daß am Ausgangsort eine Spur (als leere Kategorie) verbleibt. Sie dient zur Rekonstruktion der semantischen Zusammenhänge. Im allgemeinen reichen aber auch schon lexikalische Eigenschaften der beteiligten Einheiten dazu aus. Dies ist jedoch nicht mehr unbedingt der Fall bei 'long distance' - Abhängigkeiten, besonders wenn hierbei Ambiguitäten entstehen. Es ist nun denkbar, daß zwischen extrahierter Phrase und der jeweiligen Domäne, der sie entstammt, ein Kontrast hergestellt wird, der dann sowohl semantisch wie auch intonatorisch berücksichtigt wird. Dies wäre eine Möglichkeit, den Ort der Spur phonetisch, wenn auch indirekt, zu kennzeichnen.

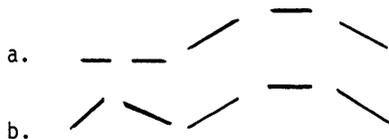
Zur Prüfung dieser Hypothese haben wir den gleich dreifach ambigen Satz (65) verwendet: wann kann hier einer der drei angezeigten Domänen zugeordnet werden.

(65) Wann [glaubst du [daß Peter gesagt hat [daß Käte kommen wird]]]

Durch Resynthese haben wir einer natürlichen Äußerung des Satzes verschiedene Intonationskonturen aufgeprägt und dann in einem Perzeptionsexperiment die Versuchspersonen angewiesen, wann einer der drei Domänen zuzuordnen. Die Resultate mit größter Übereinstimmung ergaben sich, wenn wann zusammen mit einem der drei Hauptverben einen intonatorischen Brückenakzent bildete. (Man beachte, daß das Adverbial wann als Modifikator einer VP anzusehen ist und für das Verb nicht so leicht ein extern bedingter Fokus angenommen werden kann wie für das jeweilige Subjekt.) Dies darf als Bestätigung dafür angesehen werden, daß Kontraststrukturen eine (wenn auch marginale) Rolle bei Extraktionsphänomenen spielen.

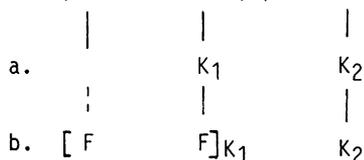
Zum Schluß ist nochmals auf die syntaktischen Klammerungen zurückzukommen, die in § 6 als rekursive Fokuszerlegungen diskutiert wurden. Dort wurde auch festgestellt, daß das Intonationsmuster in (66b) mit einem Gipfelakzent auf zwei nicht ableitbar war, sondern nur (66a).

(66) (zwei mal drei) plus vier



Nun ist ja völlig unklar, ob bei einer Äußerung von "(2 x 3) + 4" das Ganze im Fokus steht oder jede einzelne Zahl im Fokus steht, verwertet werden kann nur die syntaktische Gliederung im Gegensatz zu "2 x (3 + 4)". Es ist daher sinnvoll anzunehmen, daß die syntaktische Gliederung auf die Zerlegung in Kontrastphrasen abgebildet wird. Man kann dann sagen, daß mit (66a) ein positioneller Kontrast angezielt wird: die erste Operation endet mit drei und die zweite Operation endet mit vier, während mit (66b) ein Phrasenkontrast angezielt wird: die erste Phrase enthält zwei und drei und die zweite Phrase enthält dazu noch vier. (67) gibt die beiden Möglichkeiten in der bisher verwendeten Notation wieder.

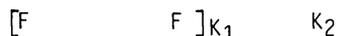
(67) (zwei mal drei) plus vier



Unsere Regeln der Tonzuweisung, insbesondere Regel (49), führen dann zu den beiden Konturen von (66).

Für das andere Beispiel aus § 6, hier als (68) wiedergegeben, macht nur der Phrasenkontrast einen Sinn.

(68) Er [zog und verlor] den Springer



Das direkte Objekt ist hier mit beiden Verben zu kontrastieren; dabei ist es unerheblich, ob die Verben primären Fokus (F) oder abgeleiteten Fokus (F*) erhalten. Nochmals hervorzuheben ist aber, daß semantisch gleichberechtigte Foki einen intonatorisch verschiedenen Akzent erhalten. Dies ist der Preis, der bei der Abbildung einer hierarchischen Struktur auf die lineare Dimension des Sprechereignisses zu zahlen ist.

Wer Spaß daran hat, den Mechanismus dieser Abbildung weiter zu ergründen, mag sich an einem zugegeben forcierten 4-Term-Problem versuchen. Ich behaupte, daß sich die in (69) bis (72) wiedergegebenen Klammerungen rein intonatorisch disambiguieren lassen (ganz abgesehen davon, daß man zusätzliche Pausen benutzen kann). Die dafür verantwortlichen Kontraststrukturen habe ich aufgeschrieben. Dabei ist in (a) jeweils die positionelle Kontraststruktur und in (b) die Phrasen-Kontraststruktur angegeben. (Bei der intonatorischen Unterscheidung von Gipfelakzent und linkem Brückenpfeiler ist es vielleicht nützlich, die Einsilber durch Zweisilber, z.B. 13, 14, 15, 16, zu ersetzen.)

$$\begin{array}{l}
 (69) \quad (2 \times 3) + (4 \times 5) \\
 \quad \quad | \quad \quad | \quad \quad | \quad \quad | \\
 \text{a.} \quad \cdot \quad \quad K_1 \quad \quad \cdot \quad \quad K_2 \\
 \quad \quad \vdots \quad \quad | \quad \quad \vdots \quad \quad | \\
 \text{b.} \quad [K_1 \quad K_2]_{K_1} \quad [K_1 \quad K_2]_{K_2}
 \end{array}
 \qquad
 \begin{array}{l}
 (70) \quad 2 \times (3 + 4) \times 5 \\
 \quad \quad | \quad \quad | \quad \quad | \quad \quad | \\
 \text{a.} \quad K_1 \quad \quad [\quad \quad K_1 \quad \quad K_2]_{K_2} \\
 \quad \quad | \quad \quad \cdot \quad \quad / \quad \quad | \\
 \text{b.} \quad K_1 \quad \quad [[F \quad F]_{K_1} \quad K_2]_{K_2}
 \end{array}$$

$$\begin{array}{l}
 (71) \quad 2 \times (3 + (4 \times 5))) \\
 \quad \quad | \quad \quad | \quad \quad | \quad \quad | \\
 \text{a.} \quad K_1 \quad \quad [K_1 \quad \cdot \quad K_2]_{K_2} \\
 \quad \quad | \quad \quad | \quad \quad \vdots \quad \quad | \\
 \text{b.} \quad K_1 \quad \quad [K_1 \quad [K_1 \quad K_2]_{K_2}]_{K_2}
 \end{array}
 \qquad
 \begin{array}{l}
 (72) \quad ((2 \times 3) + 4) \times 5 \\
 \quad \quad | \quad \quad | \quad \quad | \quad \quad | \\
 \text{a.} \quad \cdot \quad \quad K_1 \quad \quad [K_1 \quad K_2]_{K_2} \\
 \quad \quad \vdots \quad \quad / \quad \quad | \quad \quad | \\
 \text{b.} \quad (F \quad F)_{K_1} \quad [K_1 \quad K_2]_{K_2}
 \end{array}$$

Besonders aufschlußreich ist der Unterschied von (71) und (72). In der b-Variante ist es allein die Ausprägung des ersten Akzents (linker Brückenpfeiler bei (71), Gipfel bei (72)), der erkennen läßt, ob eine rechtsrekursive oder linksrekursive Klammerung vorliegt. Sozusagen legt die erste Tonfolge das Musikstück fest.

8. Phonetik der Intonation

Am Eingang der Phonetik-Komponente haben wir zwei Arten von F_0 -relevanten Informationen:

- eine prosodische (P-)Struktur mit stark-schwach-Kennzeichnung der relativen Prominenz der Konstituenten, bis hinab zur Silbe;
- zu jeder Silbe einen Tonwert H oder L.

Zunächst wird die P-Struktur abgebaut, indem die stark-schwach-Verteilung auf evtl. zusätzliche Gewichte der prosodischen Einheiten verteilt wird.

Dann wirken metrische Euphonie-Regeln über den Silben, die Akzentzusammenstöße oder rhythmische Lücken auflösen, indem sie die Gewichte evtl. anders verteilen (vgl. Féry 1986). Schließlich kommen Emphase-Regeln zur Geltung, die einigen Wörtern oder Silben ein zusätzliches Gewicht verleihen.

Alles in allem erhalten die Silben der Äußerung ihr abschließendes Gewicht, das ist eine graduelle Größe. Außerdem erhält die Äußerung u.U. als Ganzes eine bestimmte Deklination ihrer H- und/oder L-Werte (bzw. es werden Regeln für Downstep oder Downdrift aktiviert). Der graduelle phonetische Tonwert der einzelnen Silbe bestimmt sich dann aus drei Größen: ihrem phonologischen Tonwert, ihrem Gewicht und ihrer Position auf der Deklinationlinie. Das bedeutet z.B., daß auch Wortakzent-Silben, die keine Fokussilben sind, einen erhöhten phonetischen Tonwert erhalten können.

Dieser phonetische Tonwert ist nun auf eine Folge von F_0 -Werten auf den einzelnen Segmenten der Silbe abzubilden. Dafür spielen die Sonorität der Segmente und die inhärente F_0 -Beeinflussung durch die Segmente eine Rolle. Zugleich sind die Intensitätswerte und die Dauerwerte der einzelnen Segmente auszubilden, die ihrerseits mit den Frequenzwerten interagieren.

Über alles dies gewinnt der an F_0 -Dokumentationen arbeitende Linguist einige Einsicht, unterstützt durch die Kenntnis der vielfältigen phonetischen Untersuchungen zu speziellen dieser Einflußgrößen. Es scheint mir aber verfehlt, an dieser Stelle explizit irgendwelche Regularitäten dafür angeben zu wollen.

Allerdings muß sich der an solchen F_0 -Dokumentationen arbeitende Linguist immer des methodologischen Problems bewußt sein, daß er vielleicht Töne in der Melodie identifiziert, die allein phonetischen Einflußgrößen zu verdanken sind, oder Töne übersieht, die sich wegen dieser Einflußgrößen nur wenig ausprägen. Die Historie der Intonationsforschung hat uns gelehrt, daß Linguisten etwas zu hören glauben, was akustisch nicht vorhanden ist; ebenso kann der Betrachter einer akustischen Aufzeichnung etwas zu sehen glauben, was phonologisch nicht vorhanden ist. In der Melodie die Töne zu erkennen, ist keine triviale Aufgabe.

Diese Untersuchung ist im Rahmen des DFG-Projekts "Satzintonation und Fokusstruktur von W-Fragen im Deutschen" (WU 86/7) entstanden. Die Experimente, auf die ich hinweise, sind vor allem von Herbert Tropf durchgeführt worden. Dabei haben Robert Bannert bei der Herstellung des Stimulusmaterials und Jeanne Becker, Claudia Belting, Ingrid Kaufmann, Lerke Tyra, Kornelia Waag-Corrah und Heike Wöst bei der Durchführung geholfen. Die Experimente werden in Arbeitspapieren beschrieben, die Ergebnisse sollen an anderer Stelle veröffentlicht werden.

Literatur

- Altmann, Hans (1987) Zur Problematik der Konstitution von Satzmodi als Formtypen. In: Jörg Meibauer (Hrsg.) Satzmodus zwischen Grammatik und Pragmatik. Tübingen: Niemeyer, 22-56.
- Bierwisch, Manfred (1987) Semantik der Graduierung. In: Manfred Bierwisch & Ewald Lang (Hrsg.) Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. Berlin: Akademie, 91-286.
- Bolinger, Dwight (1986) Intonation and its Parts: Melody in Spoken English. London: Arnold.
- Chomsky, Noam (1986) Barriers. Cambridge (Mass.): MIT-Press.
- Clements, George N./Ford, K. (1979) Kikuyu Tone Shift and its Synchronic Consequences. In: Linguistic Inquiry 10, 179-210.
- Cooper, William/Sorensen, John (1981) Fundamental Frequency in Sentence Production. Heidelberg: Springer.
- Culicover, Peter W./Rochemont, M. (1983) Stress and Focus in English. In: Language 59, 123-165.
- Féry, Caroline (1986) Metrische Phonologie und Wortakzent im Deutschen. In: Studium Linguistik 20, 16-43.
- Gärding, Eva (1981) Contrastive Prosody: A Model and its Application. In: Studia Linguistica 35, 146-165.
- Gärding, Eva (1983) A Generative Model of Intonation. In: Anne Cutler & Robert Ladd (eds.) Prosody: Models and Measurements. Heidelberg: Springer, 11-25.
- Goldsmith, John (1976) Autosegmental Phonology. Ph.D.Dissertation MIT.
- Goldsmith, John (1978) English as a Tone Language. In: Communication and Cognition 11, 453-476.
- Halle, Morris/Vergnaud, Jean-Roger (1982) On the Framework of Autosegmental Phonology. In: v. d. Hulst & Smith (eds.), 65-82.

- Hayes, Bruce (1984) The Phonology of Rhythm in English. In: Linguistic Inquiry 15, 33-74.
- Höhle, Tilman (1986) VERUM-Fokus, Tübingen. Ms.
- Hulst, Harry van der/Smith, Norval (1982) (eds.) The Structure of Phonological Representations (Part I). Dordrecht: Foris.
- Jacobs, Joachim (1982) Neutraler und nicht-neutraler Satzakzent im Deutschen. In: Theo Vennemann (Hrsg.) Silben, Segmente, Akzente. Tübingen: Niemeyer, 141-170.
- Kiparsky, Paul (1982) From Cyclic Phonology to Lexical Phonology. In: v. d. Hulst & Smith (eds.), 131-175.
- Klein, Wolfgang/von Stechow, Arnim (1982) Intonation und Bedeutung von Fokus. SFB 99 Linguistik Universität Konstanz. Arbeitspapier 77.
- Kohler, Klaus J. (1977) Einführung in die Phonetik des Deutschen. Berlin: E. Schmidt.
- Ladd, Robert D. (1983) Phonological Features of Intonational Peaks. In: Language 59, 721-759.
- Lang, Ewald (1977) Semantik der koordinativen Verknüpfung. Berlin: Akademie.
- Lang, Ewald (1987) Parallelismus als universelles Prinzip sekundärer Strukturbildung. In: Linguistische Studien. Akademie der Wissenschaften der DDR.
- Leben, William (1976) The Tones in English Intonation. In: Linguistic Analysis 2, 69-107.
- Lieberman, Mark/Prince, Alan (1977) On Stress and Linguistic Rhythm. In: Linguistic Inquiry 8, 249-336.
- Nespor, Marina/Vogel, Irene (1982) Prosodic Domains of External Sandhi. In: v. d. Hulst & Smith (eds.), 225-255.
- Pierrehumbert, Janet Breckenridge (1980) The Phonology and Phonetics of English Intonation. Ph.D.Dissertation MIT.
- Pulleyblank, Douglas (1986) Tone in Lexical Phonology. Dordrecht: Reidel.
- Selkirk, Elisabeth (1984) Phonology and Syntax. The Relation between Sound and Structure. Cambridge (Mass.): MIT-Press.
- v. Stechow, Arnim (1985) Comparing Semantic Theories of Comparison. In: Journal of Semantics 3, 1-77.
- v. Stechow, Arnim/Uhmann, Susanne (1985) Some Remarks on Focus Projection. 6th Groningen Grammar Talks.
- Thorsen, Nina (1980) Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish. In: Annual Report. Institute of Phonetics. Univ. of Copenhagen 14, 1-29.

- Travis, Lisa (1984) Parameters and Effects of Word Order Variation. Ph.D. Dissertation MIT.
- Tropf, Herbert/Wunderlich, Dieter (1985) Zur Wahl des Grenztons in W-Fragen. Arbeitspaier DFG-Projekt Wu 86/7. Düsseldorf. Ms.
- Wiese, Richard (1986a) Syntax und Phonologie - Ein Überblicksartikel anhand von: |Selkirk 1984|. In: Linguistische Berichte 103, 252-276.
- Wiese, Richard (1986b) Nichtlineare Phonologie - Eine Fallstudie des Chinesischen. In: Linguistische Berichte 102, 93-135.
- Wunderlich, Dieter (1986) Echofragen. In: Studium Linguistik 20, 44-62.
- Wunderlich, Dieter (1987) Partizipien im Deutschen. In: Linguistische Berichte.

Rhythmische und tonale Struktur der Intonationsphrase¹

Caroline Féry , Konstanz

0. EINLEITUNG

Die vorliegende Arbeit behandelt die rhythmische und die tonale Struktur der deutschen Intonationsphrase unter phonologischen Gesichtspunkten. Der erste Teil besteht aus einer Beschreibung der rhythmischen Struktur der Intonationsphrase (IP), die dabei in zwei Takte eingeteilt wird, wobei der Nuklearakzent, der im Deutschen mit der letzten und prominentesten tonalen Hervorhebung zusammenfällt, als Trennung zwischen den beiden Takten fungiert. Die zwei Takte weisen verschiedene rhythmische Eigenheiten auf, die ausführlich dargestellt werden. Im zweiten Teil des Aufsatzes wird die tonale Struktur behandelt. Es wird gezeigt, daß die Aufteilung der Intonationsphrase in Takte auch für die tonale Struktur relevant ist. Der erste Takt besteht aus einer Reihe von Tonakzenten, die zurückgestuft werden (*downstep*), der zweite Takt besteht aus dem Nuklearakzent und einer Melodie, die keine Hervorhebungen mehr enthält. Der zweite Takt wird auch durch einen Grenzton abgeschlossen.

1. RHYTHMISCHE STRUKTUR DER INTONATIONSPHRASE

Der Terminus *Intonationsphrase* setzt sich immer mehr in der Intonationsforschung durch, weshalb er auch hier gewählt wird.

Die Betonungsverhältnisse oder die melodische Gestalt der Intonationsphrase (oder beide) dienen dabei als ausschlaggebende Kriterien. Die Intonationsphrase hat schon viele Bezeichnungen erhalten, je nachdem, welche ihrer Eigenschaften gerade betrachtet wird: *Phrasierungseinheit* bei Bierwisch (1966), *rhetorisches Syntagma* bei von Essen (1956), *Tongruppe* bei Pheby (1980) oder *Fokusdomäne* bei Gussenhoven (1983) beziehen sich in den meisten Fällen auf dasselbe Phänomen.

1. Ich danke Arnim von Stechow, Wolfgang Sternefeld, Hubert Truckenbrodt, Susanne Uhmann, Richard Wiese und Ede Zimmermann für Kommentare, Hilfe und Unterstützung.

Der Terminus *Nuklearakzent* bezeichnet die als am stärksten betont empfundene Silbe innerhalb der Intonationsphrase und muß vom lexikalischen Akzent unterschieden werden. Der lexikalische Akzent ist ein potentieller Akzent ("stress"), der auf jedem lexikalischen Wort vorhanden ist, und zwar auf einer metrisch starken Silbe (für eine metrische Behandlung der Wortakzentuierung siehe Féry 1986). Der Nuklearakzent (oder der grammatische Akzent) ist dem *designated terminal element*² einer metrischen prosodischen Domäne äquivalent und fällt normalerweise mit einem realisierten lexikalischen Akzent zusammen.

Der Nuklearakzent (wie übrigens andere Hervorhebungen der Intonationsphrase auch) wird im Deutschen vorwiegend durch eine Tonbewegung realisiert. Diese wahrgenommenen, durch eine Tonbewegung realisierten Hervorhebungen werden hier *Tonakzente* genannt.

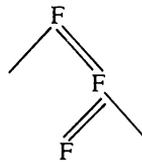
Die genaue Stelle der fallenden oder steigenden Tonhöhenveränderung in einer Silbe (und möglicherweise in den folgenden Silben) ist dabei entscheidend. (Tropf, Bannert und Wunderlich 1986, Goldbeck und Sendlmeier 1987)

Im folgenden werde ich nicht auf das Problem der Stelle des Nuklearakzents im Satz eingehen, da ich mich hier auf die Phonologie des Taktes als rhythmische und melodische Einheit beschränken möchte. Der nötige theoretische Rahmen wird hier nur skizziert: Wir wollen in dem Konstanzer Projekt Regelmäßigkeiten in der grammatischen Akzentuierung von Sätzen untersuchen. Wir unterscheiden begrifflich zwei Aspekte der Fokussierung:

- 1) "Fokus" kennzeichnet einen bestimmten syntaktischen Bereich
- 2) Fokus wird an einer bestimmten Stelle realisiert (phonetisches Fokuskorrelat), und zwar als Nuklearakzent des Satzes.

Wir beschreiben diese Doppelfunktion folgendermaßen: Einem Syntagma wird ein Fokusmerkmal zugewiesen, das an einer Stelle phonetisch realisiert wird:

(1)



2. Das 'designated terminal element' eines metrischen Baums ist das stärkste phonologische Element des Baums und wird ausschließlich von s-Knoten dominiert.

In Analogie zur morphologischen Terminologie sprechen wir von *Fokusprojektion* oder *Fokusvererbung*.

Wenn der ganze Satz fokussiert ist, d.h. wenn der ganze Satz als neue Information gilt, wird der Fokus auf einem Wort realisiert. Dieses ist dann der *Fokusexponent* (in der Terminologie von Fuchs (1976)).

In den weiter unten folgenden Beispielen wird der Fokus durch eine in Klammern gestellte Frage induziert. Ist keine Frage vorhanden, soll der Satz "ohne Kontext" gesprochen werden. Ich nehme für den Zweck dieses Papiers an, daß die Sätze in diesen Fällen ganz fokussiert sind.

Ich komme nun zu den grundlegenden rhythmischen Eigenschaften von Intonationsphrasen:

1.1. Silbenzahl

Eine Intonationsphrase besteht aus mindestens einer Silbe, wobei genau eine Silbe den Nuklearakzent trägt. Problematisch sind einsilbige Wörter in Isolation wie *ja/nein, so* usw... Betonung ist bekanntlich ein relativer Begriff, und eine betonte Silbe braucht unbetonte Nachbarsilben, um hervorgehoben zu wirken. Diese Bedingung wird bei einsilbigen Wörtern offensichtlich nicht erfüllt, obwohl diese die typische tonale Gestalt eines Tonakzents aufweisen; ich gehe deshalb im folgenden davon aus, daß die tonale Struktur eines Ausdrucks ausreicht, um eine Intonationsphrase zu bilden.

Die höchstmögliche Anzahl von Silben innerhalb einer Intonationsphrase ist vielleicht empirisch feststellbar, aber nicht sonderlich interessant. Interessanter ist vielleicht die maximale Dauer, die eine Intonationsphrase haben kann. Es besteht eine Tendenz, am Ende eines Satzes schneller zu sprechen als am Anfang; das wurde schon von Bierwisch (1966) beobachtet. Die Intonationsphrasen nehmen also gegen Satzende bezüglich der Segmentzahl zu.

Auch der Stil spielt eine Rolle: in einem Gespräch sind die Intonationsphrasen kürzer als in einer feierlichen Rede oder beim Vorlesen. Hinzu kommt noch, daß die Dauer oder die Länge der Intonationsphrase eine sprecherabhängige Angelegenheit ist: manche Sprecher können sehr lange Intonationsphrasen bilden, während andere eine Tendenz haben, jedem Wort eine eigene Intonationsphrase zuzuweisen.

1.2. Ein Nuklearakzent pro Intonationsphrase

Jede Intonationsphrase enthält einen einzigen Nuklearakzent. In manchen Fällen können aber zwei gleichstarke Tonakzente vorhanden sein, wie in dem mehrteiligen Fokus von Jacobs (1982:145):

- (2) Nicht₁ [₁Péter₁] liebt [₁Gérda₁].
(sondern Gérda liebt Péter)

Oder auch in phrasalen Koordinationen mit Gapping des Verbs wie:

- (3) || Elsa schenkt ihrem Sohn || einen Teddy || und ihrer Tochter || einen Schnuller.||

In diesen Fällen nehme ich an, daß die gleichstarken Tonakzente jeweils eine Intonationsphrase bilden. Eine Alternative bestünde darin, eine tiefere Ebene der Phrasierung zu postulieren, also so etwas wie eine Akzentphrase oder eine Akzentdomäne. Für die Existenz einer solchen Intonationseinheit gibt es aber bisher nicht genügend unabhängige Belege.

Die Phrasierung einer Äußerung wird nicht notwendigerweise von ihrer Fokusstruktur bedingt; sie kann auch rhythmisch bedingt sein. Manche Konstituenten haben eine Tendenz, eine eigene Intonationsphrase zu bilden, sobald sie zu lang werden. Das gilt insbesondere für Nominal-, Präpositional-, Adverbial- und Adjektivphrasen im Vorfeld. Vgl. dazu (4) und (5) (siehe auch (42)):

- (4) || Die Fjorde sind schön.||

- (5) || Die Fjorde in Norwegen || sind unbeschreiblich schön.||

Die Fokusstruktur dieser beiden Sätze soll gleich sein. Ihre Phrasierung ist jedoch völlig verschieden. Satz (4) besteht aus einer einzigen Intonationsphrase mit dem Nuklearakzent auf *schön* und einem zusätzlichen Tonakzent auf *Fjorde*. Satz (5) besteht aus zwei Intonationsphrasen, wobei die erste den Nuklearakzent auf *Norwegen* und einen zusätzlichen Tonakzent auf *Fjorde* hat; die zweite Intonationsphrase hat einen Nuklearakzent auf *schön* und einen Tonakzent auf *unbeschreiblich*. Diese Art von rhythmisch bedingten Phrasierungen muß sorgfältig von dem unterschieden werden, was unter dem Stichwort "Topik/Kommentar-Struktur" bekannt ist (siehe Jacobs 1984)³.

3. "Topik" hat im linguistischen Jargon mindestens drei verschiedene Bedeutungen. Hier ist Topik im Sinne von Jacobs gemeint: "ein Rahmen für den Rest des Satzes". Er gibt als Beispiel:

[Was deinen Geburtstag betrifft], kann ich leider nicht kommen.

Topik wird aber auch manchmal in einem etwas allgemeineren Sinn benutzt; so bei Krifka (1984), der neben dem Topik von Jacobs auch ein anderes Topik erkennt: "das Element worüber man redet". Beispiel:

[Der Meister] war wieder mal indisponiert.

Eine dritte übliche Benutzung des Terminus Topik bezieht sich auf eine Stelle im Satz. Topik ist dann das Element, das am Satzkopf steht.

1.3. Grenzen

Eine Intonationsphrase hat Grenzen, die oft erkennbar sind. So wird sie manchmal von ihren Nachbargruppen mithilfe eines Grenztons (im Sinne Pierrehumberts 1980) abgehoben. Der Grenzton kann dabei hoch oder tief sein; die finale Intonationsphrase einer deklarativen Äußerung erhält in den unmarkierten Fällen einen Tiefton.

Finale Silben einer Entscheidungsfrage sowie der ersten Elemente einer Aufzählung können dagegen oft durch einen hohen Grenzton charakterisiert werden (vgl. aber auch (39)):

(6) || Ist Anna gekommen? ||

(7) || Alains Bruder || Annas Schwester || und Martin sind zusammen gefahren.||

Ein Grenzton am Anfang der Intonationsphrase ist selten notwendig.

Die andere Strategie, um die Grenze einer Intonationsphrase zu markieren, ist die Pause oder die verlängerte Silbe (was aber zu einer Pause äquivalent ist). Die Pause ist allerdings nicht immer ein sicheres Grenzzeichen, weil sie auch oft willkürlich als Verzögerung zu interpretieren ist. Ich nehme auch an, daß in bestimmten Fällen eine Pause mitten in einer Intonationsphrase auftreten kann. Das gilt insbesondere bei Engfokussierungen im Vorfeld, wobei der Rest des Satzes mit einer tiefen flachen Stimme ausgesprochen wird, sowie in den im Englischen sogenannten *intonational tags*:

(8) (i) (Was hat er gekauft?)

|| Drei EIER hat er gekauft ||

(ii) || Ziemlich LANGWEILIG, dein Freund ||

(iii) || Das ist meine SCHWESTER, Maria.||

(iv) || DONNERWETTER, hat er gebrüllt.||

Die zweiten Teile der Sätze in (8), als jeweils einzelne Intonationsphrasen zu analysieren, ist kaum sinnvoll, da sie weder rhythmische noch tonale Struktur besitzen.

Die Pause ist also vorsichtig zu interpretieren, kann aber als Bestätigung einer Grenze, die auf anderer (z.B. syntaktischer) Grundlage rechtfertigbar ist, benutzt werden.

1.4. Disambiguierung

Die Einteilung in Intonationsphrasen dient manchmal der Disambiguierung von Äußerungen: in den Beispielen (8) (iii) und (iv) ist auch eine andere Interpretation möglich: setzt man einen Tonakzent auf *Maria* und *hat er gebrüllt*, erhalten die Sätze eine andere Bedeutung: *Maria* ist dann nämlich eine Apposition, und was er gebrüllt hat, war nicht unbedingt *Donnerwetter*, sondern möglicherweise etwas ganz anders. Ähnliche Fälle sind die folgenden:

- (9) (a) ||Martin bedauert,|| Maria aber nicht.||
 (b) ||Martin bedauert Maria aber nicht.||
- (10) (a) ||Der Mensch denkt;|| Gott lenkt.||
 (b) ||Der Mensch denkt, Gott lenkt.|| (Pheby 1980)
- (11) (a) ||Ede fährt nicht schnell,|| um Alain einen Gefallen zu tun.||
 (b) ||Ede fährt nicht schnell um Alain einen Gefallen zu tun || (sondern weil er sich verspätet hat)
- (12) (a) ||Rainer verfolgt den Mann,|| mit dem Motorrad.||
 (b) ||Rainer verfolgt den Mann mit dem Motorrad.||

Die Beispiele unter (a) werden jeweils als zwei Intonationsphrasen ausgesprochen, und die Beispiele unter (b) als eine einzige Intonationsphrase. Man kann sich leicht davon überzeugen, daß die Sätze dann jeweils eine andere Bedeutung haben.

1.5. Die hierarchische Struktur von Intonationsphrasen

Die Debatte um die Existenz höherer Intonationseinheiten als der Intonationsphrase läßt sich am besten durch einen kurzen Vergleich der Meinungen zweier bedeutender amerikanischer Intonationsforscher illustrieren:

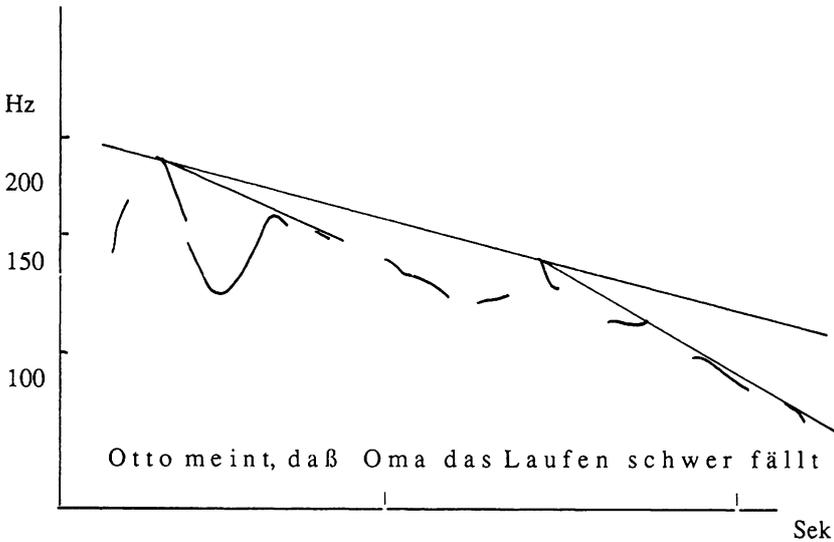
Ladd (1986) erlaubt Einbettungen von Intonationsphrasen, wie etwa in dem folgendem Satz:

- (13) My brother, who is a geologist, lives in Denver.

Die andere Möglichkeit einer hierarchischen Strukturierung der Intonation bei Ladd ist die Bildung von Super-Major-Phrasen (wobei Major-Phrasen ungefähr dasselbe wie Intonationsphrasen sind)

- (14) I asked him what he thought of it, and he said it was pretty bad.

Das stärkste Argument bei Ladd ist die *declination within declination* (Deklination innerhalb der Deklination). Die ist übrigens im Deutschen auch anzutreffen:



(15) Otto meint, daß Oma das Laufen schwerfällt⁴.

Beckman und Pierrehumbert (1986) bezweifeln dagegen die Existenz von größeren Intonationsdomänen als der Intonationsphrase und zeigen, daß es weder phonetische noch phonologische Hinweise dafür gibt. Nach ihnen ist die Deklination wahrscheinlich der kombinierte Effekt von Tonzurückstufung und *final lowering* der Stimme am Ende einer Diskurseinheit. Sie zeigen auch, daß *final lowering* eine weniger häufige melodische Erscheinung ist, als bisher angenommen wurde.

2. DIE RHYTHMISCHE STRUKTUR DER TAKTE

Die Intonationsphrase kann, so wie sie eben definiert wurde, von unterschiedlicher Länge sein; ich will sie im Folgenden in zwei Takte unterteilen, wobei der erste Takt sich bis zur letzten Silbe vor dem Nuklearakzent erstreckt und der zweite Takt aus dem Nuklearakzent und allem, was darauf folgt, besteht.

Traditionell bestehen metrische Füße (auf deutsch manchmal phonologische Takte genannt) aus einer betonten Silbe und den folgenden unbetonten Silben, wobei jeder Takt entweder mit einer betonten Silbe oder einer rhythmischen Pause anfängt: Ausdrücke wie *Vergeblich!* und *Bestimmt!* bestehen nach Giegerich (1985) und Pheby (1980) also aus zwei Takten. Ich will sie lieber jeweils als einen einzigen Takt behandeln. Alle schwachen Silben, die vor der ersten starken Silbe stehen, werden an den ersten Takt adjungiert, im Sinne einer Konvention aus Hayes (1980), die auf englisch *Stray Syllable Adjunction* (Adjungierung einzelner Silben) heißt. Damit entfällt hier die Notwendigkeit, rhythmische Pausen anzuwenden. Die einzigen Fälle, wo man sonst eine rhythmische Pause annehmen könnte, liegen bei der folgenden Art von Beispielen vor:

⁴ Die Beispielsätze wurden von zwei Personen gesprochen. Es handelt sich um einen neunundzwanzigjährigen Mann und eine dreißigjährige Frau, beide Muttersprachler des Standarddeutschen. Die Abbildungen zeigen die Frequenz der Ausdrücke.

(16) || Aber... das Auto ist kaputt.||

Das Wort *aber* bildet möglicherweise eine einzige Intonationsphrase, hat aber keine rhythmische Struktur - genau wie der Ausdruck *dein Freund* im Beispiel (8ii) auch keine rhythmische Struktur besaß. Hier sollte man m. E. den Begriff der Extrametrität bemühen: Ausdrücke, die pragmatisch als Verzögerung anzusehen sind, werden behandelt, als ob sie im Metrum, in der rhythmischen Struktur der Sprache, gar keine Rolle spielen. So könnte auf das Konzept der rhythmischen Pause ganz verzichtet werden.

Folgende Beispiele dienen zur Illustration der Einteilung in zwei Takte:

(17) || Gudrun hat ein | Lexikon gekauft.||

(18) (Wer hat ein Lexikon gekauft?)

|| Gudrun hat ein Lexikon gekauft.||

(19) || Wolfgang | schläft. ||

Die Beispiele (17) und (19) bestehen jeweils aus zwei phonetisch realisierten Takten, wobei der zweite Takt von (19) eine einzige Silbe ist, auf welcher der Nuklearakzent und der Grenzton realisiert werden. Der zweite Satz von (18) hat dagegen einen ersten phonetisch leeren Takt, da der Nuklearakzent schon auf der ersten Silbe der Intonationsphrase realisiert wird.

Die beiden eingeführten Takte weisen ganz verschiedene rhythmische Phänomene auf.

2.1. Die rhythmische Struktur des ersten Taktes

2.1.1. Eurhythmie-Prinzipien

Die Eurhythmie-Prinzipien sind für die Alternation von schwach und stark betonten Silben in der Sprache verantwortlich und sorgen auf der einen Seite dafür, daß kein Akzentzusammenstoß vorkommt; d.h. zwei stark betonte Silben werden nicht auf zwei unmittelbar benachbarten Silben realisiert, wenn sie derselben Intonationsphrase angehören. In diesen Fällen findet, womöglich, eine Akzentverschiebung statt (2.1.2.).

Die andere Aufgabe der Eurhythmie-Prinzipien ist die Vermeidung von rhythmischen Lücken. Es wird manchmal behauptet, daß die ideale rhythmische Struktur der Sprache eine ständige Alternanz von schwachen und starken Silben ist; oder auch, daß in einer akzentmetrisierten Sprache wie dem Deutschen die Füße isochron sein sollen. Dieser allzu mechanische Rhythmus ist aber in der gesprochenen Sprache praktisch gar nicht zu finden (außer in Kinderreimen).

Trisyllabische Füße sind im Deutschen genauso häufig wie bisyllabische. Lange Folgen von unbetonten Silben (z.B. Klitika) werden aber vermieden.

- (20i) Er hat' s ɾ ɣ gegeben. (Er hat es ihm gegeben)
 (20ii) Er hat' s ɾ ɣ ʃ ɔ n gegeben (Er hat es ihm schon gegeben)
 (20ii) ? Er hat' s ɾ ɣ ʃ ɔ n gegeben (Er hat es ihm schon gegeben)

Im Wort wandert die Nebenbetonung möglichst weit weg vom Hauptakzent, anstatt eine Alternation zu schaffen (wie es z.B. im Spanischen der Fall ist):

- (21) $\begin{array}{cc} 2 & 1 \\ \text{Philosophie} \end{array}$ $\begin{array}{cc} 2 & 1 \\ \text{Biologie} \end{array}$ $\begin{array}{cc} 2 & 1 \\ \text{Analphabetismus} \end{array}$

In Phrasen, die aus drei betonbaren Wörtern bestehen, erhält das erste Wort die Nebenbetonung :

- (22) $\begin{array}{cc} 2 & 1 \\ \text{die schönen jungen Männer} \end{array}$ * $\begin{array}{cc} 2 & 1 \\ \text{die schönen jungen Männer} \end{array}$

Eurhythmie-Prinzipien:

- Vermeide Akzentzusammenstöße.
- Vermeide rhythmische Lücken durch Bevorzugung bisyllabischer und trisyllabischer Füße.
- Verschiebe die Nebenbetonung so weit wie möglich weg vom Hauptakzent in der phonologischen Domäne. (Silben, die ein Schwa enthalten, werden dabei ausgeschlossen, da sie prinzipiell unbetont sind.)

Intonationsphrasen sind Barrieren für die Anwendung von Eurhythmie-Prinzipien:

- (23) $\begin{array}{cc} 2 & 1 \\ \text{Pascale arbeitet in der Universität Konstanz.} \end{array}$ || (Akzentverschiebung)

- (24) $\begin{array}{cc} 1 & 1 \\ \text{(Ihre Adresse lautet:)} \end{array}$ || Universität, || Konstanz, || Bundesrepublik ||

2.1.2. Akzentverschiebung

Die Akzentverschiebung verhindert, daß zwei stark betonte Silben zu nahe beieinander stehen (siehe Féry 1986 für eine genaue Definition dieses Begriffes). Das gilt für Wörter ebenso wie für Phrasen: in Wörtern nach Affigierung betonter Affixe und in bestimmten Kompositionen, und zwar nach rechts und nach links.

1 2 1 3 2 2 1 2 3 1
 (25) vollständig, unvollständig; Nation, national.

1 2 1 3 2 1 2 1 3 2
 Schuhmacher, Handschuhmacher. Mittag, Nachmittag.

1 2 1 2
 (26) Akzentzusammenstoß Hallenschwimmbad

In Phrasen kann die Akzentverschiebung über Wortgrenzen hinausgehen:

2 1 2 3 1
 (27) fing an ; fing an zu reden.

Kiparsky (1966) nimmt an, daß die Akzentverschiebung im Deutschen symmetrisch ist, daß sie sowohl nach links wie auch nach rechts (in Phrasen und in Wörtern) gehen kann. Das Beispiel (28) ist für ihn ein Fall von Akzentverschiebung nach rechts. Es handelt sich dabei m. E. jedoch eher um eine Deakzentuierung (siehe 2.2.) (auch wenn hier in markierten Fällen manchmal eine Akzentverschiebung realisiert wird):

1 2 1 2 2 1 3 2
 (28) anziehen; den Rock anziehen und nicht *den Rock anziehen

In Wörtern geschieht also die Akzentverschiebung im Deutschen in beide Richtungen, in Phrasen dagegen nur nach links.

2.2. Der zweite Takt

Im zweiten Takt findet in der rhythmischen Struktur die Deakzentuierung statt. Die Deakzentuierung betrifft Silben, die ansonsten betont wären, und die infolge ihrer Stelle im Satz oder im Wort ihre Betonung verlieren. Im Satz werden alle Silben, die nach dem Tonakzent im zweiten Takt vorkommen, deakzentuiert. Beispiel:

(29) (Tom hat das rosa Handtuch benutzt.)
 (Welches Handtuch?)
 || Das rosa Handtuch.||

Alle Silben, die nach dem Nuklearakzent vorkommen, sind deakzentuiert. Dieses Phänomen kann auch längere Äußerungen betreffen, wie im Beispiel (18). Es ist typisch für den zweiten Takt, kann aber auch im ersten Takt vorkommen:

- (30) (Ich habe ein wunderschönes Haus gesehen.)
 (Ein wunderschönes was?)
 || Ein wunderschönes Haus.||

Deakzentuierung in einzelnen Wörtern ist nur nach einem sog. kontrastiven Akzent zu finden, wie in (31)⁵:

- (31) entladen und nicht beladen

In diesem Beispiel tragen Silben, die normalerweise unbetont sind, den Worthauptakzent. Die zweiten Silben werden deakzentuiert.

Abschließend und zusammenfassend können die folgenden Sätze miteinander verglichen werden.

- (32) || Der Rock war gras | grün.||
 2 1
- (33) || Der grasgrüne | Rock || stand ihr | gut.||
 2 1 1
- (34) || Gras | grün war der Rock.||
 1

(32) besteht aus einer Intonationsphrase und zwei Takten (wobei der zweite Takt nur aus der hauptakzentuierten Silbe besteht); der Nuklearakzent geht auf *grün*, und das Wort *Rock* erhält einen Nebenakzent. (33) besteht aus zwei Intonationsphrasen, wenn das Wort *Rock* neben seiner normalen Betonung als letztes Element einer Nominalphrase noch einen Grenzton trägt. Der erste Takt der ersten rhythmischen Gruppe besteht aus den Wörtern *der grasgrüne*, wo infolge eines Akzentzusammenstoßes eine Akzentverschiebung stattfinden muß.

(34) enthält eine topikalisierte Konstituente, die auch den Nuklearakzent des Satzes (oder der Intonationsphrase) trägt. Infolgedessen wird der Rest des Satzes deakzentuiert.

3. TONALE STRUKTUR DER INTONATIONSPHRASE

Im folgenden werde ich die tonale Struktur von verschiedenen Intonationsphrasen skizzieren, ohne dabei in phonetische Details zu gehen. Das bitonale System, das Pierrehumbert (1980) für das Englische entwickelt hat, wurde im Rahmen des Konstanzer Intonationsprojekts für das

⁵. Es ist wohl bekannt, daß die Betonung durch andere Mittel als Tonhöhenveränderung realisiert werden kann. Im zweiten Takt (wie auch beim Flüstern) werden Wortbetonungen wahrgenommen. Dauer und Intensität spielen in solchen Fällen eine entscheidende Rolle.

Deutsche übernommen und weiterentwickelt. (In diesem Sinn schlägt Uhmann (1986, 1987) ein bitonales System vor, das unterschiedliche Fokusstrukturen in Deklarativsätzen analysieren kann.) Hier wird ein ähnliches Intonationsmodell dargestellt, das aber auf beliebige Intonationsphrasen anwendbar und mit der rhythmischen Struktur der deutschen Intonationsphrase besonders gut verträglich ist. Es werden Deklarativ-, Frage-, Imperativ-, Vokativsätze usw. berücksichtigt. Weitere Beispiele für Intonationsphrasen sind:

- (35) || Hat Tom schon gegessen? ||
 || Komm mal her! ||
 || Angelika! ||

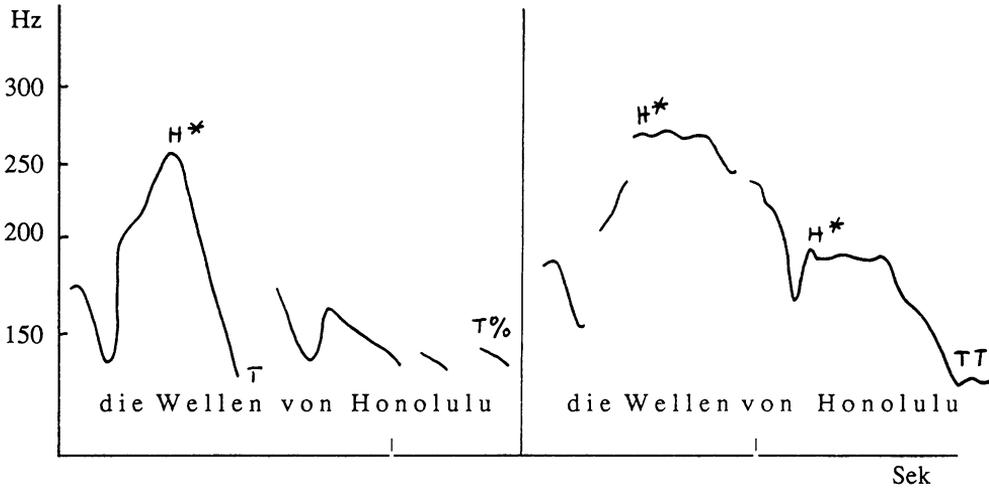
Das Intonationsmodell von Pierrehumbert wird oft als Tonsequenzenanalyse der Intonation bezeichnet, die sich von der traditionellen Konturauffassung (z.B. der Britischen Schule) unterscheidet. (Für einen kritischen Vergleich der beiden Richtungen siehe Ladd 1983.) Pierrehumbert (1980) schlägt einen endlichen Automaten für den Aufbau der Intonation vor, der nach ihrem Modell schrittweise und linear abläuft. Jeder Ton wird auf der Basis der vorangehenden Töne produziert. Die Verkettung der einzelnen Töne erzeugt dann die Melodie. Pierrehumbert beruft sich auf Bruce (1977), der für die Analyse der schwedischen Wortakzente im Satz drei verschiedene Arten von Tönen isoliert hatte, und übernimmt die drei Töne mit folgenden Funktionen:

- Tonakzente (*pitch accents*), die die wahrgenommenen Tonhöhenprominenzen ausdrücken. Pierrehumbert kontrastiert sechs wohldefinierte Akzentformen: zwei monotonale Töne H* und T*, wobei H für Hochton und T für Tiefton stehen, und vier⁶ bitonale Töne H*+T, H+T* T*+H und T+H*. Der Ton mit Stern wird mit der metrisch starken Silbe assoziiert, der andere Ton ist von dem gesterntem Ton durch einen gewissen Zeitabstand getrennt, wird aber nicht unbedingt auf derselben Silbe realisiert. Im folgenden wird kein + Zeichen mehr benutzt.

Die Töne (oder wohldefinierten Akzentformen) können im Prinzip mit jedem lexikalischen Element assoziiert werden; sie kontrastieren im Deutschen und im Englischen unterschiedliche Fokussierung und Intonationsbedeutungen (Behauptung, Frage, Erstaunen...) und nicht lexikalische Einheiten wie in Tonsprachen (z.B. Chinesisch, Igbo) oder Wortakzentsprachen (Schwedisch, Norwegisch⁷).

6. In ihrer Dissertation hatte Pierrehumbert noch H*+H, das für bestimmte Hut-Konturen relevant war. Später verzichtete sie auf diesen Ton.

7. Manche Wörter des Schwedischen und des Norwegischen tragen einen distinktiven Akzent (Akzent 1 oder 2); ansonsten wird der Modus des Satzes durch die Intonation wiedergegeben. Für eine ausführliche Beschreibung der schwedischen Wortakzente siehe Bruce (1977) und der skandinavischen Wortakzente siehe Gårding (1977).



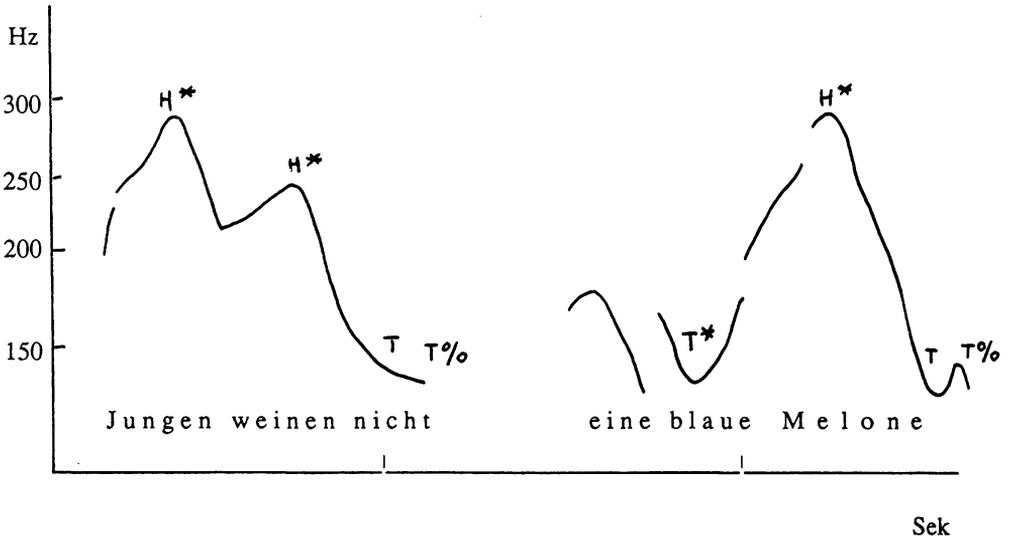
(36) || die | Wellen von Honolulu || || die Wellen von Hono | lulu ||

Linie-0-Schläge stehen für die Silben, Linie-1-Schläge für die lexikalischen Akzente und die auf Linie 2 für Nuklearakzente.

Ob die lexikalischen Akzente als Tonakzente realisiert werden oder nicht, ist also nicht nur von der metrischen Struktur abhängig, sondern auch von der Stelle des Wortes innerhalb der Intonationsphrase. Im ersten Fall trägt *Wellen* den Nuklearakzent; *Honolulu* ist Teil des zweiten Takts und erhält keinen Tonakzent. Im zweiten Fall hat *Honolulu* den Nuklearakzent; *Wellen* ist im ersten Takt und erhält einen Tonakzent.

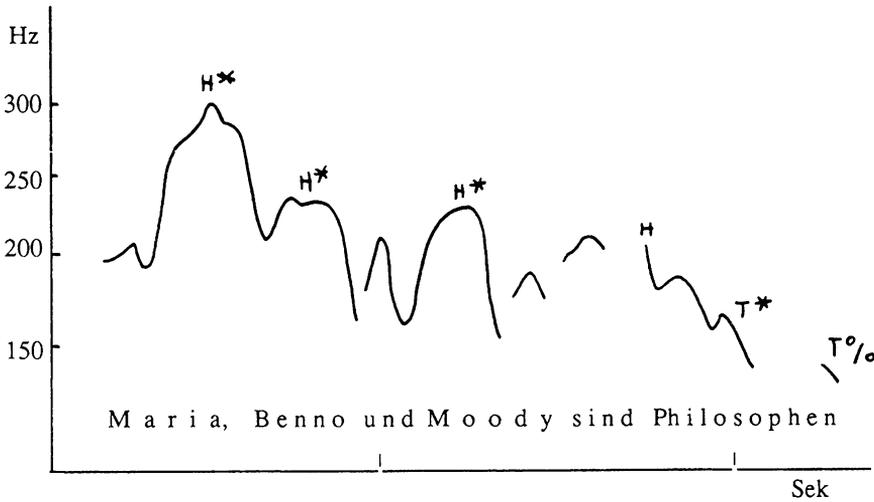
Dies kann auf alle möglichen Intonationsmuster verallgemeinert werden: Nach dem Nuklearakzent wird grundsätzlich kein Tonakzent mehr zugewiesen. Im ersten Takt sind dagegen Tonakzente üblich.

Hier sind ein paar Beispiele für Tonakzente im ersten Takt:



- (37) H* || Jungen | weinen nicht ||
- (38) T* || eine blaue Me | lone ||

(37) ist mit (36b) vergleichbar. In (38) ist der Nuklearakzent auf *Melone* und wird als H*T realisiert. Der Tonakzent auf *blau* im ersten Takt wird als Tiefton realisiert; die Sprecherin wurde dabei angewiesen, absichtlich Erstaunen oder Kontrast auszudrücken.

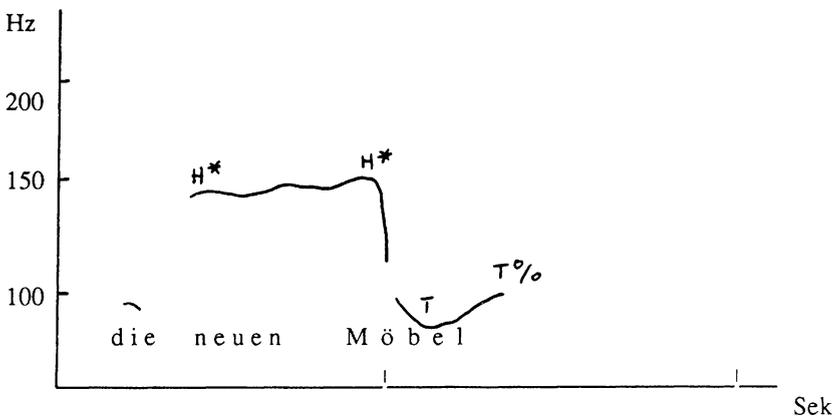


x x x x
 x xx x x x x x x x
 (39) H*T || Maria, Benno und Moody | sind Philosophen ||

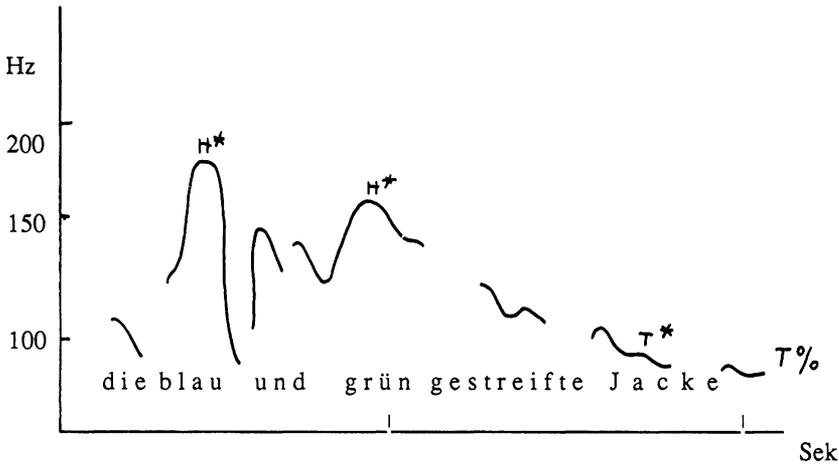
Die Auflistung wird hier als Teil des ersten Taktes analysiert. Jedes Wort hat einen bitonalen Ton, der den Eindruck vermittelt, daß jedes Wort eine eigene Phrase bildet. Beckman und Pierrehumbert reden in solchen Fällen von *intermediate phrases*. (Wie es schon für die rhythmische Struktur bemerkt wurde, konnte die Existenz einer solchen Intonationseinheit im Deutschen bisher nicht genügend belegt werden.) Bemerkenswert ist hier die Zurückstufung der Hochtöne in der ganzen Intonationsphrase. Der erste Teil des bitonalen Nuklearakzents HT* auf *Philosophen* ist tiefer als alle andere Tonakzente.

Die melodische Kontur zwischen den Tonakzenten wird nach Pierrehumbert mit sog. *Interpolationsregeln* charakterisiert. Diese Regeln sagen z.B. voraus, daß zwischen zwei Hochtönen stets ein Tal liegt, während zwischen zwei Tieftönen die Melodie ganz flach verläuft. Dabei muß die segmentale Struktur der einzelnen Äußerungen in Betracht gezogen werden.

Die oft beschriebene Hutkontur (ein Plateau zwischen zwei Hochtönen) ist typisch für das Standarddeutsche (40). Ihr Vorhandensein hängt entscheidend von der segmentalen Auffassung der Äußerung ab. Sie ist auch unwahrscheinlich, wenn zwischen den beiden Hochtönen mehrere unbetonte Silben liegen, wie in (41). Man vergleiche die beiden folgenden Konturen:



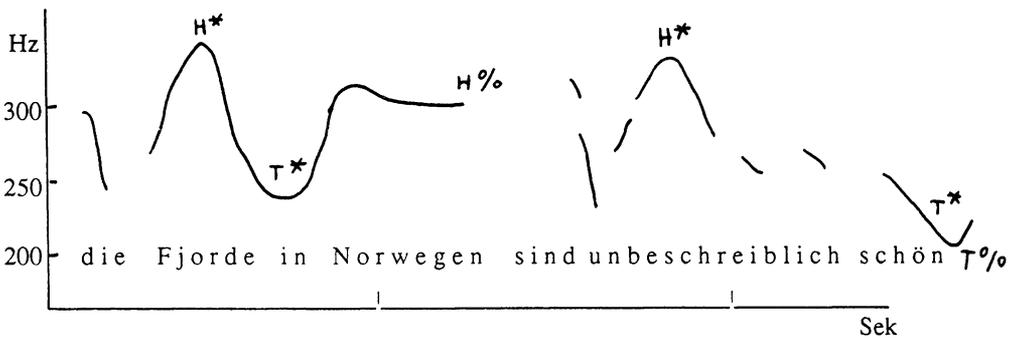
x
 x x
 x xx x x
 (40) die neuen Möbel;



x
 x x x x
 x x x x x x x x
 (41) die blau und grün gestreifte Jacke

Nochmal zur Zurückstufung: Pierrehumbert (1980) hat den *downstep* im Englischen nachgewiesen: ein Hochton wird in bestimmten Kontexten tiefer als ein vorherstehender Hochton ausgesprochen. Dieses Phänomen ist von afrikanischen Tonsprachen her bekannt. Inzwischen sprechen Beckman und Pierrehumbert (1986) für das Englische hier von einer *Katathese* ("withdrawing the term 'downstep' out of deference to the Africanist usage"), eben weil nicht bewiesen ist, daß beide Erscheinungen dieselbe Ursache und dieselbe Realisierung haben.

Das Phänomen der Zurückstufung gibt es, wie gesagt, auch im Deutschen; hierzu ein weiteres Beispiel:



x
 x x x x x x x x x x x
 x x x x x x x x x x
 (42) || Die Fjorde in | Norwegen || sind unbeschreiblich | schön. ||

Pierrehumbert nimmt in ihrem Modell an, daß ein bitonaler Ton Zurückstufung induziert, und zwar sowohl auf Tief- als auch auf Hochtönen. Ladd (1983) schlägt vor, [*Downstep*] als ein phonologisches Merkmal zu betrachten, das unabhängig von der phonetischen Realisation existiert. Es gibt also bei ihm neben H auch ein !H, den zurückgestuften Hochton. Beckman und Pierrehumbert lehnen aber diese Auffassung ab: die Hinzunahme eines weiteren Tons erweitert unnötigerweise das vorhandene Inventar.

4.2. Zweiter Takt

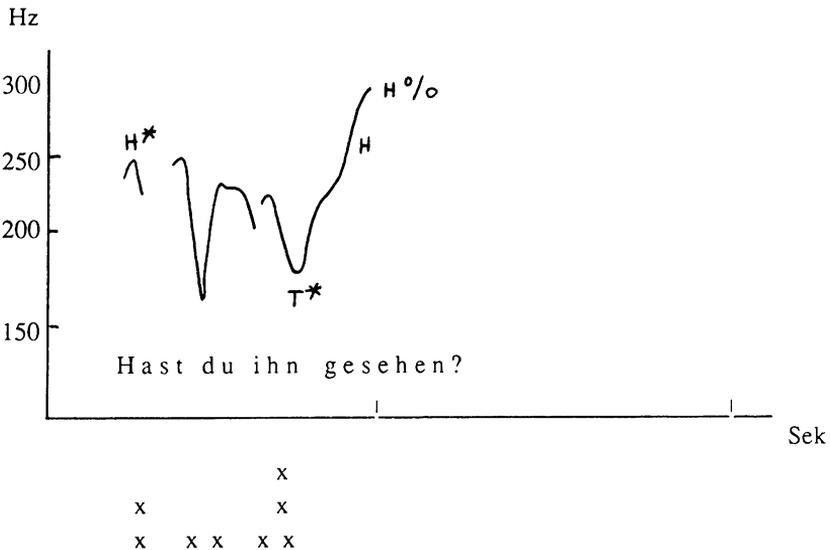
Im zweiten Takt gibt es nach dem Nuklearakzent nur noch eine Melodie und keine Akzente, also keine tonale Hervorhebung.

In dem Modell, das hier vorgeschlagen wird, stellt sich jetzt die Frage, ob der Nuklearakzent nicht eher als ein Teil des ersten Taktes anzusehen wäre, damit alle Tonakzente in den ersten Takt und die restlichen Töne in den zweiten Takt verlagert werden.

Aber das genaue Ende des Nuklearakzents ist oft schwer zu lokalisieren. Die folgenden Silben hinter der Silbe mit dem Nuklearakzent können oft als Teile eines bitonalen Tons aufgefaßt werden, besonders wenn sie zugrundeliegende unbetonte Silben sind. Vgl. dazu die Beispiele (45) und (46).

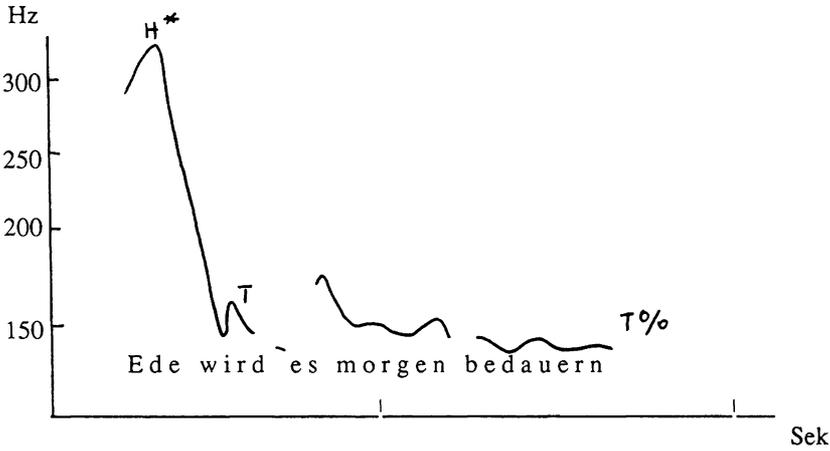
Die Lösung, die bisher gewählt worden ist, benutzt ausschließlich rhythmische Kriterien und wird weiterhin bevorzugt. Die Trennung, wie sie hier gewählt wird, hat schon eine Tradition: der zweite Takt entspricht z.B. dem *nuclear tone* der Britischen Schule.

Die Melodie des zweiten Taktes besteht also aus dem Nuklearakzent und einer terminalen Melodie. Der Nuklearakzent kann mono- oder bitonal sein.



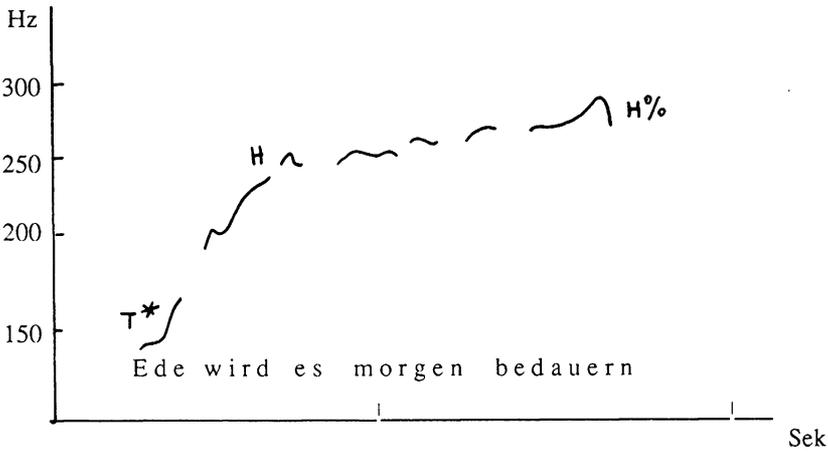
(43)H*|| Hast du ihn ge | sehen?||

(44) T* = (42) Die Fjorde in Norwegen sind unbeschreiblich schön.



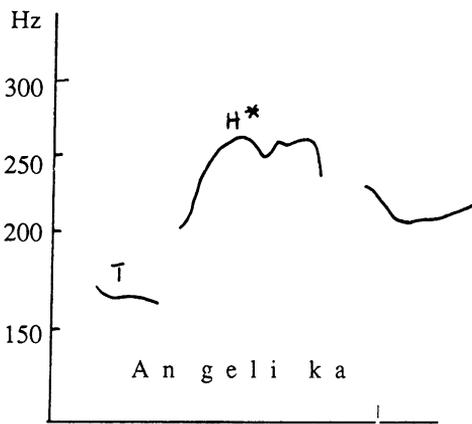
x
 x x x
 x x x x x x x x x x

(45) H*T || Ede wird es morgen bedauern ||



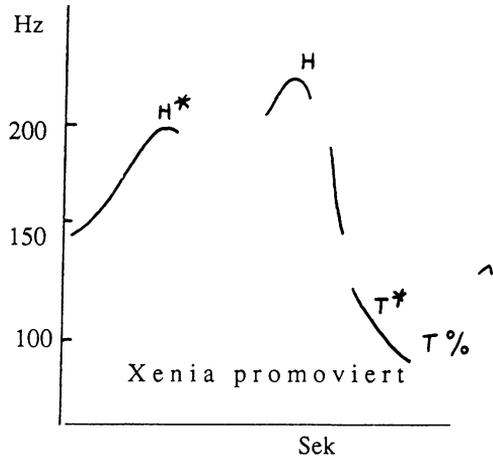
x
 x x x
 x x x x x x x x x x

(46) T*H || Ede wird es morgen bedauern ||



x
x
x x x x

(47) TH* || Angelika! ||



x
x
x x x x x

(48) HT* || Xenia promo | viert ||

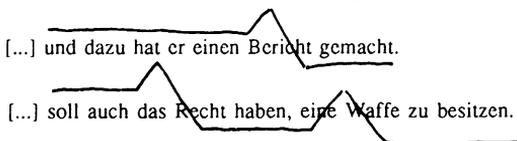
Nach Pierrehumbert werden die Tonakzente in pränukelearer wie in nuklearer Position unabhängig von der finalen Kontur gewählt. Im Standarddeutschen aber hängt T* oft davon ab, was folgt, und wird in nuklearer Position vor einem hohen Grenztone gewählt.

Goldbeck und Sendlmeier (1987) haben gezeigt, daß der Kontrast zwischen *módern* und *modérn* in finaler Stelle eines Fragesatzes dadurch ausgedrückt wird, daß der Anstieg im Falle von *módern* am Anfang der unbetonten Silbe anfängt, während in *modérn* der Anstieg spät in der zweiten betonten Silbe beginnt. In den beiden Fällen ist der Nukleartone tief.

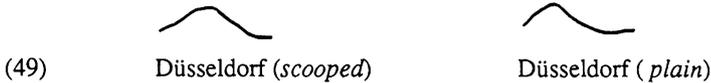
HT* und TH* wie in (47) und (48) sind Zwischenfälle: sie werden zum Teil im ersten Takt und zum Teil im zweiten Takt realisiert. Sie haben dementsprechend einen besonderen Status: Der bitonale TH* charakterisiert den Anfang einer Rufkontur.

Das Beispiel, das hier zur Illustration des Tons HT* dient, stammt aus Uhmann (1986)⁸.

8. Diese Intonation (der HT* Ton) wird häufig von Fernsehmoderatoren benutzt: z.B. hat Fritz Pleitgen in der Sendung "Wir, das Volk der Vereinigten Staaten..." ARD 4.7.87 die folgenden Sätze geäußert:



Es ist ein ziemlich markierter Fall, der vergleichbar ist mit der *scoop intonation* von Ladd (1983) (im Gegensatz zu der *plain intonation*). Die *scooped* Intonation ist für ihn ein Fall, bei dem die Assoziation des intonatorischen Gipfels mit der metrisch starken Silbe verzögert wird, und wird in Ladds Notation mit dem phonologischen Merkmal [*delayed peak*] versehen. Hier ist ein Beispiel von *plain* und *scooped* Intonation für Ladd (1983: 731):



Das hier skizzierte intonatorische Modell stellt natürlich nur einen Bruchteil der möglichen Melodien des Standarddeutschens⁹ dar, bildet aber eine brauchbare Basis für feinere Analysen der Intonation.

Zusammenfassend werden folgende Prinzipien für die Analyse der rhythmischen Struktur der Intonationsphrasen benötigt:

- 1) Die Intonationsphrase ist die Domäne der Anwendung rhythmischer und tonaler Phänomene.
- 2) Die rhythmische Struktur ist von der metrischen Struktur abhängig, die von der Folge der schwach- und starkbetonten Silben bestimmt wird.
- 3) Die Intonationsphrase enthält genau einen Nuklearakzent, der die metrisch stärkste Silbe der Intonationsphrase ist.
- 4) Die Intonationsphrase besteht aus einem bis zwei Takten, von denen der erste leer sein kann.
- 5) Der erste Takt hat eine reiche rhythmische Struktur: hier finden Euphonie-Regeln und Akzentverschiebung statt.
- 6) Der zweite Takt beginnt mit dem Nuklearakzent und enthält möglicherweise noch eine Reihe deakzentuierter Silben.

Die folgenden Prinzipien werden für die Analyse der tonalen Struktur der Intonationsphrasen benötigt:

- 7) Um die Intonation des Deutschen zu beschreiben, werden zwei Töne benötigt, ein Hochton und ein Tieftone, die die relative Tonhöhe wiedergeben. Die Töne werden entweder als Tonakzente (mono- oder bitonal) oder als Grenztöne realisiert. Sie werden mit Silben assoziiert.

9. Andere deutsche Dialekte haben eine verschiedene tonale Struktur. Schweizerdeutsch und Konstanzer Dialekt z.B. haben unmarkierte Tieftöne in Nuklearposition im Deklarativsatz. (siehe Isenberg und Ganter 1985). Das Schwäbische wiederum hat HT*H(TH) für Nukleartöne in satzfinaler Position. Siehe auch Féry und von Stechow (1987) für eine relationale Behandlung des alemannischen Akzents.

- 8) Im ersten Takt werden ausschließlich Tonakzente realisiert.
- 9) Die Töne des ersten Taktes werden meistens zurückgestuft.
- 10) Der Nuklearakzent wird meistens mithilfe eines bitonalen Tons realisiert.
- 11) Der zweite Teil eines bitonalen Tonakzents bestimmt den Melodie-Verlauf bis zum nächsten Ton (Tonakzent im ersten Takt, Grenzton im zweiten Takt).
- 12) Nach dem Nuklearakzent werden keine Tonakzente mehr realisiert.
- 13) Der Grenzton wird auf den letzten Silben der Intonationsphrase realisiert.

Literatur

- Beckman, M.E. & J.B. Pierrehumbert (1986) Intonational Structure in Japanese and English. *Phonology Yearbook* 3. 255-309.
- Bierwisch, M. (1966) Regeln für die Intonation deutscher Sätze. *Studia Grammatica* VII: 99-201.
- Bruce, G. (1977) Swedish Word Accent in Sentence Perspective. *Travaux de l' Institut de Linguistique de Lund* 12, Gleerup, Lund.
- Essen, O. von (1956) Grundzüge der hochdeutschen Satzintonation. Ratingen.
- Féry, C. (1986) Metrische Phonologie und Wortakzent im Deutschen. *Studium Linguistik* 20. 16-43.
- und A. von Stechow. (1987) Relationale Tonologie. In Vorbereitung.
- Fuchs, A. (1976) 'Normaler' und 'Kontrastiver' Akzent. *Lingua* 38, 293-312.
- Gårding, E. (1977) The Scandinavian Word Accents. *Travaux de l' Institut de Linguistique de Lund* 11, Gleerup, Lund.
- ✗Giegerich, H.J. (1985) *Metrical Phonology and Phonological Structure. German and English.* Cambridge.
- Goldbeck, T.P. & W.F. Sendlmeier (1987) Wechselbeziehung zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen. Vortrag gehalten in Augsburg bei der DGfS-Tagung, März 1987.
- Gussenhoven, C. (1983) Focus, Mode and the Nucleus. *Journal of Linguistics* 19: 377-417.
- ✗Hayes, B. (1980) *A Metrical Theory of Stress Rules.* MIT Diss.
- Isenberg, B & A. Ganter (1985) Die Intonation der Konstanzer Stadtsprache. Konstanz Ms.
- Jacobs, J. (1982) Neutraler und nicht-neutraler Satzakzent im Deutschen. In: Vennemann, Th. (Hrsg) *Silben, Segmente, Akzente.* Tübingen. 141-169.
- (1984) Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91. 25-58.
- Kiparsky, P. (1966) Über den deutschen Akzent. *Studia Grammatica* VII, 69-98.
- Krifka, M. (1984) Fokus, Topik, syntaktische Struktur und semantische Interpretation. Konstanz Ms.
- Ladd, D.R. (1980) *The structure of Intonational Meaning: Evidence from English.* Bloomington: Indiana University Press.
- (1983) Phonological Features of Intonational Peaks. *Language* 59. 721-759.
- (1986) Intonational Phrasing: the case of recursive prosodic structure. *Phonology Yearbook* 3. 311-340.
- Pheby, J.: In Heidolph, K.E., W. Flämig, W. Motsch u.a. (1980) Kapitel 6 (Phonologie: Intonation) in: *Grundzüge einer deutschen Grammatik.* Berlin: Akademie-Verlag. 839-897.
- Pierrehumbert, J.B. (1980) *The Phonology and Phonetics of English Intonation.* MIT Diss.
- Tropf, H., R. Bannert und D. Wunderlich. (1986) Wahrnehmung des Satzaketzes in

Abhängigkeit vom Fo-Verlauf in Wörtern mit variabler Akzentposition. Düsseldorf. Ms.

Uhmann, S. (1986) Akzenttöne, Grenztöne und Fokussilben. Zum Aufbau eines phonologischen Intonationssystems für das Deutsche. Dieser Band.

Uhmann, S. (1987) Tonzuweisung und Fokusprojektion. Zur Disambiguierung von Fokusstrukturen durch Tonakzente und Fo-Kontur. Vortrag gehalten in Kiel im März 1986.

Susanne Uhmann (Konstanz)

1. FOKUSZUWEISUNG UND FOKUSKONTROLLE

Grundlage für die folgenden Überlegungen ist die Annahme, daß die Fokusstruktur eines Satzes (neben anderen Faktoren, z.B. dem Modus) für die distinktiven Eigenschaften seiner Intonationskontur verantwortlich gemacht werden kann, indem nämlich fokussierte Konstituenten intonatorisch hervorgehoben werden. Diese intonatorische Hervorhebung erfolgt vor allem durch die Akzentuierung bestimmter Konstituenten im Fokusbereich des Satzes. Der nicht-fokussierte, thematische Teil der Äußerung bleibt dagegen, von einigen topologischen Besonderheiten (z.B. Vorfeld-Position) und rhythmischen Prinzipien abgesehen, in der Regel intonatorisch unmarkiert. Bringt man nun noch von der Intonation unabhängige Mittel ins Spiel, um Aussagen darüber treffen zu können, welcher Teil einer Äußerung fokussiert ist und welcher nicht, kann die behauptete Beziehung empirisch überprüft werden. Für diese intonationsexterne Kontrolle bietet sich vor allem die kontextuelle Einbettung in Frage/Antwort-Sequenzen und Korrektursequenzen an. In Ergänzungsfragen ist es die erfragte Konstituente und in Korrektursequenzen die korrigierte Konstituente, denen, im ersten Fall durch fokuskontrollierende Frageelemente und im zweiten Fall durch fokuskontrollierende intonatorische Hervorhebung, ein abstraktes **Fokusmerkmal (F)** zugewiesen wird.

In den folgenden Beispielen werden die Ausdehnung des Fokusbereichs (F-inizierte eckige Klammern) und die Akzentuierung (Großbuchstaben für die metrisch prominenteste Silbe der akzentuierten Konstituente) immer separat notiert. Auf die Notwendigkeit einer solchen Trennung wurde verschiedentlich hingewiesen (z.B. Jacobs 1984), sie wird sich auch für die hier vorgestellten Analysen als erforderlich erweisen. Zwei Beispiele für Fokuskontrolle und Akzentplazierung:

(1) Frage / Antwort-Sequenz:

F: Was_i frißt eure Katze am liebsten?

A: Unsere Katze frißt am liebsten [F_i LATzikatz].

* Überarbeitete Version eines Vortrags im Rahmen des DFG-Kolloquiums (Schwerpunkt: "Formen und Funktionen der Intonation") in Kiel 12./13. März 1986. Für Hinweise und Kommentare danke ich E.Couper-Kuhlen, C.Féry, J.Jacobs, R.Ladd, W.Oppenrieder und A.v.Stechow.

(2) Korrektursequenz:

- B: Unsere Katze frißt am liebsten Kittikat_i.
 GB: Nein, das stimmt nicht.
 Unsere Katze frißt am liebsten [_{F_i} LA_tzikat_i].

Eine Äußerung kann aber auch mehr als einen Fokus enthalten:

- (3) F: Wann_i fährt Karl wohin_{ii} ?
 A: Karl fährt [_{F_i} morgen] [_{F_{ii}} nach Berlin].

In diesem Beispiel wären dann auch zwei Akzente zu erwarten, die die beiden F-Merkmale phonologisch repräsentieren. Auch das Beispiel (4)

- (4) Nicht PE_ter liebt GER_da (, sondern Gerda liebt Peter).

könnte nur mit zwei Akzenten realisiert werden, ohne daß es hier allerdings sinnvoll wäre zu sagen, daß dieser Satz zwei Foki enthalten würde. Auf diese Möglichkeit hat Jacobs (1982:14) hingewiesen und gezeigt, daß der Fokus bestimmter fokussierender Elemente, hier der Partikel *nicht*, intonatorisch "mehrteilig" realisiert wird.

- (4') Nicht_i [_{F_i} PE_ter] liebt [_{F_i} GER_da] (sondern GER_dA liebt PE_tER)

In allen bisher vorgestellten Sätzen könnte man von einem **engen Fokus**, einem Fokus auf einer einfachen Konstituente, sprechen. Unter dem Stichwort **Fokusprojektion**¹ ist jedoch ein Phänomen beschrieben worden, das - nimmt man den Akzent zum Ausgangspunkt der Analyse - systematische Ambiguitäten in Bezug auf die Fokusstruktur zuläßt. So läßt der Satz (5) in Bezug auf die Ausdehnung des Fokusbereichs mindestens zwei Analysen zu.

- (5) Karl fährt morgen nach Ber_LIN.

(i) Fokussiert ist nur das Nomen der Präpositionalphrase. Es läge also wie in den anderen Beispielen ein enger Fokus vor, und mögliche Kontexte könnten Fragen wie *Wohin fährt Karl morgen?* oder eine Behauptung wie *Karl fährt morgen an die Nordsee.* sein.

(ii) Fokussiert ist die gesamte Verbalphrase. In diesem Fall wäre *Ber_LIN* Fokusexponent², also die Konstituente eines komplexen Syntagmas, deren Akzentuierung Fokusprojektion (weiter Fokus) erlaubt. Kontextuelle Einbettungen könnten eine Frage wie *Was macht Karl am Wochenende?* oder eine Behauptung wie *Karl will sich mal richtig ausschlafen.* sein.

Nimmt man wieder umgekehrte Perspektive ein und geht von der abstrakten Fokusmarkierung aus, so stellen sich die Fragen, wie der Fokusexponent eines komplexen Syntagmas ermittelt, und wo genau der Akzent innerhalb des Fokusexponenten realisiert wird. (Im Beispiel (5) müßte

1 Lit. u.a.: Chomsky (1971), Fuchs (1976, 1980), Halliday (1967), Höhle (1982), Jackendoff (1972), Klein & v.Stechow (1982), Ladd (1978) Selkirk (1984); v.Stechow & Uhmman (1984, 1986).

2 Vgl. Fuchs (1976,1980). Fokusexponent eines komplexen Syntagmas ist die Konstituente, deren Akzentuierung maximale Ambiguität in Bezug auf die Ausdehnung des Fokusbereichs zuläßt.

sichergestellt werden, daß der Akzent auf die zweite Silbe von Berlin projiziert wird.) Unterhalb der Konstituentenebene ist also für die Intonationsstruktur eine Beschreibungsebene relevant, die die Silbenstruktur, die Alternanz von starken und schwachen Silben, die metrische Struktur, repräsentiert und sie für die Regeln zugänglich macht, die als **Fokusregeln** bezeichnet werden. Fokusregeln lenken Akzente bei engem Fokus zunächst auf die fokussierte Konstituente oder bei weitem Fokus auf den Fokusexponenten. Unter Rekurs auf die metrische Struktur fällt der Akzent dann an die metrisch prominenteste Silbe, die Wortakzentsilbe. Diese wird für die Interpretation der Fokusstruktur einer Äußerung zur relevanten Akzent-Trägerin.

2. FOKUSZUWEISUNG, FOKUSPROJEKTION UND AKZENTDOMÄNENBILDUNG

Aus dem bereits zur Fokuszuweisung, Fokuskontrolle und zur phonologischen Materialisierung des abstrakten Fokusmerkmals Gesagten lassen sich drei Prinzipien ableiten, die zur Überführung der Oberflächenstruktur eines Satzes mit bestimmten Fokusmerkmalen in einen Satz mit einer bestimmten phonologischen Intonationsstruktur erforderlich sind:

- (6) Prinzip (I): Fokuszuweisung
Auf der Oberflächenstruktur wird einer oder mehreren Konstituenten das Fokusmerkmal **F** zugewiesen.
- (7) Prinzip (II) Phonologische Materialisierung:
Phonologisches Korrelat von **F** ist mindestens ein Akzent, der als **Akzentton (AT)** bezeichnet wird, weil die Grundfrequenz, der F_0 -Verlauf, sein wichtigstes phonetisches Korrelat ist.
- (8) Prinzip (III) Metrische Struktur und Fokusstruktur:
Der Akzentton wird auf die metrisch prominenteste Silbe des Fokusbereichs projiziert. Bei Fokussierung komplexer Phrasen ermitteln Fokusregeln zunächst den Fokusexponenten, und der AT fällt an dessen Wortakzentsilbe. Diese wird durch die AT-Zuteilung zur **Fokussilbe (Σ^*)** (weiter Fokus und Fokusprojektion). Bei fokussierten Silben oder Wörtern ist diese Silben bzw. ihre Wortakzentsilbe³ (σ^*) zugleich die Fokussilbe (Σ^*).

Fokusregeln - Regeln, die Akzente auf bestimmte Konstituenten lenken - sind nun verschiedentlich formuliert worden. Die Nuclear Stress Rule (Chomsky 1971), die den Akzent auf die rechteste Konstituente des Fokusbereichs lenkt, ist aber zumindest für Verbal- und Präpositionalphrasen des Deutschen problematisch und führt zu falschen Voraussagen. Die Beispiele (9) und (10) sollen das verdeutlichen:

- (9) F: Warum_i hat sie ihn noch nicht verlassen?
(i) A: * [_{F_i} der Kinder WEGen].
(ii) A: [_{F_i} wegen der KINder]. *oder* (iii) [_{F_i} Der KINder wegen].

3 Zur Bestimmung des Wortakzents im Rahmen einer metrischen Theorie vgl. Selkirk (1984) sowie für das Deutsche Féry (1986). Wortakzente erhalten im metrischen Gitter einen Schlag auf der zweiten Ebene.

- (10) F: Warum_i bist du wütend auf Karl?
- (i) A: * Weil er [_{F_i} an meinem Geburtstag nach Berlin FÄHRT].
- (ii) A: Er [_{F_i} fährt an meinem Geburtstag nach BerLIN]. *oder*
- (iii) Weil er [_{F_i} an meinem Geburtstag nach BerLIN fährt].

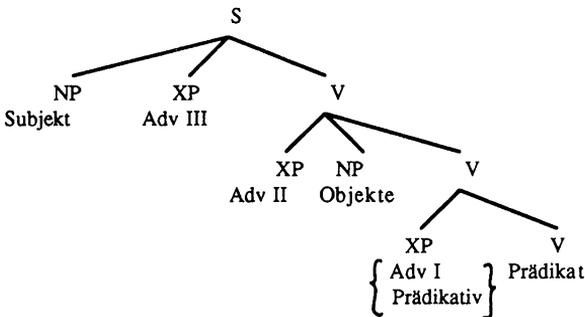
Auch Selkirks (1984: 207) Fokusregeln, die auf die Prädikat/Argument-Struktur Bezug nehmen, liefern in vielen Fällen nicht die richtigen Ergebnisse.⁴

- (11) (I) Fokus-Akzenton-Regel (Basic Focus Rule)
 "A constituent to which a pitch accent is assigned is a focus."
- (11) (II) Fokusprojektionsregel (Phrasal Focus Rule)
 "A constituent may be a focus if (i) or (ii) (or both) is true:
 (i) The constituent that is its *head* is a focus.
 (ii) A constituent contained within it that is an *argument* of the head is a focus."

Nach Selkirks Regeln (11)(III) müsste Fokusprojektion in (10i) möglich sein. In v.Stechow & Uhmann (1986) wurde versucht, Fokusregeln zu formulieren, die auch Beobachtungen von Contreras (1976), Höhle (1982) sowie Haftka (1980) und Pheby (1980) gerecht werden. Betrachtet wurde dort die Fokusprojektion in der Verbalphrase⁵; sie wurde in Zusammenhang gebracht mit der normalen linearen Ordnung (NLO) und einer rhematischen Hierarchie.

- (12) Rhematische Hierarchie (Pheby 1980 : 871):
 Adv III < V.Zus. < Verb < Subj. < Adv. II < Adv. I < Objekte < Prädikativ
Temp. Instr. Direkt. Loc.

- (13a) NLO (Vereinfachte Version nach Haftka 1980 und Pheby 1980):



4 Eine ausführliche Kritik an Selkirk (1984) findet sich in von Stechow & Uhmann (1984, 1986).

5 Fokusregeln für verbale und nominale Projektionen werden in Uhmann (1987MS) entwickelt.

Als Test zur Feststellung der normalen linearen Ordnung wird von Haftka (1980) die Bildung von Infinitivgruppen vorgeschlagen. Ihre NLO-Definition⁶ lautet:

(13b) NLO-Definition (Haftka 1980: 707)

NLO liegt dann vor, wenn die Infinitivgruppen

- " (a) keine Vorerwähnung voraussetzen und
- (b) nur einen einzigen Intonationsschwerpunkt haben, der
- (c) am Ende der Infinitivgruppe (...) liegt."

Fokusprojektion in der Verbalphrase kann also von der Erfüllung der beiden Bedingungen unter (14) abhängig gemacht werden.

(14) Fokusprojektion in der Verbalphrase

- (i) NLO ist gegeben.
- (ii) Die als Fokusexponent (AT-Träger) gewählte Konstituente ist das stärkste Element in der rhematischen Hierarchie.

Ein Beispiel:

(15) Warum_i ist Ede verletzt?

Er [_F wollte mit einer Hacke ein LOCH in eine Wand schlagen].
 Sbj. Adv.II dir. Obj. Adv. I Verb

Fokusprojektion ist möglich, weil der Akzent von den vorhandenen Elementen auf das rhematisch stärkste gelenkt wurde, außerdem ist NLO gegeben. Die Sätze (16)(a-c) erlauben hingegen keine Fokusprojektion, weil der Fokusexponent nicht das rhematisch stärkste Element ist. Satz (d) erfüllt zwar (14ii), Fokusprojektion ist aber nicht möglich, weil eine Abweichung von der NLO vorliegt.

- (16) (a) er wollte mit einer Hacke ein Loch in eine Wand schlagen
- (b) er wollte mit einer Hacke ein Loch in eine WAND schlagen
- (c) er wollte mit einer Hacke ein Loch in eine Wand SCHLAgen
- (d) er wollte ein LOCH mit der Hacke in eine Wand schlagen

Die Abweichung von der NLO in Satz (16d) kann zusätzlich zu dem von Haftka vorgeschlagenen Test dadurch gezeigt werden, daß bei dieser Reihenfolge der Stellungsglieder durch einen einzigen Akzentton stets nur die enge Fokussierung der betroffenen Konstituente erzielt wird:

- (16) (d) [_F ein LOCH] mit einer Hacke in eine Wand schlagen
- (e) ein Loch [_F mit einer Hacke] in eine Wand schlagen
- (f) ein Loch mit einer Hacke [_F in eine WAND] schlagen

6 Besonders die Punkte (b) und (c) könnten von besonderem Interesse sein. (b) wegen der Korrelation zwischen NLO und Anzahl von Akzenttönen, und (c) weil sich hier ein Phänomen andeutet, das über die Existenz einer Akzentposition möglicherweise auch auf die Existenz einer Fokusposition innerhalb der VP hinweist.

Die Fokusprojektionsregel (14) und Phebys rhematische Hierarchie (12) treffen aber auf Probleme bei Sätzen wie (17) bis (21), bei denen (14ii) verletzt, Fokusprojektion aber trotz Akzent auf einer rhematisch schwächeren Konstituente möglich ist.

- (17) Ich glaube, [F daß vielen Asthmatikern eine KNEIPPkur helfen würde].
 (18) [F Ein Junge GEIGT].
 (19) [F Vater KOCHT noch].
 (20) Ich glaube, [F daß einem Kind ein FÜLLer gestohlen wurde].
 (21) [F Weil einem Beamten ein FEHLer unterlaufen ist].

Andererseits demonstrieren Sätze wie (18)' und (19)' eine korrekte Akzentplatzierung nach (14ii).

- (18') [F Ein MÜLLwagen kommt].
 (19') [F Das WASser kocht].

Fokusprojektion ist jedoch in den Fällen (17), (20),(21), (18)' und (19)' in Einklang mit Regel (14) möglich, wenn man von der Annahme ausgeht, daß hier die von der Wahl des Verbs abhängige thematische Rollenstruktur der Verbargumente berücksichtigt werden muß, und die Fokusexponenten auf der für die Analyse der Fokusstruktur relevanten Ebene als 'Objekte'⁷ zählen, weil sie mit objektspezifischen thematischen Rollen (Nicht-Agens) versehen sind. Die Sätze (18) und (19) hingegen verstoßen gegen Phebys rhematische Hierarchie, weil Fokusprojektion trotz Akzentton auf dem rhematisch schwächeren Element, dem Prädikat, möglich ist. Dieser Datenlage könnte eine reformulierte rhematische Hierarchie gerecht werden, die die nicht-agentiven Subjekte der Gruppe der Objekte zuordnet und die agentiven Subjekte schwächer als das Prädikat einstuft.

(12)' Reformulierte rhematische Hierarchie:

Adv. III < Subjekt < Prädikat < Adv. II < Adv. I < Objekte < Prädikativ
 (agentiv) (+ nicht-agent. Subj.)

Bisher wurden nur Sätze vorgestellt, in denen ein einziger Akzentton entweder für die enge Fokussierung dieser Konstituente verantwortlich war, oder - falls Regel (14) erfüllt war - den Fokusexponenten eines Syntagmas markierte. Das Prinzip (II) war allerdings schon so formuliert, daß als phonologisches Korrelat des F-Merkmals mindestens ein Akzentton fungieren sollte. Die Beschreibung mehrfacher Akzentton-Zuweisung erfolgt in Anlehnung an Fuchs

7 Ergativ-Hypothese (vgl. Burzios (1981) und Williams (1981) Unterscheidung zwischen internen und externen Argumenten sowie von Stechow & Uhmman (1984, 1986). Diese Erklärung der Akzentuierungsunterschiede zwischen den Sätzen (18) und (18)' sowie (19) und (19)' ist aber nicht unproblematisch. Sie erfordert ja die Zuordnung des nicht-agentivischen Subjektes zur Kategorie der Objekte. Der Objektivstatus allein bringt aber keine Erklärung für das unterschiedliche Akzentuierungsverhalten, da auch die Gruppe der Objekte kein einheitliches Akzentuierungsverhalten gegenüber verschiedenen Verben zeigt (Vgl. Uhmman 1987MS).

(1976, 1980) und Gussenhoven (1983). Für die Zuteilung von Akzenttönen ist die Information relevant, ob - und wenn ja in wieviele **Akzentdomänen (ADs)** - der Fokusbereich eines Satzes aufgeteilt werden soll. Bei Nicht-Unterteilung wird das F-Merkmal durch einen einzigen Akzentton realisiert, ansonsten gibt es soviele Akzenttöne wie Akzentdomänen. Man könnte dieses Phänomen auch nach Fuchs als integrierende Akzentuierung (ein Akzentton) bzw. als isolierende Akzentuierung (mehrere Akzenttöne) bezeichnen. Primär unterschieden werden muß also zwischen:

- (22) (i) Domänenzerlegung: Akzentdomänen - Regel für isolierende Akzentuierung
Semantische Konstituenten (Prädikate, Ergänzungen und Angaben) bilden eigene, separate Akzentdomänen; ihnen wird pro AD ein Akzentton zugeteilt.
- (22) (ii) Akzentvererbung: Akzentdomänen - Regel für integrierende Akzentuierung
Semantische Konstituenten verschmelzen zu einer Akzentdomäne. Die Ermittlung und die Akzenttonzuweisung an den Fokusexponenten erfolgt nach der rhematischen Hierarchie.

Das Zusammenwirken von metrischer Prominenz und Akzentdomänenbildung wird durch das Prinzip (IV) geregelt:

- (23) Prinzip (IV) Prinzip der Wohlgeformtheit für Intonationsphrasen:
- (a) Eine Konstituente mit einem Akzentton ist prominenter als eine Konstituente ohne Akzentton (mindestens drei Schläge).
- (b) Bei mehreren Konstituenten mit jeweils eigenem Akzentton ist die am weitesten rechts stehende Konstituente prominenter als die übrigen Akzentton-Konstituenten (Nuclear Stress Rule) (vier Schläge).
- (c) Bei weitem Fokus muß der Fokusexponent (bei engem Fokus die Fokuskonstituente) die prominenteste Konstituente der gesamten Intonationsphrase sein.

Dieses obligatorische Prinzip ähnelt auf den ersten Blick stark der 'Pitch Accent Prominence'-Regel von Selkirk (1986:265ff). Es ordnet gleichfalls dem Begriff der Prominenz die abstrakte metrische Stärke (Schläge auf einem metrischen Gitter) sowie die tonale Realisierung (Akzenttonzuweisung) zu. Ebenso wie bei Selkirk wird auch im Prinzip (IV) ein 'pitch accent first'-Ansatz vertreten, da sich Prominenzrelationen zwar für Wörter und kurze Phrasen aber nicht für komplexe Phrasen und lange Sätze auf der Ebene rein abstrakter metrischer Relationen festlegen lassen. Deshalb ist eine der endgültigen Ausbuchstabierung der Prominenz vorgängige Entscheidung darüber, in wieviele Akzentdomänen mit zugehörigen Akzenttönen die Phrase oder der Satz aufgeteilt werden sollen, unbedingt erforderlich, um präzise Aussagen über ihre/seine intonationsphonologische Gestalt machen zu können. Der wichtige Unterschied zu Selkirks Betrachtungsweise liegt aber in der Einführung des Konzeptes der Akzentdomäne, da so vermieden werden kann, von der problematischen Eins-zu-eins-Beziehung zwischen Akzenttönen und Foki ausgehen zu müssen (vgl.(11)(I)). Ebenfalls bei Selkirk nicht vorhanden ist die Beschränkung (IV)(c), die in Interaktion mit den Prinzipien (a) und (b) sowie der fakultativen Regel (22) alle möglichen Akzenttonpositionen der Intonationsphrase bestimmt, ohne daß ihr Aufbau (z.B. in prä- und postnukleare Abschnitte) von einer Intonationsgrammatik des Deutschen stipuliert werden müßte. Für die Regeln zur Akzentdomänenbildung (22) muß allerdings (abweichend von Gussenhoven 1983) noch festgelegt werden, daß sie erst dann zur

Anwendung kommen, nachdem die Fokusregeln bereits den Fokusexponenten (Fettdruck) bestimmt haben (vgl. auch Uhmans 1987MS).

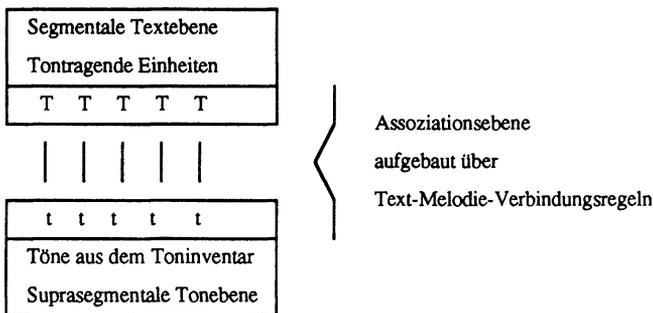
- (20) (a) Ich glaube, [F [AD daß einem Kind ein **FÜLler** gestohlen wurde]].
 (20) (b) Ich glaube, [F [AD1 daß einem **KIND**] [AD2 ein **FÜLler** gestohlen wurde]].
 (20) (c)* Ich glaube, [F [AD1 daß einem **KIND**] [AD2 ein **FÜLler**] [AD3 ge**STOH**len wurde]].
 (18) (a) [F [AD Ein Junge **GEIGT**]].
 (18) (b) [F [AD1 Ein **JUN**ge] [AD2 **GEIGT**]].
 (18) (a) [F [AD Ein **MÜLL**wagen kommt]].
 (18) (b)* [F [AD1 Ein **MÜLL**wagen] [AD2 **KOMMT**]].

In den Fällen (20)(c) und (18)(b) ist die jeweils notierte isolierende Akzentuierung ausgeschlossen, da die mit dieser Akzenttonzuweisung verbundene obligatorische Anwendung von (IV)(b) das Prominenzerhaltungsprinzip für Fokusexponenten (IV)(c) verletzt. (20)(b) und (18)(b) sind hingegen zulässige isolierende Varianten der integrierenden Akzentuierung in (20a) und (18)(a), da hier die Interaktion der Prinzipien (IV)(b) und (c) mit der Akzentdomänenbildung nicht zu Verletzungen der metrischen Wohlgeformtheitsbedingungen führt.

3. DER AUFBAU DER INTONATIONSSTRUKTUR

Die Intonationsstruktur wird durch die metrische und die tonale Struktur einer Phrase gebildet. Zur tonalen Struktur gehören die Intonationsgrammatik mit dem sprachspezifischen Toninventar, die Phrasierungsregeln zur Bildung von Intonationsphrasen und die Text-Melodie-Verbindungsregeln zur Assoziation von Tönen und Text. Der Aufbau der Intonationsstruktur steht damit in der Tradition der nicht-linearen, autosegmentalen Phonologie.

(24) Tonale Struktur



An dieser Darstellung lassen sich drei Analysekomplexe festmachen: (1.) Wie werden Töne und Text synchronisiert? (2.) Welches sind die tontragenden Einheiten der Textebene? (3.) Wie sieht das Toninventar für das Deutsche aus?

3.1 Die Synchronisierung von Ton und Text

In dem Moment, in dem tonale Merkmale auf einer von den segmentalen Merkmalen abgetrennten, unabhängigen Ebene untergebracht werden, ist die Verbindung, die Assoziation, von entscheidender Bedeutung. Da die Fragen (2) und (3) erst später beantwortet werden sollen, wird im folgenden zunächst weiter nur von Einheiten der Tonebene (repräsentiert durch t's) und von Einheiten der Textebene (repräsentiert durch T's) die Rede sein.

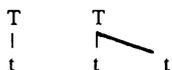
Im einfachsten Fall einer Ton-Text-Assoziation könnten die Segmente direktional (z.B. von links nach rechts) in einer Eins-zu-eins Relation aufeinander abgebildet werden.

(25) Eins-zu-eins Relation

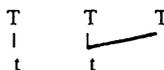


Der Einflußbereich eines Elements kann aber auch den Einflußbereich eines anderen Elements auf der anderen Ebene über- bzw. unterschreiten. Geht man von der tonalen Ebene aus, so sind zwei weitere Möglichkeiten der Ton-Text-Assoziation zu unterscheiden. Quetschung würde dann zustandekommen, wenn in einer Phrase mehr Töne als tontragende Elemente vorhanden sind und die Töne zugeordnet werden müssen. Dehnung würde in dem Fall zu fordern sein, wenn die Anzahl der tontragenden Einheiten die der Töne überschreitet.

(26) Quetschung (enger Einflußbereich)



(27) Dehnung (weiter Einflußbereich)



Welche dieser drei Assoziierungskonventionen wann Anwendung findet, ist eine empirisch und einzelsprachlich zu beantwortende Frage. Mögliche universelle Prinzipien sind natürlich nichtsdestoweniger von entscheidender Bedeutung. Ein erster Lösungsvorschlag wurde von Goldsmith (1976) mit der Well-formedness Condition (WFC) entwickelt.

(28) Goldsmiths Well-formedness Condition

- (1) Jeder Ton wird mit mindestens einem tontragenden Element assoziiert.
- (2) Jedes tontragende Element wird mit mindestens einem Ton assoziiert.
- (3) Assoziationslinien dürfen sich nicht kreuzen.

Um auch bei nicht-identischer Anzahl von Tönen oder tontragenden Elementen eindeutige Assoziation zu gewährleisten, wurde die WFC durch eine Tone-Linking-Regel ergänzt.

(29) Tone-Linking-Regel

- Töne werden von links nach rechts in der Eins-zu-eins Relation assoziiert, bis es entweder keine Töne oder keine tontragenden Einheiten mehr gibt.
- (1) und (2) der WFC wird dann über Dehnung oder Quetschung Genüge getan.

An diesen Prinzipien ist in bezug auf ihre universelle Gültigkeit und Gleichgewichtigkeit detaillierte Kritik geübt worden (vgl. Williams 1976, Halle & Vergnaud 1982, van der Hulst 1984), auf die an dieser Stelle aber nicht weiter eingegangen werden kann. Hier soll vielmehr der Versuch unternommen werden, eine Adaption der WFC für das Deutsche zu entwickeln, die den Besonderheiten einer Intonationssprache gerecht wird. Für dieses Unterfangen soll das ebenfalls von Goldsmith formulierte Akzentprinzip zur Ton-Text-Assoziation in Intonations- oder Akzentsprachen den Ausgangspunkt der Analyse bilden. Das Akzentprinzip läßt sich in zwei Subprinzipien aufteilen:

(30) Das Akzentton-Zuteilungsprinzip (tone assignment)

Einer Folge von Silben wird zunächst eine Folge von Tönen zugeteilt, dabei wird sowohl eine Silbe als auch ein Ton besonders markiert (im folgenden durch einen Stern (*))

T	T	T*	T
t	t	t*	t

(31) Das Akzentton-Verbindungsprinzip (tone association)

Die Tonverbindung nach dem Akzentprinzip assoziiert zunächst immer den besonders markierten Ton mit der besonders markierten Silbe.

T	T	T*	T
t	t	t*	t

(32) Die Verbindung der verbleibenden Töne und Silben

Diese erfolgt für den gesamten Bereich nach der WFC (28) und der Tone-Linking-Regel (29).

3.2 Die tontragenden Einheiten der Textebene

Das Akzentprinzip enthält nun, wenn seine Gültigkeit für das Deutsche zunächst einmal angenommen wird, schon eine Antwort auf die zweite Frage: Was sind die tontragenden Einheiten des Deutschen? Die tontragenden Einheiten sind Silben. Bei einer Folge von Silben gibt es noch mindestens eine auf besondere Weise ausgezeichnete Silbe; diese kann auf der Wort- und Konstituentenebene als die metrisch prominenteste Silbe des Wortes, als **Akzentsilbe** (σ^*)⁸ identifiziert werden. Auf der Phrasen- und Satzebene handelt es sich um die metrisch prominenteste Silbe des von Fokusregeln ermittelten Fokusexponenten, um die **Fokussilbe** (Σ^*) (in anderen Theorien der **Nukleus**), die prominenteste Silbe der Intonationsphrase.

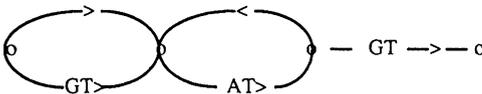
⁸ Diese Silben sind bei der Unterteilung größerer Fokusbereiche in Akzentdomänen potentielle Akzentträger.

3.3 Das Toninventar des Deutschen

Operierend auf der Distinktion zwischen Hoch-Tönen und Tief-Tönen (H und T) wurden aus den F_0 -Konturen eines Korpus von circa 2000 Testsätzen vier verschiedene und funktional distinktive⁹ **Akzenttöne** extrahiert: zwei bitonale, kinetische Töne und zwei statische Leveltöne. Bei den kinetischen Tönen handelt es sich um einen steigenden und einen fallenden Ton, die als Sequenzen von Leveltönen notiert werden: TH und HT. Fügt man nun noch die vom Akzentprinzip (24) geforderte besondere Markierung eines Tons hinzu, so kommt das folgende Akzenton-Repertoire zustande: **H*+T**, **T*+H**, **T*** und **H***¹⁰. Das durch (*) gekennzeichnete tonale Element wird im Folgenden als tropischer Ton bezeichnet werden. Darüberhinaus hat sich aber noch die Annahme eines weiteren, von dem Akzentton distinktiven Tontyps¹¹, des **Grenztons (H%, T%)**, analytisch bewährt. Die Unterscheidung zwischen diesen beiden Tontypen ist primär funktional motiviert. Während der Akzentton als phonologisches Korrelat des F-Merkmals fungiert, ist der Grenztton das phonologische Korrelat der Phrasierungsregeln. Dabei scheint es so zu sein, daß die phrasenfinale Markierung obligatorisch (Modusmarkierung), die initiale hingegen fakultativ ist.

Die Matrix (34) stellt das Tonsystem für das Deutsche mit den eingeführten Funktionstypen und ihren phonologischen Merkmalen (tonale und metrische Charakterisierung) sowie eine angenäherte phonetische Realisierung und (33) den Aufbau der Intonationsphrase dar.

(33) Intonationsphrase



9 Ihre funktionale Distinktivität läßt sich in der Vorfeld-Position zeigen (vgl. Uhmans 1987 MS).

10 Theoretisch möglich wären natürlich auch H+T* und T+H*, die als tonale Entsprechungen von Isacenko & Schädlich (1966) praktisch fallendem bzw. steigendem Tonmuster (vgl. auch Löscher 1983) zu interpretieren wären, sowie H*+H (bzw. H+H*) und T*+T (bzw. T+T*).

11 Keine Anwendung findet der Phrasenton, der von Pierrehumbert (1980) für das Amerikanische identifiziert wurde.

(34) Tonsystem - Matrix

Funktionstypen	Phonologische Merkmale	Phonetische Realisierung
Akzentton Phonologisches Korrelat des Fokus- Merkmals	H*+T	fallender Fo-Verlauf durch die akz. Silbe, bezieht nachfolgende unakz. Silben ein
	T*+H	steigender Fo- Verlauf durch die akz. Silbe bezieht nachfolgende unakz. Silben ein
	H*	Fo-Gipfel im Nukleus der akz. Silbe, kein Einfluß auf unakz. Silben
	T*	Fo-Tal im Nukleus der akz. Silbe, kein Einfluß auf unakz. Silben
Grenztton Phonologisches Korrelat der Phrasierungsregeln	T%	sprecherindividuelles Fo-Minimum (baseline) an der/den Phrasengrenzen
	H%	Fo-Werte über dem Minimum und dem sprecherindiv. mittleren Niveau
Akzentsilbe metrisch prominenteste Silbe eines Wortes	σ^* : mind. zwei Schläge: keine tonale Markierung bei drei Schlägen: H*+T, T*+H, H*, T*	(Markierung über Intensität, Dauer) Fo-Veränderungen
Fokussilbe metrisch prominenteste Silbe des Fokusexponenten	Σ^* : mind. drei Schläge bei vier Schlägen: H*+T oder T*+H	(Markierung über Intensität u. Dauer) fallender oder steigender Fo-Verl.
Grenzsilbe von Phrasierungs- Regeln ermittelte initiale und finale Silbe der Intonationsphrase	$\sigma\%$: keine besonderen metrischen Merkmale Tonale Markierung final obligatorisch: T% oder H% inital fakultativ: H%, T% oder $\emptyset\%$	<hr/> T% = sprecherindiv. Fo-Minimum H% final = Werte über Fo-Minimum H% initial = Werte über dem sprecher- indiv. mittleren Ansatzniveau $\emptyset\%$ = mittleres Ansatzniveau

4. DIE ASSOZIATION VON TÖNEN UND TEXT IM DEUTSCHEN

Im Rahmen der Entwicklung eines Intonationssystems für das Deutsche bleibt ein wesentliches Problem noch ungelöst, das in in diesem Abschnitt behandelt werden soll. Die Text-Ton-Assoziation wurde bisher nur über die allgemeinen Prinzipien der Eins-zu-eins-Verbindung, der Dehnung und Quetschung sowie über das Akzentprinzip vorgestellt. Assoziationkonventionen für das Deutsche sind das noch zu lösende Problem im Rahmen einer autosegmentalen Intonationsbeschreibung. Die folgenden Beispiele sollen zur Entwicklung und Demonstration der Assoziationsprinzipien des Deutschen dienen:

(35) (a) Xe*nia (b) Fe*rnseher (c) Freu*nd

(1.) Akzenton-Zuteilung
 (a) Xe*nia (b) Fe*rnseher (c) Freu*nd
 H*+T H*+ T H*+T

(2.) Akzenton-Verbindung
 (a) Xe*nia (b) Fe*rnseher (c) Freu*nd
 | | |
 H*+ T H* + T H*+T

(3.) Verbindung der verbleibenden Töne und Silben
 (a) Xe*nia (b) Fe*rnseher (c) Freu*nd
 | | | |
 H*+T H* + T H*+T

Für (35)(b) ist diese Ableitung aber nicht die einzige Möglichkeit, der WFC zu folgen. Zugelassen wäre auch:

(35) (b') Fe*rnseher
 / \
 H*+ T

Beziehungweise bei nicht-wortinitialer Akzentuierung:

(36) (a) Mari*a (b) Veronika
 \ \ / \
 H*+T H*+T

Aufgrund empirischer Beobachtungen hat bereits Goldsmith (1981) die Ableitung (35) (b') ausschließen wollen. (36) (a und b) würden bei ihm in dieser Form nicht zustandekommen, weil er die Tonmelodie MH*T als die zugrundeliegende Kontur für die Produktion isolierter Wörter im Deklarativmodus ansetzt. Muß man ihm also folgen und das Tonrepertoire um tritonale, kinetische Töne erweitern? Zwei Überlegungen sprechen gegen eine solche Lösung.

Zum einen kann (35)(b') dadurch ausgeschlossen werden, daß die Dehnungs-Konvention dahingehend modifiziert wird, daß nicht-tropische Töne tropische in Bezug auf die Dehnung dominieren, daß also Zuweisungsambiguitäten immer zugunsten der Dehnung von nicht-tropischen Tönen gelöst werden müssen. Dazu müßte das Prinzip (1) der WFC um die Unterscheidung zwischen tropischen und nicht-tropischen Tönen erweitert werden. Das Prinzip (1) der WFC könnte dann also lauten:

- Alle tropischen Töne werden mit genau einer Silbe, mit der Akzent- bzw. Fokussilbe, verbunden. (Dehnung verboten.)
- Alle nicht-tropischen Töne müssen mit mindestens einer Silbe assoziiert werden. (Dehnung gefordert, wenn mehr Silben als Töne zur Verfügung stehen.)

Zum anderen kann die tonale Unterspezifizierung, die aus dem umformulierten Prinzip (1) für die Beispiele (36)(a,b) folgt, dahingehend aufgelöst werden, daß statt tritonaler Akzenttonzuweisung eine M-Default Regel angenommen wird - eine Lösung, die auch Leben (1976) vorgeschlagen hat.

(37) M-Default Regel (Leben)

Nachdem (1) und (2) der WFC erfüllt worden sind, erhalten die unakzentuierten Silben zur Linken der Fokus- oder Akzent-Silbe, die noch keinen Ton tragen, einen M-Ton.

- (36) (a) Mari*a (b') Ver*nika
 | | \ | | \
 M H*+T M H*+T

Daß auch diese Lösung problematisch ist, soll später motiviert werden. Zunächst ist an dieser Stelle ein anderes Problem relevant, das auch bereits von Goldsmith im Zusammenhang mit der tonalen Realisierung von Akzentsilben angesprochen wurde. In den Beispielen (35)(b') und (36)(a u. b) wurde Links- und Rechtsdehnung des H*-Tons ausgeschlossen. Die Rechtsdehnung des nicht-tropischen Tons in (35)(b) und die Linksquetschung des gleichen Tons in (35)(c) waren aber die einzigen Möglichkeiten, das Prinzip (1) der WFC zu erfüllen; sie wurden deshalb zugelassen. Linksquetschung nicht-tropischer Töne auf bereits tontragende Akzentsilben beschreibt auch Goldsmith (1981) als eine tonale Realisierungsmöglichkeit, die selbst dann Anwendung finden kann, wenn Silben für die Realisierung dieses Tons zur Verfügung stehen, die aber die WFC nicht auf minimale Weise erfüllt, sicher nicht universal ist, sondern von der Tongrammatik einer Sprache besonders stipuliert werden muß. Die Tonzuweisung in den bisher vorgestellten Beispielen würde nach dieser Stipulierung aussehen wie unter (38)(a-e):

- (38) (a) Xe*nia (b) Fe*rnscher (c) Freu*nd (d) Mari*a (e) Ver*nika
 | | \ | | \ | | \ | | \ | | \
 H*+T H* +T H*+T M H*+T MH*+T

Meiner Meinung nach gibt es gute Gründe, eine solche Tonzuweisung für das Deutsche festzulegen. Tonbewegungen (nicht Tonbrüche) auf den Akzent- und Fokussilben sind auffällige Eigenschaften deutscher Intonationskonturen. Eine solche Analyse hätte neben möglicher größerer deskriptiver Adäquatheit den weiteren Vorteil, daß sich die Aufnahme der Unterscheidung zwischen tropischen und nicht-tropischen Tönen in das Prinzip (1) der WFC im Prinzip (2) in der Unterscheidung zwischen akzentuierten Silben (Akzent- bzw. Fokussilben (σ^* und Σ^*) und unakzentuierten Silben (σ) wiederholen und sich ebenfalls in unterschiedlichen Assoziationskonventionen manifestieren würde. Das Prinzip (2) der WFC könnte dann also folgendermaßen lauten:

- Alle Σ^* und σ^* -Silben müssen mit mindestens einem Ton, dem tropischen, assoziiert werden. (Quetschung bei der Assoziation kinetischer Töne gefordert.)
- Alle unakzentuierten Silben (σ) erhalten maximal einen Ton. (Quetschung verboten.)

Diese Prinzipien lassen sich zu einer reformulierten WFC mit einer Stipulierung für das Deutsche zusammenfassen:

(39) Reformulierte WFC

- (1) (a) Alle tropischen Töne werden mit genau einer Silbe assoziiert, und zwar mit einer σ^* - oder Σ^* -Silbe.
- (b) Stipulierung für das Deutsche:
 t^*+t wird auf eine Silbe projiziert; d.h. es gibt eine Silbe, die sowohl mit t^* als auch mit t assoziiert wird.
- (2) Alle σ^* oder Σ^* -Silben müssen mit einem tropischen Ton assoziiert werden.
- (3) Alle nicht-tropischen Töne müssen mit mindestens einer Silbe verbunden werden.
- (4) Alle Silben erhalten einen Ton.
- (5) Assoziationslinien dürfen sich nicht kreuzen.

Für das korrekte Operieren der WFC ist darüberhinaus eine generelle Restriktion erforderlich, die die Ausdehnung der **Tondomäne** (# #) als den Bereich festlegt, in dem die Tonzuweisungsprinzipien der WFC für einen Akzentton Anwendung finden. Im Gegensatz nämlich zu Ton-sprachen, die von ihrem Toninventar überwiegend lexikalischen Gebrauch machen und deren Tondomäne zumeist das Morphem ist, muß für Akzent- bzw. Intonationssprachen die Tondomäne nach anderen Kriterien festgelegt werden. In diesen Sprachen bietet sich für die Festlegung der Tondomänengrenzen vor allem die Präsenz tonal spezifizierter Silben¹³ an. Tonal spezifiziert sind solche Silben, die über die Anwendung der Akzenttonverbindung (30) tonale Information erhalten (vgl. auch (1) und (2) der reformulierten WFC).

(40) Generelle Restriktion für Tondomänen

Tonal spezifizierte Silben (σ^* - bzw. Σ^* - und phrasenfinale Grensilben $\sigma\%$) sowie Fokusbereichsgrenzen sind Barrieren für Tondomänen.

Die Anweisungen zur Quetschung oder Dehnung von Tönen folgen damit aus den Prinzipien der reformulierten WFC und aus der Restriktion für Tondomänenbegrenzung. So folgt aus (1) und (2) der WFC, daß tropische Töne auf keinen Fall gedehnt werden dürfen; aus (1)(a), (1)(b), der Stipulation für das Deutsche, und (2), daß bestimmte σ^* - bzw. Σ^* -Silben zwei Töne erhalten, aus (1)(b) und (5), daß unakzentuierte Silben nur einen Ton erhalten können und aus (3), (4) und (5), daß nicht-tropische Töne nach rechts gedehnt werden können. Außerdem folgt aus (5), daß Silben zur Linken einer σ^* - bzw. Σ^* -Silbe von dem auf diese Silbe projizierten Akzentton kein tonales Merkmal erhalten können. (4) und (5) der WFC können in diesen Fällen nur über eine zusätzliche Tonzuweisungskonvention (z.B. M-Default (Leben)) erfüllt werden. Dasselbe gilt in Tondomänen, denen ein monotonaler Akzentton (t^*) zugewiesen wurde, auch für die Silben zur Rechten von σ^* - bzw. Σ^* - Silben. Dazu zwei Beispiele:

¹³ Tonal spezifiziert sind neben den Σ^* -Silben natürlich auch die Silben, die nach der Anwendung von Phrasierungsregeln als erste bzw. letzte Silben einer Phrase mit dem phonologischen Korrelat der Phrasierungsregeln, dem Grenzton, assoziiert werden.

(41) (a) # σ σ σ^* σ σ # Σ^* σ σ #
 T* H*+T

(b) # σ σ σ^* σ σ # Σ^* σ σ # folgt aus (1)(a) u.(2) der WFC
 | |
 T* H*+T

(c) # σ σ σ^* σ σ # Σ^* σ σ # folgt aus (1)(b)
 | / \
 T* H*+T

(d) # σ σ σ^* σ σ # Σ^* σ σ # folgt aus (1)(b),(3), (4) und (5)
 | / \ /
 T* H*+T

(e) # σ σ σ^* σ σ # Σ^* σ σ # folgt aus (4), (5) und M-Default
 / \ / \ / \ /
 M T* M H*+T

(42) (a) # σ σ^* σ σ # σ^* σ σ # Σ^* σ σ σ σ #
 T*+H H* H*+T

(b) # σ σ^* σ σ # σ^* σ σ # Σ^* σ σ σ σ # folgt aus (1)(a) u. (2)
 | | |
 T*+H H* H*+T

(c) # σ σ^* σ σ # σ^* σ σ # Σ^* σ σ σ σ # folgt aus (1)(b)
 / \ | / \
 T*+H H* H*+T

(d) # σ σ^* σ σ # σ^* σ σ # Σ^* σ σ σ σ # folgt aus (1)(b), (3) , (4) und (5)
 / \ / \ | / \ / \ /
 T*+H H* H*+T

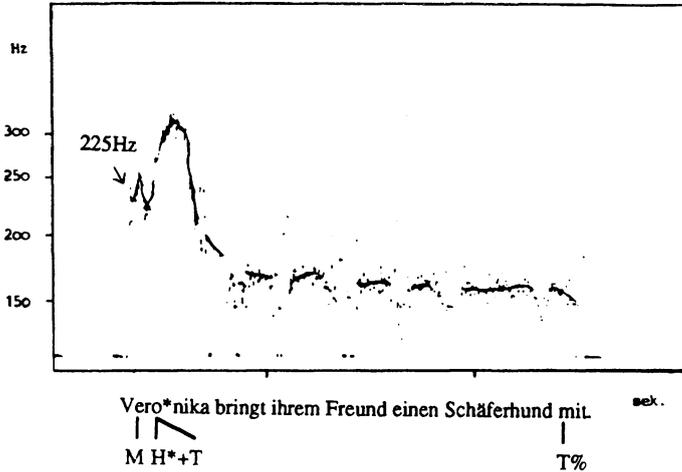
(e) # σ σ^* σ σ # σ^* σ σ # Σ^* σ σ σ σ # folgt aus (4) und (5) bei
 | / \ / \ | / \ / \ / \ /
 M T*+H H* M H*+T Anwendung von M-Default

Kommen wir aber noch einmal zur M-Default-Tonzuweisung zurück. Die Beispiele (43) und (44), F_0 -Konturen aus einem Testkorpus zur Fokusprojektion, scheinen auf die Notwendigkeit hinzuweisen, einen M-Ton in das Tonrepertoire aufzunehmen, damit der

phraseninitialen Silbe das mittlere Tonniveau, das in der Gesamt-F₀-Kontur aufscheint, zugewiesen werden kann.

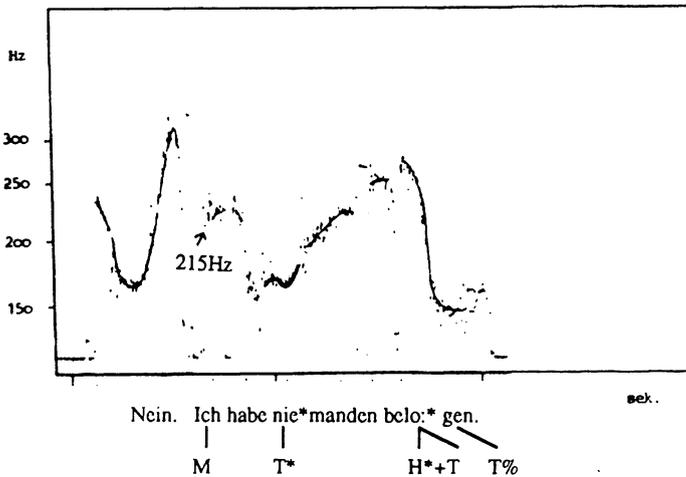
(43) VII 203 GL 52

Wer bringt seinem Freund einen Schäferhund mit?



(44) VII 176 GL 141

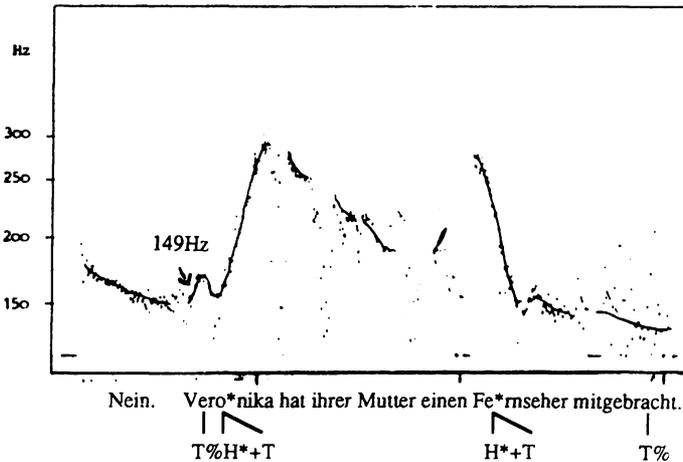
Haben sie noch etwas hinzuzufügen?



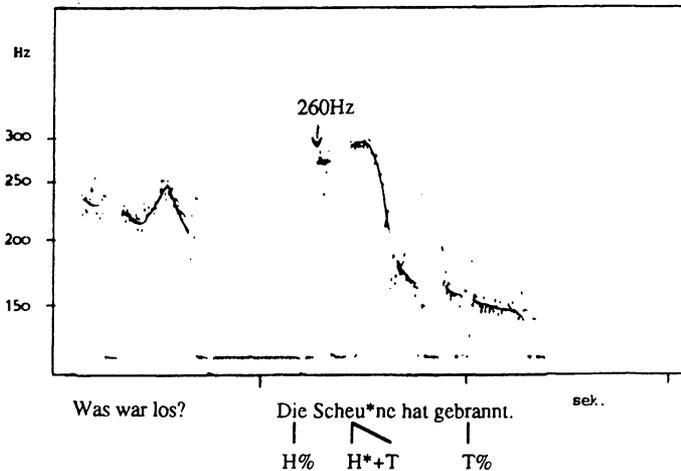
Betrachtet man jedoch alle Testsätze der Sprecherin GL, ca. 230, so stellt sich heraus, daß sie fast alle ihre Äußerungen in einem Frequenzbereich zwischen 185 Hz und 230 Hz ansetzt. Die Ausnahmen sind deutliche Abweichungen nach oben bzw. nach unten.¹⁴

(45) VII 125 GL 69

Ich habe gehört, Veronika hat ihrer Mutter einen Schäferhund mitgebracht.

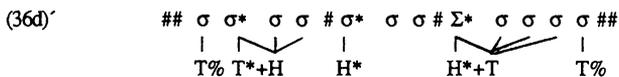
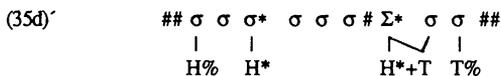


(46) VII 165 GL 7



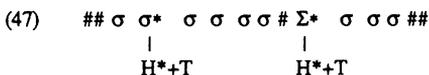
14 Tiefe bzw. hohe initiale Grenzöne verlangen in der Regel erste Akzentöne mit einem identischen tropischen Ton (T% - T*+H, H% - H*+T). Das Verhältnis zwischen dem letzten Akzenton und dem finalen Grenzton ist hingegen komplementär (H*+T - T%, T*+H - H%).

Bestimmen die Phrasierungsregeln noch die Grenzen der Intonationsphrasen (## ##), sodaß die obligatorische finale und die fakultative initiale Grenztonzuweisung erfolgen kann,

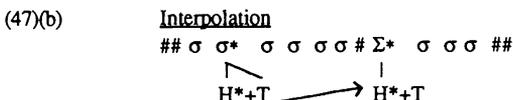
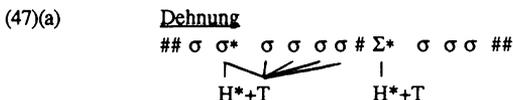


dann könnten diese Ableitungen mit den tonal markierten und den tonal unmarkierten Silben die für das Deutsche bei dieser Silben-, Grenz- und Akzentton-Struktur phonologisch relevante Toninformation zeigen.

Ein letztes Problem ergibt sich aber dann, wenn man in den Ableitungen (35d) und (36d)´ die letzte Tondomäne (Entsprechendes gilt natürlich auch für nicht-letzte Tondomänen) unter dem Aspekt der phonologischen Relevanz tonal kodierter Information noch einmal genauer betrachtet: In geeigneten Kontexten (bei einer großen Anzahl von Silben zur Rechten von Σ* bis zur nächsten Tondomänengrenze) wird über (1) und (4) der WFC die Dehnung eines nicht-tropischen Tons (falls vorhanden) über den gesamten Bereich hinweg erzwungen. Eine Alternative wäre, auch diesen Silben über Interpolation und nicht über tonale Markierung intonatorische Gestalt zu verleihen, indem unter Rekurs auf die phonologische Relevanz tonale Markierung auf Akzent- und Fokussilben sowie Grenzsilben beschränkt würde (Reformulierung von (4) der WFC). Diese Beispiele zeigen zwar das Problem, aber geben keinen Hinweis auf die Assoziationsmechanismen, da die Konturen in beiden Fällen flach wären. Konstruieren wir also ein anderes Beispiel:

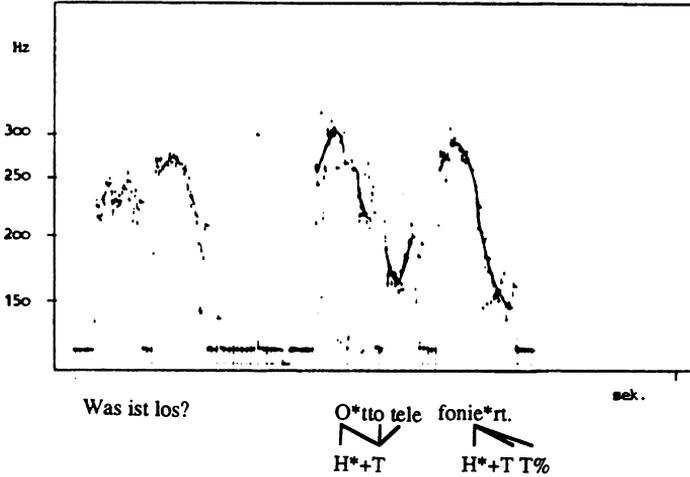


Relevant für das Argument ist hier die erste Tondomäne. Entsprechend der beiden Hypothesen sind die folgenden Ableitungen denkbar.



(47a) würde eine fallende Kontur zwischen den beiden Akzenttönen erzeugen, (47b) hingegen eine steigende. Einer Entscheidung zugunsten der einen oder der anderen Lösung kann uns eine Kontur, in der diese beiden Akzenttöne verwendet wurden, näherbringen.

(48) VII 54 GL 126



Folgt man diesem Beispiel (und auch andere Testsätze aus dem schon genannten Korpus legen diesen Schluß nahe), so muß eine dritte Lösung für die konkurrierenden Ansprüche von Dehnung und Interpolation gefunden werden: es wird gedehnt - aber nicht bis zur nächsten gestemten Silbe; es wird interpoliert - aber nicht von σ^* zu σ^* .

(48) O* tto te le fo nie*rt.
 $\sigma^* \sigma \sigma \sigma \Sigma^*$
 H*+T → H*+T T%

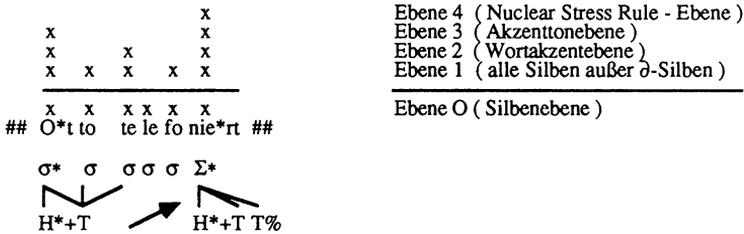
Bei dieser "gemischten" Lösung, muß die Restriktion für Tondomänen präzisiert und um eine empirische Generalisierung (basierend auf der metrischen Struktur) erweitert werden.

(40) Restriktion für Tondomänen
 Akzent-, Fokus- und phrasenfinale Silben und Fokusbereichsgrenzen sind maximale Begrenzungen für Tondomänen.
Empirische Generalisierung
 Die Dehnung nicht-tropischer Töne innerhalb einer Tondomäne erfolgt lediglich bis zur nächsten metrisch prominenten Silbe (mindestens zwei Schläge).

In der metrischen Struktur finden im Anschluß an Selkirk (1984) (vgl. auch Féry 1986 und Uhmman 1987MS) zwei Typen von Regeln Anwendung: die 'Text-Gitter-Regeln' bestimmen die Wortakzente und die 'Wohlklangsregeln' arbeiten ein "wohlgeformtes metrisches Gitter" aus.

Für die Wortakzentbestimmung ist die Unterscheidung zwischen "nativen" und "nicht-nativen" Wörtern (Wurzel 1980) zentral. Verkürzend dargestellt erhalten native Wörter den Wortakzent auf der ersten Stammsilbe und nicht-native auf der letzten schweren Silbe (CVC, CVV) über die Zuweisung eines **Schlags (x)** auf der 2.Ebene des metrischen Gitters. Die weiteren Ebenen werden durch Schlagzuweisung über das Prinzip IV (23) aufgebaut. Außerdem wachen Wohlklangsregeln darüber, daß das nach Selkirk (1984) universale Prinzip der "rhythmischen Alternation" gewahrt wird (so erhält die Silbe *te-* im Beispiel einen Schlag auf der 2. Ebene).

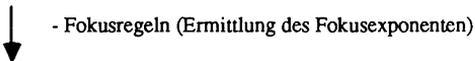
(48)''



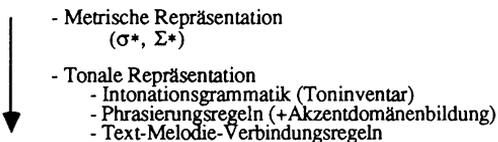
Durch diese Restriktion für die Tondomänenbildung wird die Anzahl tonal spezifizierter Silben weiter reduziert. Damit wird einem Anliegen Rechnung getragen, das im Einklang mit der geforderten phonologischen Relevanz eines Tonsystems für das Deutsche steht. Die Abbildung (49) zeigt abschließend den Gesamtaufbau des hier entwickelten Intonationssystems:

(49) Das Intonationssystem

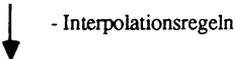
Oberflächenstruktur
 (+Argumentstruktur und Zuweisung von Fokusmerkmalen)



Fokusstruktur



Intonationsstruktur (Phonologische Repräsentation)



Phonetische Repräsentation

LITERATUR:

- BRAZIL, D. (1975), Discourse Intonation (Part I and II). English Language Research. Birmingham 1975.
- BURZIO, L. (1981), Intransitive Verbs and Italian Auxiliaries. MIT Diss. 1981.
- CHOMSKY, N. (1971), "Deep Structure, Surface Structure and Semantic Interpretation", in: D. Steinberg & L. Jakobovits (eds.), Semantics: An Interdisciplinary Reader in Philosophy, Cambridge 1971: 183-216.
- CONTRERAS, H. (1976), A Theory of Word Order with Special Reference to Spanish, Amsterdam 1976.
- COULTHARD, M. & D.BRAZIL (1982), "The Place of Intonation in the Description of Interaction", in: D. Tannen (ed.), Analysing Discourse: Text and Talk, Georgetown 1982: 94-112.
- COUPER-KUHLEN, E. (1983), "Intonatorische Kohäsion. Eine makroprosodische Untersuchung", Zeitschrift für Literaturwissenschaft und Linguistik 49, 1983:74-100.
- CRYSTRAL, D. (1969), Prosodic Systems and Intonation in English, Cambridge 1969.
- FÉRY, C. (1986), "Metrische Phonologie und Wortakzent im Deutschen", Studium Linguistik 20, 1986: 16-43.
- FUCHS, A. (1976), "Normaler und kontrastiver Akzent", Lingua 38, 1976:293-312.
- (1980) "Accented Subjects in "All -New" Sentences, in: Wege zur Universalienforschung (Festschrift für H.J. Seiler). Tübingen 1980: 449-461.
- GOLDSMITH, J. (1976), "An Overview of Autosegmental Phonology", Linguistic Analysis 2, 1976: 23-68.
- GOLDSMITH, J. (1981), "English as a Tone Language", in: D. Goyvaerts (ed.), Phonology in the 1980's. Gent 1981: 287-308.
- GUSSENHOVEN, C. (1983), "Focus, Mode and the Nucleus", Journal of Linguistics 19, 1983: 377-417.
- HAFTKA, B. (1980), "Reihenfolgebeziehungen im Satz (Topologie)", Kap. 4 in: Grundzüge einer deutschen Grammatik. Berlin Akademieverlag 1980: 702-764.
- HALLE, M. & J.R.VERGNAUD (1982), "On the Framework of Autosegmental Phonology" in: H.v.d. Hulst & N.Smith (eds.), The Structure of Phonological Representation. 2 Bde. Dordrecht 1982: 65-82.
- HALLIDAY, M.A.K. (1967), "Notes on Transitivity and Theme in English", Journal of Linguistics 3, 1967: 37-81, 177-274.
- HEIKE, G. (1962), "Suprasegmentale Merkmale der Stadtkölner Mundart. Ein Beitrag zur Rheinischen Schärfung", Phonetica 8, 1962: 147-165.
- HÖHLE, T.N. (1982), "Explicationen für 'normale Betonung' und 'normale Wortstellung' ", in: W.Abraham (ed.), Satzglieder im Deutschen. Tübingen 1982: 75-152.
- HULST, H.v.d. (1984), Syllable Structure and Stress in Dutch, Dordrecht 1984.
- ISACENKO, A.V. & H.J. SCHÄDLICH (1966), "Untersuchungen über die deutsche Satzintonation", Studia Grammatica VII, 1966: 7-64.
- JACKENDOFF, R.S. (1972), Semantic Interpretation in Generative Grammar, Cambridge 1972.
- JACOBS, J. (1982), "Neutraler und nicht-neutraler Satzakzent im Deutschen", in: Th.Vennemann (ed.), Silben, Segmente, Akzente. Tübingen 1982: 141-169.

- JACOBS, J. (1984), "Funktionale Satzperspektive und Illokutionssemantik", Linguistische Berichte 91, 1984: 25-58.
- KLEIN, W. & A. v. STECHOW (1982), Intonation und Bedeutung von Fokus, Arbeitspapiere des SFB 99 Nr. 77, Konstanz 1982.
- LADD, D.R. (1978), The Structure of Intonational Meaning, PhD Diss. University Microfilm International 1978.
- LEBEN, W.R. (1976), "The Tones in English Intonation", Linguistic Analysis 2, 1976: 69-107.
- LÖTSCHER, A. (1983), Satzakzent und funktionale Satzperspektive im Deutschen, Tübingen 1983.
- PHEBY, (1980), "Phonologie : Intonation", Kap. 6, in: Grundzüge einer deutschen Grammatik, Berlin Akademie-Verlag 1980: 839-897.
- PIERREHUMBERT, J. (1980), The Phonology and Phonetics of English Intonation, MIT Diss. 1980.
- SELKIRK, E. (1984), Phonology and Syntax: The Relation between Sound and Structure, Cambridge 1984.
- STECHOW, A.v. & S.UHMANN (1984), "On the Focus - Pitch Accent Relation", Groninger Arbeiten zur Germanistischen Linguistik 25, 1984: 223-263.
- & S.UHMANN (1986), "Some Remarks on Focus Projection", in: Abraham, W. & S.J. de Mey (eds.), Topic, Focus and Configurationality, Amsterdam 1986: 295-320.
- UHMANN, S. (1987 MS), Fokussierung und Intonation, Eine Untersuchung zum Deutschen anhand von Frage/Anwort - Sequenzen in experimentellen Dialogen. September 1987.
- WILLIAMS, E. (1976), "Underlying Tone in Margi and Igbo", Linguistic Inquiry 7, 1976: 463-484.
- WILLIAMS, E. (1981), "Argument Structure and Morphology", The Linguistic Review 1, 1980: 81-114.
- WURZEL, W. U. (1980), "Der deutsche Wortakzent: Fakten - Regeln - Prinzipien", Zeitschrift für Germanistik 3, 1980: 299- 318.

FOKUS-HINTERGRUND-GLIEDERUNG UND GRAMMATIK

Joachim Jacobs (München)

0. EINLEITUNG

Im folgenden soll skizziert werden, welche Rolle die Fokus-Hintergrund-Gliederung (FHG) auf verschiedenen Ebenen der Grammatik spielt. Dies soll einerseits ein Beitrag zur grammatischen Fundierung von Intonationsuntersuchungen sein. Andererseits werden auch Fragen von allgemeinem sprach- oder grammatiktheoretischen Interesse berührt, so die nach dem Einfluß traditionell als 'pragmatisch' eingestufte Zusammenhänge auf Sprachsysteme.

Ich stütze mich dabei auch auf Ergebnisse eigener Arbeiten (siehe Literaturverzeichnis). Was folgt, ist jedoch keine reine Zusammenfassung dieser Ergebnisse. Vielmehr werde ich früher vertretene Ansichten z.T. erheblich modifizieren und Probleme diskutieren, denen ich bisher zu wenig Aufmerksamkeit geschenkt habe.

Aus Platzgründen muß ich es mir versagen, andere in der Literatur vertretene Ansichten über die Rolle der FHG in der Grammatik ausführlich zu diskutieren. Dies bedaure ich insbesondere im Fall der Vorschläge von A. Lötscher (1983), E.O. Selkirk (1984) und M.S. Rochemont (1986), denen ich zahlreiche Anregungen verdanke. Alle Unterschiede zwischen diesen Konzeptionen und der nun darzustellenden sind beabsichtigt.

1. FOKUS-HINTERGRUND-GLIEDERUNG UND SEMANTIK

1.1 Das Phänomen

Die FHG ist, grob gesagt, eine Gliederung von Sätzen in hervorgehobene und nicht-hervorgehobene Teile. Dies exemplifizieren (1) - (7):

- (1) Gé[́]rda hat das neue Buch von Chomsky gelobt.
- (2) Gerda hat das né[́]ue Buch von Chomsky gelobt.
- (3) Gerda hat das neue Buch von Chomsky geló[́]bt.
- (4) Hat Gé[́]rda das neue Buch von Chomsky gelobt?
- (5) Lobe das né[́]ue Buch von Chomsky!
- (6) Wenn sie das neue Buch von Chomsky doch nur geló[́]bt hätte!
- (7) Gé[́]rda hat das neue Buch geló[́]bt (und Pé[́]ter hat es getá[́]delt).

In jedem dieser Sätze werden in der durch die Setzung des Akuts angedeuteten Akzentuierung bestimmte Teile als hervorgehoben, bestimmte andere als nicht-hervorgehoben wahrgenommen: In (1) ist Gerda hervorgehoben, der Restsatz nicht-hervorgehoben, in (2) ist

neue hervorgehoben, der Restsatz nicht-hervorgehoben, in (3) ist das Hauptverb hervorgehoben und wiederum der Restsatz nicht hervorgehoben. Entsprechendes gilt für (4), (5) bzw. (6), die deutlich machen, daß das hier zu diskutierende Phänomen nicht an einen bestimmten Satzmodus gebunden ist. Daß die Hervorhebung nicht immer nur ein einziges Element betrifft, zeigt (7), wo Gerda und gelobt gegenüber dem Restsatz hervorgehoben sind. Wir bezeichnen die hervorgehobenen Teile in Sätzen wie diesen als Fokus, die entsprechenden nicht-hervorgehobenen als Hintergrund. Wenn es, wie in (7), mehrere einzeln hervorgehobene Teile gibt, sprechen wir von einem mehrteiligen Fokus.

1.2 FHG in der semantischen Struktur

Für das Folgende wird es nützlich sein, eine syntaktische und eine semantische Ebene der FHG zu unterscheiden. Beginnen wir mit der Letzteren. Die mit der FHG intuitiv verbundene Funktion der Gegenüberstellung von hervorgehobenem und nicht-hervorgehobenem Material (s. 1.1) läßt sich auf der Ebene der semantischen Struktur durch eine Zweiteilung der semantischen Repräsentation des jeweiligen Satzes darstellen. Wenn wir z.B. annehmen, (1) habe vor Berücksichtigung der FHG als semantische Repräsentation die Formel (1'),

(1') $\text{PERF}(\text{LOBEN}(\text{GERDA}, \text{DAS NEUE BUCH VON CHOMSKY}))$

dann können wir die FHG von (1) durch folgende Umformung von (1') zum Ausdruck bringen:

(1'') $\# \lambda X(\text{PERF}(\text{LOBEN}(X, \text{DAS NEUE BUCH VON CH.})), \text{GERDA}\#$

(1'') wird hier also zerlegt in einen dem Hintergrund und einen dem Fokus entsprechenden Teil, wobei man den ersteren aus der ursprünglichen Formel dadurch erhält, daß man das Fokuselement durch eine Variable des entsprechenden Typs ersetzt und diese durch einen Lambda-Operator bindet. (1'') enthält dieselbe Information wie (1') und darüberhinaus die Information, welche Teile von (1) Fokus und welche Hintergrund sind. Allgemein ergeben sich Formeln wie (1'') nach der folgenden syntaktischen Regel:

(R1) Wenn α eine Formel ist, x_1, \dots, x_n Variablen der Typen a_1, \dots bzw. a_n sind und β_1, \dots, β_n Ausdrücke der Typen a_1, \dots bzw. a_n sind, dann ist $\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n\#$ eine Formel.

Danach ist (1'') eine Formel (mit $n = 1$), aber auch (7') (mit $n = 2$), das als Darstellung der FHG von (7) dienen kann:

(7') $\# \lambda R, X(\text{PERF}(R(X, \text{DAS NEUE BUCH VON CHOMSKY})), \text{LOBEN}, \text{GERDA}\#$

Die nach (R1) gebildeten Formeln müssen nun natürlich noch eine Deutung erhalten. Eine Deutung von Formeln semantischer Repräsentationssprachen besteht üblicherweise in einer Festlegung ihrer Wahrheitsbedingungen. Doch wie soll man die Wahrheitsbedingungen von (R1)-Formeln festlegen? Wenn man z.B. (1)

und (7) vergleicht, kann man keinen Unterschied in den intuitiven Wahrheitsbedingungen feststellen: Sie sind beide genau dann zutreffende Beschreibungen der Sachlage, wenn Gerda das neue Buch von Chomsky gelobt hat. Solche Beobachtungen legen es nahe, die Wahrheitsbedingungen von (R1)-Formeln als nicht 'sensitiv' für die FHG festzulegen, was etwa so aussehen könnte:

(R2) Der Wahrheitswert von $\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n \#$ bezüglich eines Wirklichkeitsausschnitts (einer 'möglichen Welt') w und eines Äußerungskontexts k ist gleich dem Wahrheitswert von $\lambda x_1, \dots, x_n(\alpha)(\beta_1, \dots, \beta_n)$ bezüglich w und k .

Die Formel $\lambda x_1, \dots, x_n(\alpha)(\beta_1, \dots, \beta_n)$ wiederum ist im allgemeinen wahr bezüglich w und k gdw. diejenige Formel wahr bezüglich w und k ist, die man aus α dadurch erhält, daß man für alle i ($1 \leq i \leq n$) jedes freie Vorkommnis von x_i durch ein Vorkommnis von β_i ersetzt. (Das Prinzip der Lambda-Konversion. Um die Ausnahmen zu diesem Prinzip brauchen wir uns hier nicht zu kümmern.)

Damit sind also (1') und (7') äquivalent: Sie haben beide dieselben Wahrheitsbedingungen wie (1').

Nun haben wir (R1)-Formeln zwar ordnungsgemäß eine Deutung zugeordnet, aber das Ergebnis ist noch recht mager. Wir machen mit (R2) zwar keine falschen Aussagen über die Wahrheitsbedingungen von Sätzen wie (1) oder (7), aber wir sagen auch nichts über die intuitiven inhaltlichen Unterschiede zwischen solchen Sätzen und damit über die inhaltlichen Effekte der FHG aus. Worin bestehen diese inhaltlichen Effekte? Sätze wie (1), (2) oder (7), die sich nur durch die FHG unterscheiden, sind mit verschiedenen typischen Äußerungskontexten verknüpft. Typische Äußerungskontexte für (1) sind solche, in denen mögliche alternative Besetzungen der Subjektstelle von (1) zur Debatte stehen, z.B., indem solche Besetzungen durch die vorangehende Äußerung eines Wer-Interrogativsatzes erfragt wurden, oder indem Sprecher und Adressat sich gerade darüber streiten, welche der möglichen Besetzungen dieser Stelle zutreffend sind, u.a.m. Dagegen stehen in typischen Äußerungssituationen für (2) mögliche alternative Besetzungen der Stelle des attributiven Adjektivs zur Debatte (im gerade angedeuteten Sinn), und in solchen für (7) mögliche alternative Paare aus einer Subjektbesetzung und einer Besetzung der Position des Hauptverbs. Diese Beobachtungen kann man generalisieren und - einer Idee von M. Rooth (1985) folgend - die inhaltliche Funktion der FHG ganz allgemein als Herstellung eines Bezugs zu Alternativen bestimmen. Jeder Satz mit FHG ist danach in gegebenen w und k (s.o.) nicht nur mit einem bestimmten Wahrheitswert, sondern auch mit einer bestimmten Menge von Alternativen verbunden. Welche dies ist, variiert mit der FHG. Diesen Zusammenhang von FHG und Alternativenbezug kann man durch folgende Regel festhalten:

(R3) Die Menge der mit $\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n \#$ bezüglich w und k verbundenen Alternativen ist w, k -alt($b(\lambda x_1, \dots, x_n(\alpha))$).

Dabei ist w, k -alt eine durch den Kontext k gegebene Funktion, die der Bedeutung eines n -stelligen Prädikats \mathcal{P} (notiert durch " $b(\mathcal{P})$ ") ein Element der Menge der möglichen Extensionen von \mathcal{P} in w zuordnet. (Die λ -Ausdrücke in den (R1)-Formeln sind ja n -stellige Prädikate. Die Bedeutung eines solchen Prädikats könnte man sich - mit Montague - als seine Intension vorstellen, also als eine Funktion von der

Menge der möglichen Welten in die Menge der möglichen Extensionen des Prädikats. Die Menge der möglichen Extensionen eines n -stelliger Prädikats ist die Menge aller Mengen von typmäßig passenden n -tupeln. Also sind die Werte von w, k -alt Mengen von n -tupeln.)

Der Alternativenbezug für $\langle 1' \rangle$ und $\langle 7' \rangle$ in einem konkreten Wirklichkeitsausschnitt w_1 und einem konkreten Äußerungskontext k_1 könnte danach z.B. folgendermaßen aussehen:

Die Menge der mit $\# \lambda X \langle \text{PERF} \langle \text{LOBEN} \langle X, \text{DAS NEUE BUCH VON CHOMSKY} \rangle \rangle \rangle, \text{GERDA} \#$ in w_1 und k_1 verbundenen Alternativen = w_1, k_1 -alt $\langle b \langle \lambda X \langle \text{PERF} \langle \text{LOBEN} \langle X, \text{DAS NEUE BUCH VON CHOMSKY} \rangle \rangle \rangle \rangle \rangle = \{ \text{Gerda, Paul, Heidrun, Frau Rumpel} \}$.

Die Menge der mit $\# \lambda R, X \langle \text{PERF} \langle R \langle X, \text{DAS NEUE BUCH VON CHOMSKY} \rangle \rangle \rangle, \text{LOBEN, GERDA} \#$ in w_1 und k_1 verbundenen Alternativen = w_1, k_1 -alt $\langle b \langle \lambda R, X \langle \text{PERF} \langle R \langle X, \text{DAS NEUE BUCH VON CHOMSKY} \rangle \rangle \rangle \rangle \rangle = \{ \langle \text{loben, Gerda} \rangle, \langle \text{tadeln, Gerda} \rangle, \langle \text{tadeln, Peter} \rangle, \langle \text{loben, Peter} \rangle \}$.

Daß die formale Rekonstruktion der simplen Idee, daß FHG einen Alternativenbezug herstellt, so relativ aufwendig ist, liegt nicht zuletzt daran, daß die Kontextabhängigkeit der Alternativen berücksichtigt werden muß. Sie manifestiert sich besonders deutlich in Beispielen wie dem folgendem:

(8) A: Wer kömmt, Háns oder Péter? []

B: Péter kömmt.

Hier engt A durch die Formulierung seiner Frage die Menge der für $\# \lambda X \langle \text{KOMMEN} \langle X \rangle \rangle, \text{PETER} \#$ in Frage Kommenden Alternativen auf eine Menge mit genau zwei Elementen ein, nämlich Hans und Peter. (Solche Zusammenhänge zwischen Frageformulierungen und Alternativenbezug werden in einem ähnlichen formalen Rahmen in v.Stechow 1986 untersucht.)

Für das Folgende wird es nützlich sein, eine objektsprachliche Notation für die mit $\langle R1 \rangle$ -Formeln verbundenen Alternativen zur Verfügung zu haben. Wenn f der Lambda-Ausdruck einer solchen Formel ist, dann sei $/f/$ ein Ausdruck der semantischen Repräsentationssprache, der w, k -alt $\langle b \langle f \rangle \rangle$ als Extension hat. Da der Wert von w, k -alt immer eine Menge von n -tupeln ist (s.o.) und da Ausdrücke, deren Extension Mengen von n -tupeln sind, n -stellige Prädikate sind, ist $/f/$ also ein n -stelliges Prädikat.

1.3 Die relationale FHG-Auffassung

Eine Bestimmung der Wirkung der FHG, die unsere bisherigen Überlegungen zusammenfaßt, könnte so lauten: Die FHG gliedert die semantische Struktur von Sätzen zum Zwecke der Herstellung eines Alternativenbezugs in hervorgehobene und nicht-hervorgehobene Teile. Diese Bestimmung trifft tatsächlich auf alle Fälle von FHG zu, aber sie gibt natürlich wenig Aufschluß über die genauen Verwendungsbedingungen von Sätzen mit FHG. Tatsächlich ergibt sich aus dem bisher Gesagten nur eine einzige Verwendungsrestriktion für Sätze mit FHG, nämlich die, daß die FHG so eingerichtet werden muß, daß die Menge der jeweiligen Alternativen mehr als nur die vom Fokus

bezeichnete Entität umfaßt. Wenn es keine Alternativen zum Fokus gibt, wäre ja der dem Alternativenbezug dienende 'Formulierungsaufwand' der FHG überflüssig. (Wir könnten diese Bedingung in (R3) einbauen, aber angemessener ist es wohl, sie als eine durch Prinzipien rationalen Kommunizierens ableitbare pragmatische Bedingung aufzufassen.) Damit sind insbesondere FHGn ausgeschlossen, für die es in keiner möglichen Sprechsituation (d.h. bei keinem möglichen w, k) mehr als höchstens eine Alternative gibt, unter anderem also FHGn wie die folgenden:

(9) ??Ich kenne den Mann, der dich beleidigt hat.

(10) ??Er interessiert sich für Autos.

(Man beachte aber, daß (9) - (10) akzeptabel sind, wenn sie als Zurückweisung einer vorangehenden falschen Formulierung verstanden werden. In diesem Fall gibt es im Kontext ja eine Alternative zum Fokus.)

Schon die flüchtigste Betrachtung der Daten zeigt jedoch, daß dies nicht die einzige FHG-abhängige Verwendungsrestriktion ist. Beginnen wir beim deutlichsten Fall: Es gibt Sätze, bei denen, anders als bei (1) - (7), die FHG in die Wahrheitsbedingungen, allgemeiner: in den propositionalen Gehalt, eingreift:

(11) Gerda kennt nur das neue Buch von Chomsky.

(12) Gerda kennt nur das neue Buch von Chomsky.

(11) und (12) unterscheiden sich formal nur durch die FHG-anzeigende Akzentuierung, und sie haben verschiedene Wahrheitsbedingungen: (11) kann dadurch falsifiziert werden, daß Gerda außer dem neuen Buch auch den neuen Aufsatz von Chomsky kennt. (12) kann dadurch nicht falsifiziert werden.

Der formale Unterschied zwischen (11) - (12) einerseits und (1) - (7) andererseits besteht in der Anwesenheit der Gradpartikel nur. Also liegt es nahe, den unterschiedlichen Effekt der FHG an dieser Partikel festzumachen, etwa indem man ihren Einfluß auf die Satzbedeutung als sensitiv für die FHG des Materials in ihrem Skopus expliziert:

(R4) $\text{NUR}(\#\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n\#)$ ist wahr bezüglich w, k gdw.
 $\#\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n\#$ wahr bezüglich w, k ist und
 $\forall \gamma_1, \dots, \gamma_n (\lambda x_1, \dots, x_n(\alpha) / (\gamma_1, \dots, \gamma_n) \rightarrow$
 $(\lambda x_1, \dots, x_n(\alpha) (\gamma_1, \dots, \gamma_n) \leftrightarrow (\gamma_1 = \beta_1 \ \& \ \dots \ \& \ \gamma_n = \beta_n)))$
 wahr bezüglich w, k ist.

Für (11) und (12) ergibt das (mit bereits vollzogener Auswertung von (R2) und Lambda-Konversion):

$\text{NUR}(\#\lambda P(\text{KENNEN}(\text{GERDA}, \text{DAS NEUE P VON CHOMSKY})), \text{BUCH}\#)$ ist wahr bez. w, k gdw.

$\text{KENNEN}(\text{GERDA}, \text{DAS NEUE BUCH VON CHOMSKY})$ wahr bez. w, k ist und

$\forall Q (\lambda P(\text{KENNEN}(\text{GERDA}, \text{DAS NEUE P VON CHOMSKY})) / (Q) \rightarrow$

$(\text{KENNEN}(\text{GERDA}, \text{DAS NEUE Q VON CHOMSKY}) \leftrightarrow Q = \text{BUCH}))$ wahr bez. w, k ist.

$NUR(\#AP(KENNEN(GERDA, DAS P BUCH VON CHOMSKY)), NEU\#)$ ist wahr bez. w, k gdw.

$KENNEN(GERDA, DAS NEUE BUCH VON CHOMSKY)$ wahr bez. w, k ist und

$\forall Q (\neg AP(KENNEN(GERDA, DAS P BUCH VON CHOMSKY)) / (Q) \rightarrow$

$(KENNEN(GERDA, DAS Q BUCH VON CHOMSKY) \leftrightarrow Q = NEU))$ wahr bez. w, k ist.

(11) ist danach also genau dann wahr, wenn Gerda das neue Buch von Chomsky kennt und sie außerdem keines von den anderen neuen Werken Chomskys kennt, die durch die in Frage kommenden alternativen Besetzungen der Stelle von Buch in (11) bezeichnet werden (etwa den neuen Aufsatz). (12) dagegen ist nach obiger Festlegung genau dann wahr, wenn Gerda das neue Buch von Chomsky kennt und sie außerdem keines von den anderen Büchern Chomskys kennt, die durch die in Frage kommenden alternativen Besetzungen von neu in (12) bezeichnet werden. Damit ist der Unterschied in den Wahrheitsbedingungen von (11) und (12) richtig charakterisiert und allgemein der Tatsache Rechnung getragen, daß in Sätzen mit nur die FHG Einfluß auf den propositionalen Gehalt hat.

Es gibt zahlreiche andere Elemente, die im selben Sinne wie nur sensitiv für FHG sind, d.h. in deren Bedeutungsregel essentieller Bezug auf die FHG des Materials in ihrem Skopus gemacht werden muß. So ist die in (13) exemplifizierte Art der Negation (replazive Negation, s. Jacobs 1987) FHG-sensitiv, indem sich ihre Bedeutung nach der Regel (R5) ergibt:

(13) Nicht Gerda ist mit Ottheinrich verlobt.

(R5) $NEG(\#\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n\#)$ ist wahr bezüglich w, k gdw.

$\sim \#\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n\#$ wahr bezüglich w, k ist und

$\exists \gamma_1, \dots, \gamma_n (\neg \lambda x_1, \dots, x_n(\alpha) / (\gamma_1, \dots, \gamma_n) \&$

$\lambda x_1, \dots, x_n(\alpha) (\gamma_1, \dots, \gamma_n))$ wahr bezüglich w, k ist.

Damit wäre (13) genau dann wahr, wenn es nicht der Fall ist, daß Gerda mit Ottheinrich verlobt ist, und es außerdem eine Person gibt, die als alternative Besetzungen der Subjektstelle in (13) in Frage kommt (z.B. Luise, Adelheid, Kurt) und mit Ottheinrich verlobt ist.

Weitere im selben Sinn FHG-sensitive Elemente sind die anderen Gradpartikeln (sogar, auch, selbst, nicht einmal, auch nur, ..), bei denen zusätzlich anzunehmen ist, daß die durch FHG in Zusammenwirken mit dem Kontext gesetzte Alternativenmenge in einer ebenfalls kontextabhängigen Weise geordnet ist, daß also manche Alternativen höher, manche niedriger eingestuft werden. Auf diese Alternativenordnung, die man als Skala auffassen kann, wird in einer partikelspezifischen Weise Bezug genommen. (Sogar fordert z.B., daß der Fokus hoch in der Skala rangiert, während der Fokus von auch nur einen niedrigen Rang einnehmen muß. Vgl. zu Einzelheiten Jacobs 1983, wo zudem gezeigt wird, daß ein Bezug auf bestimmte Skalenpositionen auch bei nur anzunehmen ist. Insofern ist (R4) eine Simplifizierung.)

Auch Satzadverbien (leider, versehentlich, zufälligerweise, ..) sind wohl FHG-sensitive Elemente. Es liegen für sie jedoch bisher keine plausiblen semantischen Analysen vor, die die FHG-Sensitivität (etwa so wie (R4) oder (R5)) transparent machen.

Schließlich sind Einstellungsverben (bedauern, bezweifeln, sich freuen, ..) FHG-sensitiv, vgl.:

(14) Luise bedauert, daß Gerda Ottheinrich geheiratet hat.

(15) Luise bedauert, daß Gerda Ottheinrich geheiratet hat.

Analysen von Einstellungsverben, die auf einer (R1)-ähnlichen Strukturierung der Formeln in ihrem Skopus beruhen (wobei allerdings unklar ist, ob diese Strukturierung als Repräsentation der FHG aufzufassen ist), finden sich in Cresswell/v.Stechow 1982, v.Stechow 1982 und Cresswell 1985.

Es gibt also zahlreiche FHG-sensitive Elemente. (Die eben gegebene Liste ist sicher nicht vollständig.) Dies legt die Frage nahe, ob die FHG nicht vielleicht immer auf solche Elemente bezogen ist. In Termini der semantischen Struktur Könnte man dies als Hypothese so formulieren:

(RFHG) (R1)-Formeln liegen in korrekten und vollständigen semantischen Repräsentationen immer im Skopus eines Operators, der für ihre interne Gliederung sensitiv ist.

Prima facie Gegenbeispiele zu (RFHG) sind semantische Strukturen wie (1'') und (7'), die Sätze repräsentieren, die keine Lexeme enthalten, die man als FHG-sensitive Operatoren deuten kann. Doch (1'') und (7') sind keine vollständigen semantischen Repräsentationen von (1) bzw. (7). Sie explizieren nur den propositionalen Gehalt dieser Sätze, sagen aber nichts über die dem Satzmodus entsprechenden Bedeutungsaspekte. Wenn wir letztere im Anschluß an Zaefferer 1984 mit dem Illokutionstyp identifizieren und diesen mit entsprechenden FHG-sensitiven Satzoperatoren (ASS, DIR, ERO, OPT .., für Aussage, Aufforderung, Frage, Wunsch ..) darstellen, erhalten wir für Sätze wie (1) Repräsentationen, die im Einklang mit (RFHG) stehen:

(1''') ASS(# λ X(PERF(LOBEN(X,DAS NEUE BUCH VON CH.))),GERDA#)

Eine Bedeutungsregel für ASS wäre:

- (R6) ASS(# λ x₁, ..., x_n(α), β ₁, ..., β _n#) ist wahr bezügl. w, k gdw. der Sprecher durch seine Äußerung in k zum Ausdruck bringt, daß er annimmt,
- daß # λ x₁, ..., x_n(α), β ₁, ..., β _n# wahr bezügl. w, k ist,
 - daß in k zur Debatte steht, auf welche Elemente von λ x₁, ..., x_n(α) das Prädikat λ x₁, ..., x_n(α) bezügl. w, k zutrifft,
 - daß es nicht zu den in k bereits etablierten Annahmen gehört, daß # λ x₁, ..., x_n(α), β ₁, ..., β _n# wahr bezügl. w, k ist.

Diese Regel müßte ergänzt werden durch eine geeignete Explikation der Begriffe 'in k zur Debatte stehen' und 'in k etablierte Annahme'. Sicher steht das Zutreffen eines Prädikats λ x₁, ..., x_n(α) dann zur Debatte, wenn im Kontext explizit die Frage gestellt wird, worauf dieses Prädikat zutrifft. Außerdem sollte man sagen können, daß das Zutreffen des Prädikats zur Debatte steht, wenn sein Zutreffen auf bestimmte Entitäten im Kontext behauptet wurde. Darüberhinaus muß man aber wohl noch zusätzliche, weniger direkt an Sprechakten festmachbare Kontextkonstellationen annehmen, die hinreichend für das Zur-Debatte-Stehen des jeweiligen Prädikats sind.

Die Rede von den 'in k etablierten Annahmen' in c) setzt voraus, daß Kontexte von Sprechhandlungen unter anderem dadurch gekennzeichnet sind, welche Annahmen allen Kommunikationsteilnehmern als unkontrovers und für den jeweils anderen 'leicht zugänglich' (etwa im Sinne eines Enthaltenseins im Kurzzeitgedächtnis) erscheinen. Möglicherweise läßt sich die Forderung, daß die assertierte Proposition nicht zu diesen Annahmen gehören soll, aus allgemeinen Prinzipien rationalen Kommunizierens ableiten. In diesem Fall könnte sie in (R6) wegfallen.

Eine entsprechende Analyse für (5) wäre:

(5') DIR(# λ P(LOBEN(DU, DAS P BUCH VON CHOMSKY)), NEU#)

- (R7) DIR(# $\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n$ ##) ist wahr bezügl. w, k gdw.
- der Sprecher durch seine Äußerung in k zum Ausdruck bringt,
- daß er möchte, daß der Adressat herbeiführt, daß # $\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n$ ## wahr bezügl. w, k ist,
 - daß er annimmt, daß in k zur Debatte steht, für welche Elemente aus $\lambda x_1, \dots, x_n(\alpha)$ der Adressat herbeiführen soll, daß das Prädikat $\lambda x_1, \dots, x_n(\alpha)$ auf sie bezügl. w, k zutrifft,
 - und daß er annimmt, daß es nicht zu den in k bereits etablierten Annahmen gehört, daß der Adressat herbeiführen wird, daß # $\lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n$ ## wahr bezügl. w, k ist.

Sätze ohne FHG-sensitive Lexeme sind also (unter Voraussetzung einer Klärung der in solchen Regeln verwendeten Begriffe) keine Gegenbeispiele zu (RFGH). Die mit dieser Hypothese verbundene Sicht kann man als **relationale FHG-Auffassung** (RFA) bezeichnen, weil nach ihr ja sowohl Fokus als auch Hintergrund in der semantischen Struktur immer auf ein weiteres Element, den FHG-sensitiven Operator, bezogen sind. Entsprechend kann man 'Fokus' und 'Hintergrund' selbst als relationale Begriffe verwenden: In (1) ist Gerda der Fokus der Assertion, der Restsatz der Hintergrund der Assertion, in (12) ist neu der Fokus von nur, der Restsatz Hintergrund von nur, usw. (In Jacobs 1984 wird die RFA in die Definition der FHG eingebaut. So wie diese Auffassung eben eingeführt wurde, gilt sie dagegen nicht per definitionem, sondern könnte empirisch widerlegt werden. Außerdem entgeht der in Jacobs 1984 vorgeschlagenen Theorie die Gemeinsamkeit zwischen allen Typen von FHG, nämlich der Alternativenbezug.)

1.4 Vergleich der relationalen mit der traditionellen FHG-Auffassung

Die RFA steht in deutlichem Gegensatz zur traditionellen FHG-Auffassung (TFA, wo für die einschlägigen Phänomene meist nicht "Hintergrund/Fokus" verwendet wurde, sondern "Thema/Rhema", "Topik/Fokus", "Topik/Kommentar", "Präsupposition/Fokus" etc.). Die TFA ist dadurch gekennzeichnet, daß sie der FHG einen bestimmten konstanten inhaltlichen Effekt zuordnet, nämlich dem Hintergrund 'alte', d.h. im Kontext vorerwähnte oder irgendwie anders gegebene Information und dem Fokus die entsprechende 'neue' Information (wobei diese Unterscheidung von 'alter' und 'neuer' Information in ganz verschiedenen Varianten auftaucht). Eine Relativierung dieses

inhaltlichen Effekts der FHG auf bestimmte grammatische Umgebungen ist in aller Regel nicht vorgesehen, und wenn, dann nur als marginales, kaum einer genaueren Analyse wertiges Phänomen (Ausnahme: Lieb 1984). Die RFA nimmt dagegen einen viel abstrakteren konstanten FHG-Effekt an, nämlich den Alternativenbezug (s.1.2), und sagt darüberhinaus, daß dieser inhaltliche Effekt grundsätzlich und in jedem Fall mit der grammatischen Umgebung der jeweiligen FHG bei der Konstituierung der konkreten Verwendungsbedingungen von Sätzen zusammenwirkt (s.1.3). Damit machen TFA und RFA deutlich verschiedene Voraussagen. So sagt die TFA im Gegensatz zur RFA, daß der Fokus immer 'neue', also im Kontext irgendwie noch nicht vorhandene Information beinhaltet. Diese Voraussage ist eindeutig falsch (wenn man den Begriff 'neu' nicht völlig vage verwenden will). In einem Satz wie

(16) Nicht Gerda hat Ottheinrich geheiratet.

beinhaltet der Fokus 'alte' Information: Der natürlichste Kontext für (16) ist eine vorangehende Äußerung des entsprechenden affirmativen Satzes, also eine, durch die Gerda vorerwähnt wird. Ebenso wenig ist es sinnvoll, zu sagen, in

(17) Oh, hätte doch Gerda Ottheinrich geheiratet!

beinhaltet Gerda 'neue' Information. Ein typischer Kontext für (17) ist ja einer, in dem die Antwort auf die Frage, ob Gerda Ottheinrich geheiratet hat, zu den im Kontext bereits etablierten Annahmen gehört.

Die RFA sagt zu (16) und (17), daß der hervorgehobene Satzteil Fokus der Negation bzw. des dem Optativ-Modus entsprechenden Illokutionstypoperators ist, und damit, daß in (16) und (17) andere FHG-Effekte vorliegen als in (18),

(18) Gerda hat Ottheinrich geheiratet.

wo das Subjekt Fokus des Assertionsoperators ist, dessen Zusammenwirken mit der FHG eine Wirkung hat, die eher in das Schema 'alte vs. neue Information' paßt, vgl. (R6). Eine inhaltliche Gemeinsamkeit zwischen der FHG-Wirkung in (16) - (18) sieht die RFA nur darin, daß mit Äußerungen all dieser Sätze ein Bezug zu Alternativen zum Subjekt hergestellt wird.

Eine weitere falsche Voraussage der TFA ist, daß die FHG sich immer auf derselben semantischen Ebene auswirkt, nämlich eben dort, wo die Unterscheidung von 'alter' und 'neuer' Information zu lokalisieren ist. (Welche Ebene das ist, wurde in der TFA kaum jemals genau bestimmt, Ausnahme wiederum Lieb 1984.) Daß dies falsch ist, sahen wir am Vergleich von (11) - (12), wo die FHG den propositionalen Gehalt beeinflusst, mit (1) - (7), wo dies nicht der Fall ist. Die RFA hat für diesen Unterschied eine einfache Erklärung: Die Ebene der FHG-Effekte koinzidiert mit der der FHG-sensitiven Elemente, durch deren Einfluß diese Effekte zustande kommen. In (11) ist die Ebene des FHG-sensitiven Elements die des propositionalen Gehalts, in (1) - (7) ist es die des Illokutionstyps.

Ein Nebenprodukt dieser unterschiedlichen Voraussagen über die (Nicht-)Gleichartigkeit der FHG-Effekte in verschiedenen Satztypen ist, daß die RFA weit weniger Probleme mit den leidigen Tests

für die FHG hat als die TFA. Die beiden bekanntesten in der TFA entwickelten FHG-Tests sind der Fragetest und der Widerspruchstest. Beim ersten geht man davon aus, daß das, was in einem Satz S, der als natürliche Antwort auf eine Ergänzungsfrage geäußert werden kann, der W-Phrase im Fragesatz entspricht, Fokus von S ist. Beim Widerspruchstest geht man davon aus, daß das, was in einem Satz S, der als natürlicher Widerspruch zur Behauptung eines anderen Satzes geäußert werden kann, den Unterschied zu letzterem ausmacht, Fokus von S ist. Mit diesen Tests kann man u.a. FHG-Ambiguitäten aufdecken. Daß man mit (19)

(19) Gerda hat Ottheinrich geheiratet.

sowohl (20) als auch (21) beantworten bzw. sowohl (22) als auch (23) widersprechen kann,

(20) Wen hat Gerda geheiratet? (21) Was hat Gerda gemacht?

(22) Gerda hat Paul geheiratet. (23) Gerda ist in Kloster gegangen.

zeigt, daß man in (19) sowohl Ottheinrich als auch Ottheinrich geheiratet als Fokus interpretieren kann (und den jeweiligen Restsatz als Hintergrund).

Eine unangenehme Eigenschaft dieser Tests ist nun, daß sich mit ihnen nur die FHG einfacher assertiver Sätze wie (19) ermitteln läßt. Sie funktionieren weder bei nicht-assertiven Sätzen noch bei assertiven Sätzen mit FHG-sensitiven Lexemen. Für dieses Problem wurde in der TFA keine überzeugende Lösung gefunden. Die RFA dagegen läßt von vorneherein gar nicht die Erwartung aufkommen, daß ein Typ von Tests auf alle Satztypen anwendbar ist. Wenn der allgemeinste FHG-Effekt des Alternativenbezugs je nach FHG-sensitivem Element ganz verschiedene konkrete Ausformungen erhält, müssen auch die Tests, die sich ja auf die konkreten und damit operationalisierbaren Ausformungen des FHG-Effekts zu stützen haben, ganz verschieden sein. Entsprechend kann man eine Zuordnung von FHG-sensitiven Operatoren und Testtypen vornehmen:

a) ASS: Frage- und Widerspruchstest;

b) DIR: Soll-Fragetest, z.B.:

A: Wenn soll ich heiraten? A: Was soll ich tun?

B: (Heirate) Ottheinrich! B: Heirate Ottheinrich!

c) ERO: Nachfragetest, z.B.:

A: Gerda hat geheiratet.

B: (Hat sie) Ottheinrich (geheiratet)?

A: Gerda hat etwas Dummes gemacht.

B: Hat sie Ottheinrich geheiratet?

d) NEG: Test der anschließbaren Sondern-Phrasen, z.B.:

Gerda hat nicht Otthéinrich geheiratet, sondern (sie hat) Kláus (geheiratet).

Gerda hat nicht Otthéinrich geheiratet, sondern ist ins Klóster gegangen.

e) NUR: Test des Alternativenausschlusses, z.B.:

A: Gerda will in Madrid alle möglichen Museen besuchen.

B: Nein, (sie will) nur den Prádo (besuchen).

A: Gerda will in Madrid alles mögliche tun.

B: Nein, sie will nur den Prádo besuchen.

f) AUCH: Test der Alternativenhinzufügung, z.B.:

A: Gerda will nur den Prado besuchen.

B: Nein, (sie will) auch die Armería (besuchen).

A: Gerda will nur einen Einkaufsbummel machen.

B: Nein, sie will auch die Armería besuchen.

(Die Beispiele zeigen, daß FHG-Ambiguitäten aufgelöst werden können, wenn man die möglichen Ellipsen berücksichtigt.)

Eine Voraussage der RFA, zu der es in TFA nichts Entsprechendes gibt, ist, daß es in Sätzen mehrere sich überlagernde FHGn geben kann, nämlich so viele, wie es FHG-sensitive Elemente in der semantischen Struktur gibt. Dies ist tatsächlich der Fall, vgl. (24):

(24) A: Warum hat Gerda nur den Prádo besucht?

B: Sie hat nur den Prádo besucht, weil sie verkátert war.

Im B-Satz ist (in einer Lesart) den Prado Fokus von nur und, wie der Kontext deutlich macht, der Kausalsatz Fokus der Assertion. Semantisch können wir das so repräsentieren:

(24B') ASS<#XP<P<NUR<#X<PERF<BESUCHEN<SIE,X>>>,DEN PRADO#>>>, WEIL<PRAT<VERKATERT<SIE>>>>#>

Die vorangehenden Bemerkungen zur TFA (mehr dazu in Jacobs 1984) sollten nicht so verstanden werden, daß wir an dieser auch heute noch weitverbreiteten Auffassung kein gutes Haar lassen wollen. Die TFA analysiert eine große Teilmenge der einschlägigen Fälle im Prinzip durchaus richtig, nämlich die FHG von Aussagesätzen ohne FHG-sensitive Oberflächenelemente. Tatsächlich unterscheidet sich (R6) nicht grundsätzlich von dem, was die TFA schon immer über die Wirkung der FHG in solchen Sätzen sagte. Der Fehler der TFA besteht aus unserer Sicht nur darin, sich bei dem Versuch einer allgemeinen Kennzeichnung der inhaltlichen Wirkung der FHG zu sehr auf diese einfachsten und häufigsten Fälle konzentriert zu haben.

Auch der bekannten Annahme der TFA, daß die Vorerwähntheit von Konstituenten – oder allgemeiner: ihre kontextuelle Präsenz – eine wichtige Rolle bei der Determination der auf die Informationsstruktur bezogenen Formaspekte von Sätzen spielt, wollen wir hier durchaus nicht widersprechen. So wird insbesondere die Akzentuierung durch diesen Faktor beeinflußt. Wie wir in 3.4 zu zeigen versuchen werden, muß man diesen Faktor jedoch, anders als es die TFA tut, bei der Analyse der Akzentuierung als einen von der (semantischen) FHG unabhängigen behandeln.

1.5 Einige Probleme

Ein Faktum, dem bisher weder in der TFA noch in der RFA ausreichend Rechnung getragen wurde, ist, daß Koordination mit FHG zusammenwirkt. Genauer: Koordinierte Satzteile haben im allgemeinen (s.u.) parallele FHGn, vgl.:

- (25) Sie hat Otthéinrich eingeladen, und (sie hat) Kláus (eingeladen).
- (26) Gélda hat Otthéinrich eingeladen, und Luíse (hat) Kláus (eingeladen).
- (27) Sie hat Otthéinrich éingeladen, und (sie hat) Kláus áusgeladen.
- (28) Sie hat sich entweder das grüne Kleid oder das róte (Kleid) gekauft.

Um dem Rechnung zu tragen, könnten wir versuchen, koordinierende Operatoren als FHG-sensitiv zu deuten, z.B.:

- (R8) $UND(\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n \#, \# \lambda x_1, \dots, x_n(\alpha), \gamma_1, \dots, \gamma_n \#)$ ist wahr bezügl. w, t gdw.
- a) sowohl $\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n \#$ als auch $\# \lambda x_1, \dots, x_n(\alpha), \gamma_1, \dots, \gamma_n \#$ wahr bezügl. w, t sind und
- b) sowohl $\# \lambda x_1, \dots, x_n(\alpha) / (\beta_1, \dots, \beta_n)$ als auch $\# \lambda x_1, \dots, x_n(\alpha) / (\gamma_1, \dots, \gamma_n)$ wahr bezügl. w, t sind.

Damit schließen wir Konstruktionen wie (29) – (31) aus:

- (29) ?Sie hat den Prádo besucht, und er hat den Prádo besucht.
- (30) ?Síe hat den Prado besucht, und sie hat die Armería besucht.
- (31) ?Síe ist géstern gekommen, und ér ist úngern gekommen.

(29) – (30) haben nicht die von der Syntax der UND-Formel in (R8) geforderte parallele FHG. Die beiden Teilsätze von (31) haben zwar eine parallele FHG, aber ungern und gestern sind in keinem denkbaren Kontext Alternativen. Also ist b) von (R8) verletzt. (Diese Bedingung erfaßt also vielleicht die Regel aus Lang 1984, daß es für koordinierte Elemente immer eine gemeinsame Einordnungsinstanz geben muß.)

Eines der Probleme von (R8) ist, daß Beispiele wie (31) intuitiv nicht notwendigerweise den Wahrheitswert 'falsch' haben. Also müßte

man (R8) so formulieren, daß die Verletzung von b) die Gesamtformel nicht falsch, sondern irgendwie kontextuell unangemessen macht.

Ein weiteres Problem ist, daß anscheinend nicht bei allen Formen der Koordination strikt parallele FHGs vorliegen. Beispiele wie die folgenden sind durchaus akzeptabel:

(32) Péter und núr Peter ist gekommen.

(33) Er war es, der die Scheune angezündet hat, und dann ist er auch noch in den Keller eingebrochen.

Während man (32) noch als Pseudo-Koordination abtun könnte (man beachte, daß das Finitum, anders als normalerweise bei koordinativen Subjekten, im Singular steht), weist (33) darauf hin, daß es neben einer fokussierenden Koordination (im Sinne von (R8)) auch eine nicht-fokussierende gibt (genauso wie fokussierende und nicht-fokussierende Negation, vgl. Jacobs 1982, 1987). Unter welchen Bedingungen die eine oder die andere Koordinationsart gewählt wird, ist mir gänzlich unklar.

Schließlich muß man es als unbefriedigend empfinden, daß Regeln wie (R8) nicht erklären, warum Koordination FHG-sensitiv ist. Dies ist ein Problem aller hier vorgeschlagenen Regeln für FHG-sensitive Elemente: Sie beschreiben die FHG-abhängigen Restriktionen für Verwendungen der entsprechenden Sätze, aber sie sagen nichts darüber aus, warum es solche Restriktionen überhaupt gibt. M.a.W.: Sie lassen offen, warum Sprachen überhaupt eine Informationsgliederung im Sinne der FHG vorsehen. Zur Beantwortung dieser Frage müßte man die Ebene der Sprachsysteme verlassen und sich den kognitiven Prozessen zuwenden, die die sprachliche Kommunikation begleiten. Die Vermutung liegt nahe, daß FHG ein Mittel zur Steigerung der kognitiven Relevanz von Äußerungen für den Adressaten ist. Recht deutlich ist dies im Fall des Negationsfokus. Bei der Verarbeitung der Äußerung eines Satzes mit nicht-fokussierender Negation wie (34)

(34) Es stimmt nicht, daß Gerda Ottheinrich gestern geheiratet hat.

muß der Adressat, wenn er die Äußerung als wahr akzeptiert und sie seine einzige Informationsquelle über den fraglichen Wirklichkeitsbereich ist, eine bestimmte Menge von Aussagen über diesen Wirklichkeitsbereich als unsicher in ihrem Wahrheitswert betrachten, d.h. als weder sicher wahr noch sicher falsch, nämlich bei (34) unter anderem die folgenden:

- (a) Jemand hat Ottheinrich gestern geheiratet.
- (b) Gerda hat jemand gestern geheiratet.
- (c) Jemand hat jemanden gestern geheiratet.
- (d) Gerda hat Ottheinrich irgendwann geheiratet.
- (e) Gerda hat irgendwann geheiratet.
- (d) Ottheinrich hat irgendwann geheiratet.

Bei einer entsprechenden Äußerung mit fokussierender Negation dagegen

- (35) Nicht Gerda hat Ottheinrich gestern geheiratet.

ist unter gleichen Voraussetzungen die Zahl der für den Adressaten unsicheren Aussagen deutlich geringer: Im gegebenen Fall sind es nur (b) und (e) (sowie (d), wenn Gerda und Ottheinrich früher schon mal verheiratet gewesen sein könnten).

Unter der plausiblen Annahme, daß von zwei Äußerungen mit etwa dem gleichen Formulierungsaufwand diejenige die kognitiv relevantere ist, die mehr Informationen als sicher präsentiert (vgl. Sperber/Wilson 1986), bewirkt also die Fokussierung im Umfeld der Negation eine Steigerung der kognitiven Relevanz. Es muß der weiteren Forschung überlassen bleiben, ob sich Entsprechendes auch für die FHG im Bereich anderer fokussierender Operatoren zeigen läßt.

2. FOKUS-HINTERGRUND-GLIEDERUNG UND SYNTAX

2.1 Semantische und syntaktische FHG

In 1. haben wir die Fokus/Hintergrund-Gliederung (FHG) als eine Gliederung der semantischen Struktur diskutiert. Für viele Zwecke ist es jedoch nützlich, die FHG auch als eine Gliederung der syntaktischen Struktur behandeln zu können. Dies kann durch eine geeignete Merkmalszuweisung an syntaktische Konstituenten geschehen. Im folgenden werden Regeln dafür formuliert, wie diese Merkmalszuweisung in Abhängigkeit von der semantischen FHG einzurichten ist. (Für mehrstellige FHG-sensitive Operatoren wie UND - s. 1.5 - müßten die Regeln entsprechend verallgemeinert werden.) Der Status dieser Regeln ist der einer Beschränkung für den Algorithmus, der in der Grammatik einer beliebigen Sprache syntaktische und semantische Strukturen miteinander in Beziehung setzt. Die Regeln sind jedoch nicht selbst Teil der Definition eines solchen Algorithmus.

- (R1) Wenn S' die semantische Repräsentation einer Lesart des Satzes S , K die S' entsprechende syntaktische Struktur von S und $P(\# \lambda x_1, \dots, x_n(\alpha), \beta_1, \dots, \beta_n \#)$ eine Sub-Formel von S' ist, wobei P die Bewertungszahl i (s. (R2)) hat, dann erhält für jedes j ($1 \leq j \leq n$) jedes Element der β_j entsprechenden K -Konstituentenmenge (s. (R3)) das Merkmal F_j .
- (R2) Die Bewertungszahlen von Vorkommnissen von FHG-sensitiven Operatoren P in semantischen Repräsentationen S' spiegeln ihren relativen SKopus wider: Der P mit dem weitestem SKopus erhält Bewertungszahl 1, der mit dem zweitweitesten SKopus Bewertungszahl 2, usw.
- (R3) Die einem Teilausdruck α der semantischen Repr. S' von S entsprechende Menge von K -Konstituenten ($K =$ die S' entsprechende syntaktische Struktur von S) ist $\{k_1, \dots, k_n\}$ gdw. für alle i ($1 \leq i \leq n$) gilt: es gibt einen Inhaltsbestandteil von k_i , der durch α (mit)repräsentiert wird und der nicht gleichzeitig ein Inhaltsbestandteil einer echten Teilkonstituente von k_i ist, und es gibt keine Konstituente k_j , die k_i als echte Teilkonstituente enthält und deren Gesamtbedeutung von α repräsentiert wird.

Dazu einige Beispiele: Nehmen wir an, der Satz (1)

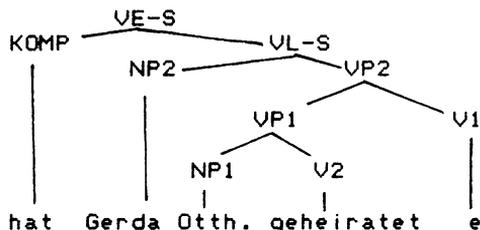
(1) Hat Gerda Ottheinrich geheiratet?

habe die semantische Struktur (1') (vgl. 1.3):

(1') $ERD(\# \lambda R(\text{PERF}(R(\text{GERDA}, \text{OTTHEINRICH}))), \text{HEIRATEN}\#)$

Nehmen wir weiter an, die (1') entsprechende syntaktische Struktur von (1) sei (k-1) (VE-S, VZ-S, VL-S = Satz mit Verberst-, Verbzweit- bzw. Verbletzstellung; die Verberst- und Verbzweitstellung wird einer verbreiteten Auffassung folgend als Ergebnis einer Bewegung des Verbs in die Position des Komplementierers beschrieben):

(k-1)



Dann erhält V2 in (k-1) das Merkmal F_1 , denn die Bewertungszahl von ERO ist 1 (s.(R2)) und die HEIRATEN in (1') entsprechende (k-1)-Konstituentenmenge ist $\{V2\}$. Es gibt ja einen Inhaltsbestandteil von V2, der durch HEIRATEN repräsentiert wird und der nicht gleichzeitig ein Inhaltsbestandteil einer Teilkonstituente von V2 ist, nämlich die Relation des Heiratens, und es gibt keine Konstituenten in (k-1), die V2 als echte Teilkonstituente enthalten und deren Gesamtbedeutung von HEIRATEN repräsentiert wird.

Ein weiteres Beispiel. Der Satz (2)

(2) Hat Gerda Ottheinrich geheiratet?

habe die semantische Struktur (2'):

(2') $ERO(\# \lambda P(\text{PERF}(P(\text{GERDA}))), \lambda X(\text{HEIRATEN}(X, \text{OTTHEINRICH}))\#)$

(M.a.W.: Hier ist der Inhalt der Verbalphrase Ottheinrich geheiratet im Fokus.)

Die syntaktische Struktur von (2) sei wiederum (k-1). In diesem Fall erhält VP1 das Merkmal F_1 : Bewertungszahl von ERO ist 1, und die $\lambda X(\text{HEIR.}(X, \text{OTT.}))$ entsprechende Konstituentenmenge ist $\{VP1\}$. Es gibt ja einen Inhaltsbestandteil von VP1, der durch $\lambda X(\text{HEIR.}(X, \text{OTT.}))$ repräsentiert wird und der nicht gleichzeitig Inhaltsbestandteil einer Teilkonstituente von VP1 ist, nämlich die Eigenschaft des Ottheinrich-Heiratens, und es gibt keine Konstituenten, die VP1 als echten Teil enthalten und deren Gesamtbedeutung von $\lambda X(\text{HEIR.}(X, \text{OTT.}))$ repräsentiert wird. (Insbesondere wird die Bedeutung von VP2 dadurch nicht vollständig repräsentiert.)

Der große Aufwand bei der Formulierung der Bedingungen der F_i -Zuordnung trägt Früchte bei Beispielen wie den folgenden, die von keiner bisherigen Theorie der syntaktischen FHG zufriedenstellend behandelt werden:

(3) Peter kömmt.

(3) hat in einer durch Kontexte wie (4) evozierbaren Lesart die semantische Repräsentation (3'):

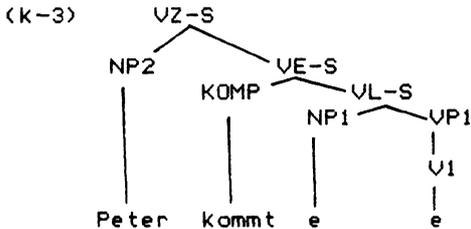
(4) A: Kommt Peter?

B: Peter kömmt.

(3') $ASS(\# \lambda P(P(\text{KOMMEN}(PETER))), \text{AFF}\#)$

AFF repräsentiert den hier von der Position des Finitums signalisierten uneingeschränkten Geltungsanspruch des Satzes (vgl. Höhle 1986). Alternativen zu AFF wären eingeschränktere Formen des Geltungsanspruchs.

Die syntaktische Struktur von (3) sei (k-3), wobei wir annehmen, daß auch die Vorfeldbesetzung Ergebnis eines transformationellen Prozesses ist:



Hier erhält nun KOMP das Merkmal F_1 : Bewertungszahl von ASS ist 1, und die AFF entsprechende Konstituentenmenge ist {KOMP}: Es gibt einen Inhaltsbestandteil von KOMP, der durch AFF repräsentiert wird und nicht Inhaltsbestandteil einer Teilkonstituente von KOMP ist, nämlich der uneingeschränkte Geltungsanspruch, und es gibt keine Konstituenten, die KOMP als echten Teil enthalten und deren Gesamtbedeutung von AFF repräsentiert wird.

Daß F_1 auch bei der semantischen Struktur (3'')

(3'') ASS(# λ P(P(PETER))),KOMMEN#)

an KOMP zugewiesen wird, entspricht einer Ambiguität von (3), vgl. (4) mit (5):

(5) A: Was tut Peter?

B: Peter kommt.

Im Beispiel (6) schließlich ist in der Lesart (6'), die in Kontexten wie (7) evoziert wird,

(6) Luise hat Gerda Ottheinrich ausgespannt.

(6') ASS(# λ P(PERF(P(GERDA))),
 λ X(AUSSPANNEN(LUISE,X,OTTH.))#)

(7) A: Was ist Gerda geschehen?

B: Luise hat Gerda Ottheinrich ausgespannt.

die dem semantischen Fokus entsprechende Konstituentenmenge für die F_1 -Zuordnung {NP2,VP1}, wobei NP2 = Luise und VP1 = Ottheinrich ausgespannt: Es gibt ja Inhaltsbestandteile von NP2 und VP1, die durch λ X(AUSSPANNEN...) (mit)repräsentiert werden und nicht gleichzeitig Inhaltsbestandteile einer Teilkonstituente von NP2 bzw. VP1 sind, nämlich Luise bzw. das Ottheinrich-Ausspannen, und es gibt keine Konstituenten, die NP2 bzw. VP1 als echten Teil enthalten und deren Gesamtbedeutung von

gebundenen syntaktischem Fokus, andernfalls von **freiem** Fokus. In den meisten Fällen koinzidieren freier und gebundener Fokus:

(10) Sie liebt nur₂ Kláus.

$$\begin{array}{c} F_2 \\ F_1 \end{array}$$

(Als Reaktion auf Sie liebt Klaus und Peter.)

1.2 FHG und Mittelfeldstellung

Eines der vielen syntaktischen Phänomene, die mit der FHG - syntaktisch repräsentiert durch die F_i -Zuordnung - interagieren, ist die Stellung von Satzgliedern im Mittelfeld deutscher Sätze, d.h. in dem Abschnitt zwischen dem Finitum (bei Verbletzstellung: der Konjunktion) und dem Beginn des Verbalkomplexes. Es ist oft behauptet worden, daß dafür das Prinzip gilt, daß der Fokus möglichst spät kommt. (So kann man schon das 'zweite Behaghelsche Gesetz' deuten, nach dem "das Wichtigere später steht als das Unwichtigere", vgl. Behaghel 1932, § 1426.) In unserer Terminologie könnten wir das so ausdrücken:

(R6) Wenn das Mittelfeld genau ein F_i -Satzglied K enthält, so folgt K auf alle anderen Satzglieder.

Als Manifestationen von (R6) kann man Beispiele wie die folgenden anführen:

(11a) Er hat das Buch $\overset{F_1}{\text{mir}}$ geschenkt.

(11b) ??Er hat das $\overset{F_1}{\text{Buch}}$ mir geschenkt.

(12a) Er hat mit Gerda am $\overset{F_1}{\text{Montag}}$ telefoniert.

(12b) ?Er hat mit $\overset{F_1}{\text{Gerda}}$ am Montag telefoniert.

(13a) Diesmal gewannen das Rennen die $\overset{F_1}{\text{Deutschen}}$.

(13b) ??Diesmal gewannen das $\overset{F_1}{\text{Rennen}}$ die Deutschen.

Die meisten Sprecher des Deutschen finden die a)-Sätze, die (R6) erfüllen, akzeptabler als die entsprechenden b)-Sätze, die (R6) verletzen.

Zu den Problemen von (R6) gehört die Frage des **Geltungsgrades** dieses Prinzips. Es gibt viele Beispiele, die (R6) verletzen und doch vollkommen akzeptabel sind:

(11c) Er hat $\overset{F_1}{\text{mir}}$ das Buch geschenkt.

(12c) Er hat am $\overset{F_1}{\text{Montag}}$ mit Gerda telefoniert.

(13c) Diesmal gewannen die $\overset{F_1}{\text{Deutschen}}$ das Rennen.

Man kann sich mit Uszkoreit 1984 solche Daten so erklären, daß die

Verletzung von (R6) durch die dadurch gesicherte Einhaltung anderer Wortstellungsprinzipien kompensiert wird, hier DatObj < AKKObj, TempAdv < PräpObj bzw. Subj < Obj (außerdem in (11c) noch Personalpronomen < Voll-NP). In den b)-Sätzen dagegen 'gewinnt' man durch die Verletzung von (R6) nichts, deswegen ist ihre Akzeptabilität gemindert. In den a)-Sätzen schließlich wird die Verletzung der eben genannten Prinzipien durch die dadurch gesicherte Einhaltung von (R6) kompensiert.

Bei dieser Sicht der Interaktion von Wortstellungsprinzipien wäre außerdem anzunehmen, daß die Prinzipien irgendwie gewichtet sind. Darauf weist die Tatsache hin, daß es Prinzipien gibt, deren Verletzung nicht durch die Einhaltung von (R6) kompensiert werden kann, z.B. AKKObj < GenObj:

(14) ??Sie beschuldigten des Verrats den General.

F₁

AKKObj < GenObj ist also anscheinend 'stärker' als (R6). Eine einigermaßen vollständige Theorie der Gewichtung von Wortstellungsprinzipien gibt es jedoch noch nicht. (Erste Ansätze dazu in Jacobs 1987a.)

Ein schöner Nebeneffekt dieser Sicht von der Interaktion zwischen (R6) und anderen Wortstellungsprinzipien ist, daß sie die Unmöglichkeit bestimmter Fokusplatzierungen voraussagt, z.B. derjenigen in (15), die in Kontexten wie (16) vorliegt:

(15) ??Er hat das Buch mir geschénkt.

F₁

(16) A: Was hat er mit dem Buch gemacht?

B: ??Er hat das Buch mir geschénkt.

Da in (15) die Einhaltung von (R6) und damit die Zuordnung von F₁ an das DatObj die einzige Rechtfertigung für die Spätstellung dieses Satzglieds ist (vgl. (11a)), senkt jede Zuordnung von F₁ an andere Konstituenten (z.B. an den Komplex aus DatObj und Hauptverb) die Akzeptabilität. Diese Erklärung wird bestätigt durch die Akzeptabilität von (17) (in Kontexten wie (18)):

(17) Er hat es mir geschénkt.

F₁

(18) A: Was hat er mit dem Buch gemacht?

B: Er hat es mir geschénkt.

Hier gibt es ja eine andere Rechtfertigung für die Spätstellung des DatObj, nämlich das Prinzip, daß bei Realisierung durch Personalpronomina daß AKKObj vor dem DatObj stehen soll. - Wir haben hier also das Muster einer Erklärung dafür, daß bestimmte 'unnormale' Wortstellungen 'enge' Fokussierung erfordern (vgl. z.B. v. Stechow/Uhmann 1986). In Reis 1987 wird allerdings dafür argumentiert, das Prinzip (R6) so umzuformulieren, daß es eine möglichst späte Position des fokusmarkierenden Hauptakzents im Mittelfeld fordert. (Zum Verhältnis von Fokussierung und Akzentuierung vgl. 3.) Auch das so umformulierte Prinzip scheint die fraglichen Akzeptabilitätsunterschiede voraussagen zu können.

Ein weiteres Problem von (R6) ist, ob es wirklich die möglichst späte Stellung des Fokus (oder des fokusmarkierenden Akzents)

ist, auf die es ankommt. Die Betrachtung der Verhältnisse in anderen Sprachen läßt daran Zweifel zu. Zwar ist ein Trend zur Spätstellung von fokussierten Konstituenten durchaus belegt (vgl. allgemein Hetzron 1975, speziell fürs Swahili Krifka 1983, fürs Englische Rochemont 1986), aber es gibt noch einen anderen belegten Stellungstrend, zu dem die von (R6) erfaßten Daten ebenfalls passen würden. In mehreren Sprachen scheint die (oder eine) präferierte Position für Fokus-Konstituenten eine unmittelbar verbadjazente zu sein. In OV-Sprachen ist das natürlich gleichzeitig eine 'späte' Position (vgl. für Japanisch und Türkisch Kuno 1980). In Nicht-OV-Sprachen kann diese verbadjazente Position dagegen durchaus 'früh' sein. Bekannt und viel diskutiert (vgl. Abraham/de Meij 1986) ist der Fall des Ungarischen, dessen normales Anordnungsmuster (nach Kiss 1980) $Top_1, \dots, Top_n F_i V K_1, \dots, K_m$ ist, wobei Top_i eine Topikkonstituente und K_i ein nicht-topikalischer, nicht-fokussierter Verbspezifikator ist. Eine Analyse, die im Einklang mit diesem Trend für das Deutsche nicht Spätstellung des Fokus, sondern Verbvoranstellung fordert, trüge den Mittelfelddaten Rechnung, wenn man von der Grundposition des Verbs am Ende des Mittelfeldes ausgeht, und könnte im Gegensatz zur Spätstellungstheorie erklären, warum die Fokussierung allein kein hinreichender Grund für eine Herausstellung ins Nachfeld ist:

(19) *Sie hat geheiratet Ottheinrich. *Schließliche B. +*
 F_1

(Vgl. die Analysen der Funktion von Herausstellungsmustern in Altman 1981.)

Schließlich findet sich in neueren Arbeiten von T. Givón (z.B. Givón 1985) die Behauptung, daß der statistisch am besten belegbare Stellungstrend für informationell wichtiges Material (also auch für den Fokus) universell der zur Erststellung ist. Die Verhältnisse im Deutschen bestätigen diese Behauptung sicher insofern, als man fokussierte Satzglieder beliebiger Kategorie ins Vorfeld (d.h. vor das Finitum bei Verbzweitstellung) setzen kann. Givóns Hypothese würde aber darüberhinaus voraussagen, daß dies eine häufigere Position für Fokuskonstituenten ist als die 'späten' Mittelfeldpositionen. Falls sich dies durch statistische Untersuchungen bestätigen läßt, müßte der Status von (R6) (in beiden oben angesprochenen Fassungen) neu überdacht werden.

2.3. Gebundener Fokus

(R6) ist nicht die einzige Restriktion für F_i -Plazierung im Deutschen. Wenn der Fokus **gebunden** ist (vgl. 2.1), unterliegt er gewissen Beschränkungen bezüglich seiner relativen Position zum 'Binder'. In Jacobs 1986 wurden u.a. die folgenden Regeln aufgestellt:

(R7) F_i muß im semantischen Bereich des Binders sein.

(R8) Der Binder muß F_i k-Kommandieren.

Ad (R7): Der semantische Bereich einer Konstituente K ist die syntaktische Entsprechung des SKopus der Entsprechung von K in der

semantischen Repräsentation. Vereinfacht gesagt ist im Deutschen eine Konstituente K_1 im semantischen Bereich einer Konstituente K_2 gdw. in der Grundstruktur a) kein 'Grenzknoten' (wie S oder NP) K_2 dominiert, der K_1 nicht dominiert, und b) K_2 vor K_1 steht. (Genauer in Jacobs 1982, 1986.) Damit erklärt (R7) die Inakzeptabilität von (21) - (22):

(21) *Daß nur₂ Peter gekommen ist, ärgert dich.

(22) ?weil Wackersdorf vielen CSÜ-Wählern sogar₂ ein Graus ist

(21) und (22) (wenn man dafür keine 'flache' Struktur annimmt) verletzen auch (R8). Dies weist darauf hin, daß (R8) vielleicht überflüssig ist. Auch Beispiele wie das folgende,

(23) *weil er von nur₂ einer reichen Frau träumt

die prima facie (R8) nötig erscheinen lassen, können wahrscheinlich allein mit (R7) erklärt werden. Zwar wird hier nach der Analyse von Jacobs 1983, 1986, reichen tatsächlich von nur nicht K-kommandiert (weil die GP hier als Ad-Artikel fungiert), aber es ließe sich wohl auch der Standpunkt vertreten, daß das Adjektiv hier nicht im semantischen Bereich der Partikel liegt.

Die Bedingungen für den semantischen Bereich (s.o.) sind auf die Grundstruktur (oder 'D-Struktur') bezogen, schließen also F_i -Verteilungen wie (24) - (25) nicht aus, wenn wir annehmen, daß Verbzweitstellung und Vorfeldfüllung das Resultat transformationeller Prozesse sind:

(24) Er telefonierte auch₂ mit ihr.

(25) Peter kam nur₂.

Problematisch ist jedoch, daß bei einer bestimmten Teilklasse der möglichen Fokusbinder des Deutschen der Fokus dem Binder auch im Mittelfeld vorangehen kann. Deutlichstes Beispiel ist auch:

(26) weil Peter manche der einfacheren Fragen auch₂ nicht beantworten konnte.

Die oben skizzierten Annahmen und die Annahme, daß es keine Transformation gibt, die Partikeln nach hinten bewegt (Jacobs 1983), legen eine Analyse nahe, nach der hier das AKKObj aus einer auf die Partikel folgenden Position herausbewegt wurde (vgl. Jacobs 1986). Es gibt aber kaum unabhängige Evidenz für eine solche Analyse, ja sogar einiges, was dagegen spricht: In einem Fall wie (24), wo die Voranstellung des Fokus vor die Partikel aus unabhängigen Gründen als Bewegung aus einer ursprünglich nachgestellten Position betrachtet werden kann, bleibt die Partikel wie bei Fokuspriorität unakzentuiert. In Fällen wie (26) muß sie dagegen betont werden. Insofern wäre eine Analyse dieser Fälle vorzuziehen, die nicht auf Transformationen beruht, sondern davon ausgeht, daß bei bestimmten zweigipfligen Akzentuierungsmustern die Präzedenzbedingung für den semantischen Bereich (s.o.) einfach aufgehoben ist. (Vgl. die Bemerkungen zur 'I-Topikalisation' in Jacobs 1982. In dieselbe Richtung geht der Vorschlag aus Bayer 1987,

in Sätzen wie (26) die Partikel als eine Art Anapher ihres Fokus zu betrachten.)

Eine weitere Restriktion für gebundene F_i ist (R9):

(R9) Der Binder kommt so spät im Satz, wie es der gegebene propositionale Gehalt des Satzes und die jeweils einschlägigen grammatischen Regeln (z.B. (R7)) erlauben.

(R9) schließt (27) und (28) aus, aber nicht (29) und (30):

(27) ??weil Gerda das Buch nur₂ Luise gelie^hhen hatte

(28) *weil sogar₂ sie Otthe^hinrich liebt^{F₂}

(29) weil er nur₂ mit einem der Inhaftierten spr^hächen wollte^F

(30) weil sie sogar₂ Otthe^hinrich liebt^{F₂}

(27) und (28) werden ausgeschlossen, weil uns weder der propositionale Gehalt noch irgendwelche grammatischen Regeln daran hindern, die Partikel weiter ans Satzende zu rücken:

(27a) weil Gerda das Buch Luise nur₂ gelie^hhen hat^{F₂}

(28a) weil sie sogar₂ Otthe^hinrich liebt (= (30))^{F₂}

(30) wird nicht ausgeschlossen, denn eine spätere Partikel-Position verstößt gegen (R7)/(R8):

(30a) ?weil sie Otthe^hinrich sogar₂ liebt^{F₂}

(29) wird nicht ausgeschlossen, weil (29a) einen anderen propositionalen Gehalt hat:

(29a) weil er mit einem der Inhaftierten nur₂ spr^hächen wollte^{F₂}

Die Geltung von (R9) macht einen wichtigen Unterschied zwischen Gradpartikeln und Modalpartikeln aus, die beliebig weit vor dem Fokus stehen können, vgl. Jacobs 1987b. Dies hängt mit der Tatsache zusammen, daß Modalpartikeln nicht im eigentlichen Sinne fokusbindend sind, sondern sich mit dem Fokus des Illokutionstypoperators assoziieren, vgl. ebd.

2.4 Syntaktische Unterscheidung semantischer Fokustypen

Manche syntaktischen Phänomene in natürlichen Sprachen beziehen sich nicht auf die FHG allgemein, sondern auf bestimmte semantisch definierbare Untertypen von FHG. So macht es syntaktisch sehr oft einen Unterschied aus, ob ein freier Fokus **exhaustiv** zu verstehen ist oder nicht. Exhaustiver freier Fokus liegt dann vor, wenn zum Ausdruck kommt, daß die vom freien Fokus denotierte Entität die einzige ist, die das Hintergrund-Prädikat erfüllt. Sehr oft kommt dies rein pragmatisch, nämlich durch die Vermittlung von Prinzipien des rationalen Kommunizierens, zum Ausdruck, so etwa in

(31):

(31) A: Mit wem gehst Du heute abend essen?

B: Ich gehe heute abend mit Frau Rumpel essen.

F₁

Falls B der Äußerung nichts hinzufügt, ist diese so zu verstehen, daß B mit Frau Rumpel und niemandem sonst essen geht.

In vielen Sprachen gibt es jedoch auch die Möglichkeit, eine exhaustive Fokusinterpretation konventionell zu signalisieren, nämlich durch den Einsatz bestimmter syntaktischer Konstruktionen. Dies scheint eine der möglichen Funktionen der verschiedenen Arten von Spalt-Sätzen ('cleft-sentences') in natürlichen Sprachen zu sein, deren gemeinsame syntaktische Eigenschaft es ist, daß sie zwischen dem Fokus und dem Restsatz eine prädikative Beziehung herstellen (vgl. Harries-Delisle 1978, Rochemont 1986.) Die semantische Struktur eines Spaltsatzes wie (32) kann man etwa wie in (32') repräsentieren ("!" bedeutet hier 'genau 1'):

(32) Es ist Luise, der ich das verraten habe.

F₁

(32') ASS(#X(31!Y(PERF(VERRATEN(ICH,Y,DAS)) & Y = X)),LUISE#)

(32') macht deutlich, daß hier die Exhaustivität des Fokus zum konventionellen propositionalen Gehalt gehört. Eine Repräsentation des letzteren erhält man durch Weglassen des Illokutionstypoperators. Das Resultat dieser Operation ist äquivalent mit (32''):

(32'') 31!Y(PERF(VERRATEN(ICH,Y,DAS)) & Y = LUISE)

Außerdem müßte der Inhalt der Formel, die man daraus durch Weglassen von Y = LUISE erhält, als präsupponiert gekennzeichnet werden, d.h. als einer Verneinung, In-Frage-Stellung etc. im Normalfall nicht zugänglich (vgl. Atlas/Levinson 1981. Hintergrundinformation ist aber nicht generell dasselbe wie präsupponierte Information, vgl. Reis 1977.)

Diese Interpretation der Bedeutung von Spalt-Sätzen erklärt unter anderem, warum bei ihnen der Fokus nicht durch eine Partikel aus der AUCH-Gruppe (auch, sogar ..) gebunden werden kann:

(33) *Es ist auch₂ Luise, der ich das verraten habe.F₂

Die Bedeutung von AUCH ist nämlich zu der der Präsupposition von (32') konträr: AUCH besagt, daß außer dem Fokus noch weitere Alternativen die Hintergrund-Prädikation erfüllen.

Kompliziert wird die Analyse von Spalt-Konstruktionen dadurch, daß das Element, das konstruktivell als der Einzigkeitsbedingung unterliegend gekennzeichnet wird, nicht immer identisch mit dem Fokus ist:

(34) Es ist die jüngste von Luisens Schwestern, der ich das verraten habe.

F₁

Nach dem oben angedeuteten Muster lassen sich jedoch auch solche Fälle semantisch repräsentieren:

(34') ASS(# λ P(\exists !X(PERF(VERRATEN(ICH,X,DAS)) &
X = DIE P SCHWESTER VON LUISE)),JÜNGSTE#)

Eine weitere inhaltliche Unterscheidung, die in vielen Sprachen Konsequenzen in der syntaktischen Kodierung der FHG hat, ist die von Kontrastivem und nicht-Kontrastivem Fokus. Kontrastiv ist ein Fokus dann, wenn er im jeweiligen sprachlichen Kontext explizit irgendwelchen Fokusalternativen gegenübergestellt wird. Grundsätzlich Kontrastiv ist z.B. der Negationsfokus, vgl. (35),

(35) Dieses Buch hat mir nicht₂ Kláus empfohlen, sondern Gërda.
F₂

Vom jeweiligen Kontext hängt es dagegen ab, ob der Fokus von Illokutionstypoperatoren Kontrastiv ist. Er ist es z.B. in (36), in (37) ist er es nicht:

(36) A: Ich fürchte, daß wieder die Schwéden gewonnen haben.

B: Keine Angst! Diesmal haben wir gewonnen.

(37) Weißt du, wer gewonnen hat? Wir (haben gewonnen)!
F₁

In der deutschen Syntax spielt diese Unterscheidung anscheinend nur eine marginale Rolle. Für manche Sprecher scheint (R6) (s.1.2) ein bißchen 'stärker' zu sein, wenn der Fokus Kontrastiv ist. Sie finden z.B. spätgestellte Subjekte in Kontexten wie (38) etwas akzeptabler als in solchen wie (39):

(38) Den Mülleimer hat dorthin nicht₂ der Kéllner gestellt, sondern die Putzfrau.
F₂

(39) A: Wer hat denn den Mülleimer dorthin gestellt?

B: Den Müllereimer hat dorthin der Kéllner gestellt.
F₁

Deutlichere syntaktische Relevanz hat die Unterscheidung von Kontrastivem und nicht-Kontrastivem Fokus in vielen anderen Sprachen. Im Efik (Niger-Kongo) z.B., das nicht-Kontrastivem Fokus durch Tonvariation auf dem Verbstamm signalisiert (verschiedene Töne, je nachdem der Fokus vor oder nach dem Verb steht), wird Kontrastiver Fokus zusätzlich durch Reduplikation des Verbstamms angezeigt (Beispiele aus de Jong 1980; der Akut bezeichnet hier einen hohen, der Gravis einen niedrigen Ton, der Apostroph einen 'downstep'):

(40) é-tim ǝ-kǝ-'bǝb à-kàm-bá ú-fòk.
Etim prät-bau-PräVFok groß Haus
'Etim baute ein großes Haus.' (nicht-Kontrastiv)

(41) é-tim ǝ-kǝ-'bǝ-bǝb ú-fòk ó-kò
Etim prät-bau-KontrPräVFok Haus das
'Etim baute das Haus.' (Kontrastiv)
F₁

3.FOKUS-HINTERGRUND-GLIEDERUNG UND AKZENTUIERUNG

3.1 Akzentuierung als Mittel der FHG-Markierung

Die bisher diskutierten Beispiele haben bereits deutlich gemacht, daß einer der mit der FHG zusammenwirkenden Formaspekte die Akzentuierung ist. Im folgenden sollen die Grundzüge dieses Zusammenwirkens diskutiert werden. Dabei sollte man aber im Auge behalten, daß die Akzentuierung weder universell noch auch nur im Deutschen das einzige Markierungsmittel für die FHG ist. Wenn man als Markierungsmittel für einen Bedeutungsaspekt B jeden Formaspekt bezeichnet, mit dem B im jeweiligen Sprachsystem in einer regelhaften Korrelation steht, so muß man nach dem in 2. Gesagten im Deutschen mindestens noch die **Satzgliedstellung** (vgl. (R6)), die **Anwesenheit und die Position von Fokusbindern** (vgl. (R7) - (R9)) sowie die Wahl bestimmter **Konstruktionstypen** (Spaltsätze etc.) als FHG-Markierungsmittel betrachten. (Im Schriftlichen kommt dazu noch die Möglichkeit, den Fokus durch Unterstreichung, Sperrung usw. anzuzeigen. Dafür fällt die intonatorisch signalisierte Akzentuierung weg.)

In anderen Sprachen müssen **morphologische Aspekte** als FHG-Markierungsmittel betrachtet werden (etwa in 2.4 erwähnten Efik), und schließlich haben manche Sprachen **fokusmarkierende Lexeme**, die anders als die Fokusbinder des Deutschen keine deutlich wahrnehmbare Eigenbedeutung haben. So verwendet das Swahili (nach Krifka 1983) Formen der Kopula ni- zur Fokusmarkierung:

(1) A: Nani alitoka Marekani? B: Ndi-ye Robert.
Wer kommt-von Amerika Welcher-ist R.

(Wie hier stehen fokusmarkierende Lexeme häufig in einer Beziehung zu 'cleft'-Konstruktionen.)

Das Verhältnis dieser Markierungsmittel untereinander ist ähnlich wie beim Satzmodus: Es können mehrere der jeweils zur Verfügung stehenden Mittel gleichzeitig verwendet werden, wenn sie in dieselbe Richtung weisen (d.h. hier: nicht verschiedene FHGn signalisieren), und es kommt in der Regel mindestens ein Mittel zum Einsatz, das zusammen mit eventuellen anderen Markierungsmitteln und dem Kontext den fraglichen Inhaltsaspekt hinreichend deutlich zum Ausdruck bringt. Dabei scheint allerdings zumindest bei der freien Fokussierung der Kontext stärker belastbar zu sein als beim Satzmodus. Wenn man einen Satz wie

(2) Der Kanzler hat den Wirtschaftsminister gerügt.

liest, der in seiner Form keine Hinweise auf die FHG enthält, wird man diese doch daran erkennen, welches Thema im Kontext gerade diskutiert wird. Wenn es um die Frage geht, worüber Kanzler und Wirtschaftsminister miteinander gesprochen haben, ist gerügt Fokus, wenn es darum geht, was der Kanzler gestern gemacht hat, ist den Wirtschaftsminister gerügt Fokus, usw. Diese große Belastbarkeit des Kontexts bei der Verdeutlichung der freien Fokussierung hängt natürlich damit zusammen, daß die FHG im Skopus von Illokutionstypoperatoren in der Regel auf im Kontext bereits zur Debatte stehende Fragestellungen verweist (vgl. 1.3).

Akzentuierung ist also nicht das einzige Markierungsmittel für FHG. Zudem ist die FHG auch nicht der einzige Bedeutungsaspekt, der mit der Akzentuierung korreliert. In Exklamativsätzen gibt es eine Akzentuierung, die offensichtlich in keiner Beziehung zur FHG steht:

(3) Bist dú aber schmutzig!

(4) Wie siehst dú denn aus?

Daß diese Akzente keine fokusmarkierenden sind, sieht man daran, daß hier kein Bezug zu Alternativen hergestellt wird, also zu anderen Personen, deren Schmutzigkeit bzw. Aussehen zu debattieren wäre. Solche Akzente kommen anscheinend nur in exklamativen Sätzen vor. Wir könnten hier deswegen von Exklamativakzenten sprechen (vgl. dazu auch den Beitrag über Exklamative von A. Batliner in diesem Band). Ein syntaktischer Unterschied zwischen Exklamativ- und fokusmarkierenden Akzenten scheint zu sein, daß erstere zu einer relativ frühen Stellung neigen, letztere dagegen nicht, s.u. Nicht alle Akzente in Exklamativsätzen sind jedoch Exklamativakzente:

(5) Hast du schon wieder das grüne Kleid angezogen (statt das rôte)!

Wie es scheint, blockiert die Anwesenheit eines fokusmarkierenden Akzents die eines Exklamativakzents (und umgekehrt).

Problematisch ist die funktionale Einordnung von Akzenten wie dem auf Peter in (6):

(6) Was nun Péter betrifft, so war ér es doch, der die Abmachung gebrochen hat.

Freie Thema-Phrasen, linksversetzte Phrasen und ähnliches (vgl. Altmann 1981) tragen grundsätzlich einen Akzent (auch in anderen Sprachen, vgl. Chafe 1976). Man kann hier von themamarkierenden Akzenten sprechen, wenn man die Funktion solcher Phrasen als Themasetzung bezeichnet. Dabei darf man aber den Terminus 'Thema' nicht im Sinne unseres Terminus 'Hintergrund' verstehen. Alle FHG-Tests zeigen, daß dies verschiedene Dinge sind. In (6) gibt es zudem durch die Spalt-Konstruktion des So-Satzes auch einen topologischen Hinweis auf die Verschiedenheit von Thema und Hintergrund. Darüberhinaus können Themasetzungsphrasen auch eindeutig fokusmarkierende Akzente enthalten:

(7) Was nun die jüngste von Gerdas Schwestern betrifft, so ...

(7) erfordert Kontexte, in denen auch über andere Schwestern Gerdas gesprochen wird.

Diesen Fakten kann man Rechnung tragen, wenn man annimmt, daß Themasetzungsphrasen in der semantischen Struktur im Skopus eines speziellen Illokutionstypoperators stehen, der den Vorgang der Themasetzung semantisch repräsentiert und FHG-sensitiv ist (vgl. Jacobs 1984). Danach wären also themamarkierende Akzente nichts anderes als fokusmarkierende eines speziellen Typs, nämlich solche, die den Fokus des Themasetzungsoperators markieren. (In (6) wäre Peter Fokus dieses Operators, in (7) das attributive Adjektiv.)

Damit bleibt aber immer noch fraglich, wie man Akzentuierungen wie die folgende erklären soll:

(8) A: Wen hat Kunigunde geheiratet?

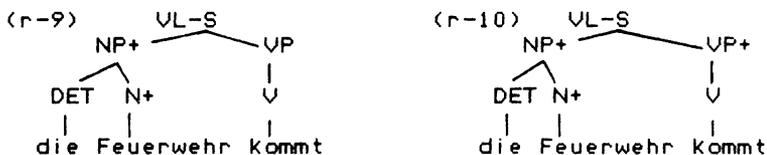
B: Kunigunde hat Ottheinrich geheiratet.

Das Subjekt ist hier eindeutig im Hintergrund und dennoch betont. Eine Analyse wie die für (7) scheidet aus, denn mit der Äußerung von Kunigunde im B-Satz vollzieht man keinen eigenen Sprechakt. Es bleiben zwei Möglichkeiten: a) Man betrachtet die Akzentuierung des Subjekts als satzphonologisches Phänomen. Dafür spricht, daß der Satz ohne den Subjektsakzent rhythmisch weniger ausgeglichen wäre. Dem Hauptakzent würde eine große Zahl unakzentuierter Silben vorausgehen, und dies wird – vor allem bei relativ langsamem Sprechen – generell vermieden, z.B. auch durch Ellipsen. b) Man nimmt eine weitere Akzentfunktion an, nämlich die der **Topikmarkierung**. Das Topik eines Satzes wäre dabei das, worüber mit dem Satz eine Aussage gemacht wird. (Der zum Topik komplementäre Teil wäre der Kommentar.) Für diese Analyse spricht unter anderem, daß man einen leichten Bedeutungsunterschied zwischen (8B) und der elliptischen Antwort Ottheinrich wahrnimmt, besonders, wenn auf den beiden Akzenten in (8B) ein steigend-fallendes Tonmuster realisiert wird (vgl. Jacobs 1982, Uhmann 1987).

Für welche der beiden Analysen man sich entscheidet, hängt davon ab, ob es gelingt, den Begriff 'Topik' einigermaßen handhabbar zu machen. In der bisherigen Forschung ist dies noch nicht gelungen. (Ein interessanter Vorschlag findet sich jedoch in Reinhart 1981.)

3.2 Akzentuelle Fokusmarkierung im Deutschen

Bevor man Akzentregeln diskutiert, muß man eine hinreichend präzise Notation für die Akzentuierung einführen. Im folgenden wird von zwei Notationen Gebrauch gemacht: Eine repräsentiert die relative **Hervorgehobenheit** von Konstituenten, die andere die **Betonungsstärke** von Silben. Zwei Beispiele dazu:



(n-9) die ¹Feuerwehr kommt (n-10) die ¹Feuerwehr ¹kommt

Die Interpretation von (n-9) und (n-10) ergibt sich nach dem Prinzip, daß in jeder Menge von Kokonstituenten a) die mit "+" gekennzeichneten Elemente gegenüber denen ohne "+" hervorgehoben sind und b) kein mit "+" gekennzeichnetes Element gegenüber einem anderen mit "+" gekennzeichneten Element hervorgehoben ist. (Als Kokonstituentenmengen ziehen wir hier und im folgenden nur solche Mengen von Schwesterkonstituenten in Betracht, deren Elemente zusammen wieder eine Konstituente bilden.)

Die Zahlen in (n-9) und (n-10) markieren mit zunehmender Größe abnehmende Betonungsstärkegrade. (Silben ohne Zahl gelten als unakzentuiert.)

Eine ausführliche Begründung für diese (an die metrische

Phonologie angelehnte) Unterscheidung zwischen zwei Akzentnotationen kann hier nicht gegeben werden. Man wird jedoch gleich sehen, daß sich damit ein recht flexibler Apparat zur Akzentbeschreibung ergibt.

Die Beziehung zwischen den zwei Akzentnotationen wird durch folgende Regeln festgelegt:

- (R1) a) In jeder Menge von KoKonstituenten haben die stärksten Silben jeder +-Konstituente eine stärkere numerische Betonung als alle anderen Silben.
 b) In jeder Menge von KoKonstituenten ist die numerische Betonung der stärksten Silben jeder +-Konstituente genauso stark wie die numerische Betonung der stärksten Silben jeder anderen +-Konstituente.
- (R2) Die stärksten Silben einer Konstituente K sind die wortakzenttragenden Silben derjenigen Wörter, die in K von der geringsten Anzahl von Knotenkette ohne +-Markierung dominiert werden.

Eine Knotenkette ist eine maximale, ununterbrochene Folge von einander dominierenden, aber nicht verzweigenden Knoten. So ist <VP,V> in (r-9) eine Knotenkette. In der Regel ist eine Knotenkette einfach ein einzelner Knoten, z.B. DET in (r-9). Eine Knotenkette ohne +-Markierung ist eine, bei der kein Element eine +-Markierung trägt.

(R2) sagt, daß die stärkste Silbe von NP in (r-9) Feu ist: Sie ist nach den Wortakzentregeln die akzenttragende Silbe von Feuerwehr, und dies ist das Wort, das in NP von der geringsten Anzahl von Knotenkette ohne +-Markierungen dominiert wird, nämlich von gar keiner. Damit erhält Feu die stärkste numerische Betonung in (r-9), denn dort muß nach (R1a) die stärkste Silbe von NP eine stärkere numerische Betonung bekommen als die stärkste Silbe von VP, nämlich kommt. (Das Wort kommt wird zwar von einer Knotenkette ohne +-Markierung dominiert, doch gibt es kein Wort in VP, daß von weniger derartigen Knotenkette dominiert wird.) In (r-10) muß dagegen die stärkste Silbe von NP nach (R1b) eine gleich starke Betonung bekommen wie die stärkste Silbe von VP. (n-9) bzw. (n-10) erfüllen diese Bedingungen, aber auch z.B. (n-9'):

(n-9') die ¹Feuerwehr ²kommt

Die durch (R1) - (R2) festgelegte Relation zwischen relativer und numerischer Akzentstärke ist also keine Funktion. Dies entspricht der Tatsache, daß die 'Bewegtheit' mit der eine gegebene Verteilung relativer Hervorgehobenheit realisiert wird, mit Faktoren wie dem Sprechtempo oder dem Stil variieren kann. Der Zusammenhang wird weiter gelockert durch die Tatsache, daß die letzte von mehreren 1-Betonungen, wenn sie sich am Ende einer vollständigen Intonationsphrase befindet (also in der Regel am Ende eines Satzes), oft als etwas prägnanter wahrgenommen wird als die vorangehenden 1-Betonungen (nämlich dann, wenn der Sprecher nicht durch den Einsatz bestimmter intonatorischer Mittel 'gegensteuert'). Das liegt wohl daran, daß das letzte von mehreren ungefähr gleichstarken Signalen generell das perzeptiv auffallendste ist, aber auch daran, daß bei der letzten Hauptbetonung intonatorisch 'am meisten los ist' (vgl. die anderen Beiträge in diesem Band). Wenn wir dies in unserer

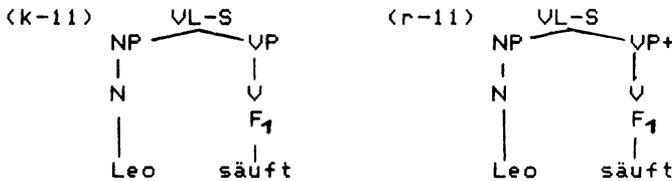
numerischen Notation berücksichtigen, können wir $\langle n-10' \rangle$ als eine Variante von $\langle n-10 \rangle$ betrachten:

$\langle n-10' \rangle$ die Feuerweh²r kömmt¹

Die zentrale Regel für die fokusmarkierende Akzentuierung ist nun die folgende:

(R3) Wenn in einer Menge von Kokonstituenten K die Elemente k_1, \dots, k_n mit F_i ($1 \leq i$) markiert sind oder eine mit F_i markierte Konstituente dominieren, wenn außerdem kein Element von K mit F_j ($j < i$) markiert ist oder eine damit markierte Konstituente dominiert und wenn zudem die Elemente von K nicht gemeinsam von einem Knoten mit einem solchen Merkmal F_j dominiert werden, dann werden k_1, \dots, k_n und keine anderen Elemente von K mit "+" markiert.

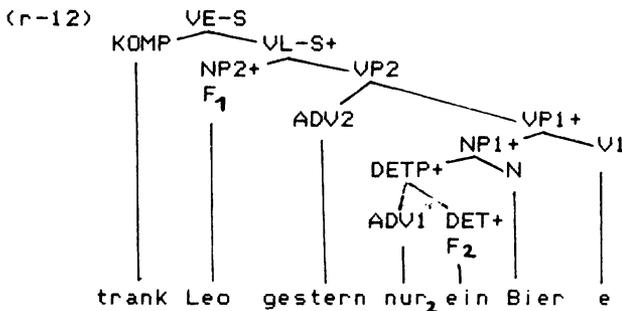
(R3) ist obligatorisch, d.h. muß immer angewandt werden, wenn sie angewandt werden kann. Einige Beispiele dazu. Der Struktur $\langle k-11 \rangle$ entspricht nach (R3) $\langle r-11 \rangle$:



$\langle r-11 \rangle$ wiederum entspricht nach (R1) und (R2) z.B. $\langle n-11 \rangle$:

$\langle n-11 \rangle$ Leo² säuft¹.

Analog erhält man nach (R3) die Struktur $\langle r-12 \rangle$ aus der entsprechenden Struktur ohne +-Markierung (Kontext: Nachfrage auf Jemand trank gestern nur ein Bier.):



In der Kokonstituentenmenge $\{KOMP, UL-S\}$ dominiert VL-S eine mit F_1 markierte Konstituente, kein Element ist mit F_j ($j < 1$) markiert, dominiert eine damit markierte Konstituente oder wird von ihr dominiert. Also erhält VL-S ein "+", KOMP dagegen nicht. In der Kokonstituentenmenge $\{NP2, VP2\}$ erhält aus analogen Gründen NP2, aber nicht VP2, ein "+". In der Kokonstituentenmenge $\{ADV2, VP1\}$ dominiert

VP1 eine mit F_2 markierte Konstituente, kein Element ist mit F_j ($j < 2$) markiert, dominiert eine so markierte Konstituente oder wird von ihr dominiert. Also erhält VP1, aber nicht ADV2, ein "+". Analoges gilt für die +-Markierung von NP1 in {NP1, V1}, von DETP in {DETP, N} und von DET in {ADV1, DET}.

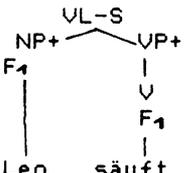
Nach (R1) - (R2) entspricht (n-12) z.B. (n-12):

(n-12) Trank ¹Leo gestern nur ²ein Bier?

Le erhält also die numerische Hauptbetonung des Satzes, obwohl Leo nur von zwei "+" dominiert wird. Nach (R1a) muß ja die stärkste Silbe von VL-S einen stärkere Betonung erhalten als jede andere Silbe in {KOMP, VL-S}. Damit muß Le numerisch stärker als alle anderen Silben im Satz sein, denn das Wort Leo wird in VL-S von der geringsten Zahl von Knotenkette ohne +-Markierung dominiert, nämlich von gar keiner. (Man beachte, daß alle anderen Wörter in VL-S von mindestens einer Knotenkette ohne +-Markierung dominiert werden, nämlich von VP2.)

In {ADV2, VP1} lenken die +-Markierungen die stärkste Betonung auf ein. Sie muß aber nach dem eben Gesagten schwächer sein als die von Le.

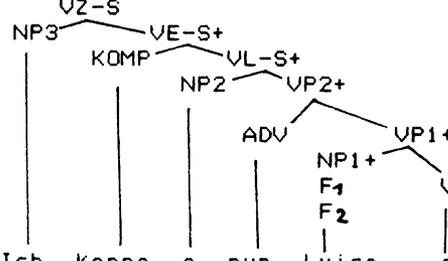
Wenn, anders als in (12), mehrere Fokusmerkmale mit dem gleichen Index anwesend sind, wird keine Akzentabstufung, sondern eine Reihung gleich starker Akzente induziert (Kontext des Beispiels: weil Leo säuft, Luise raucht und Peter drogensüchtig ist):

(n-13)  (n-13) ¹Leo ¹säuft

Neben (n-13) kann nach dem oben zu (n-10') Gesagten auch (n-13') als mögliche Realisierung von (n-13) gelten:

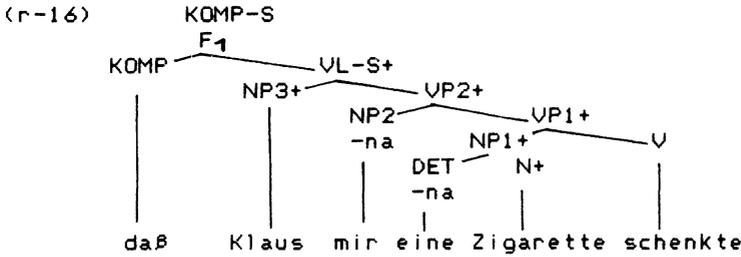
(n-13') ²Leo ¹säuft

In bestimmten Fällen kann es bei Anwesenheit mehrerer syntaktischer Foki mit verschiedenem Index nur einen einzigen fokusmarkierenden Akzent geben (Kontext des Beispiels: Antwort auf Welche von Gerdas Freundinnen kennst du?):

(n-14)  (n-14) Ich kenne e nur₂ Luise e

(R5) Wenn {k,X} eine fokusfreie, integrierbare (s.u.) Kokonstituentenmenge ist, X = V, A, ADV, P, DET oder KOMP, und wenn k nicht -na ist, dann erhält k (und nur k) die Markierung "+".

Nehmen wir an, in (k-16) seien {DET,N} , {NP1,V} und {VL-S,KOMP} integrierbar. Dann erhalten wir für (k-16) mit (R4) und (R5) (r-16) bzw. (n-16) oder (n-16')



(n-16) daß Klaus¹ mir eine Zigarette¹ schenkte

(n-16') daß Klaus² mir eine Zigarette¹ schenkte

Das Problem dabei ist natürlich der in (R5) vorausgesetzte Begriff der Integrierbarkeit. ('Integrierbarkeit', weil die in (R5) beschriebene AKzentuierung häufig den Eindruck erweckt, als seien die jeweiligen Kokonstituenten zu einer semantischen Einheit integriert, vgl. Fuchs 1976.) Als erste Eigenschaft der Integrierbarkeit ist festzuhalten, daß jedes {k,X} integrierbar ist, bei dem X = P, DET oder KOMP und k die Ergänzung von X ist (also die die Adposition ergänzende NP, der den Determinator ergänzende nominale Komplex bzw. der von KOMP eingeleitete VL-S). Dies könnte man zu der Hypothese verallgemeinern, daß Integrierbarkeit nur dann vorliegt ist, wenn k ein 'internes' oder 'direktes' Argument von X ist (zu diesen Begriffen vgl. z.B. Czepluch 1987). Dazu paßt, daß bei X = V keine Integrierbarkeit besteht, wenn k das Agens von V beinhaltet (vgl. aber 3.4): agentivische Verbargumente gelten als 'extern' bzw. 'nicht-direkt' (vgl. v.Stechow/Uhmann 1984). Man vergleiche dazu (17a,b), wo Integrierbarkeit gegeben ist (das Subjekt ist ein nicht-agentivisches Argument des Verbs), mit (18a,b), wo das Subjekt Agens ist. In den b-Sätzen ist das Verb akzentgesenkt, in den a-Sätzen nicht. Entsprechend ist in (17) der b-Satz, in (18) der a-Satz neutral akzentuiert (es sei denn, man versteht in (18b) das Subjekt nicht als Agens, sondern als den Gegenstand, der zum Kochen gebracht wird, vgl. Die Suppe kocht.):

- | | |
|---|-----------------|
| (17a) Peter ¹ kommt ¹ . | (nicht-neutral) |
| (17b) Peter ¹ kommt ² . | (neutral) |
| (18a) Peter ¹ kocht ¹ . | (neutral) |
| (18b) Peter ¹ kocht ² . | (nicht-neutral) |

Natürlich reicht die Bedingung der Nicht-Agentivität nicht aus. Z.B. leistet sie nichts für die Unterscheidung zwischen integrierbaren und nicht-integrierbaren Kombinationen aus Verben und

Objekten, vgl. (K-16) vs. (K-15). Man könnte natürlich sagen, daß es sich in (K-15) im Gegensatz zu (K-16) um ein 'externes' Argument handelt. Ob es dafür unabhängige Evidenz gibt, ist jedoch fraglich. Daß es nicht-integrierbare Komplexe aus direktem Objekt und Verb gibt, ist im übrigen in der Literatur (z.B. Allerton/Crutenden 1979, Fuchs 1980, v.Stechow/Uhmann 1984, Krifka 1984) weitgehend ignoriert worden. Man hat sich hauptsächlich auf die (Nicht-)Integrierbarkeit von Subjekt-Verb-Komplexen konzentriert und deshalb auch übersehen, daß es als Erklärung für die gelegentliche Integrierbarkeit eines solchen Komplexes nicht ausreicht, das Subjekt als 'zugrundeliegendes' direktes Objekt zu analysieren (wofür es im übrigen durchaus gute Argumente gibt, vgl. Grewendorf 1986). Das wäre ja nur dann eine Erklärung, wenn alle direkten Objekte mit dem Verb integrierbar wären.

Ein weiteres Problem von (R5) ist, daß diese Regel die strukturellen Konfigurationen, in denen es zu integrierender Akzentuierung kommt, möglicherweise zu eng bestimmt. Schon (17b) wird eigentlich durch (R5) nicht mehr erfaßt, da nach der hier vorausgesetzten Strukturanalyse von Verbzweitstellungssätzen (vgl. 2.) die involvierte KoKonstituentenmenge {NP,VE-S} ist. Wir müssen also wohl für X in (R5) auch Kategorien zulassen, die die Kategorie ψ (bzw. P, A etc.) dominieren und die Integrierbarkeit unter bestimmten Bedingungen von ihr 'erben' können. Das zeigen auch Fälle, auf die S. Uhmann (1987) aufmerksam gemacht hat:

(19) Er hat ein Loch in eine Wand geschlagen. (neutral)

In Uhmann 1987 findet sich darüberhinaus eine interessante Hypothese zur Klärung der Frage, ob und wann die Wahl einer integrierenden Akzentuierung obligatorisch ist. Man vergleiche dazu die folgenden Beispiele:

(20a) wegen der Regierungskrise (neutral)

(20b) wegen der Regierungskrise (neutral)

(21a) der Regierungskrise wegen (neutral)

(21b) der Regierungskrise wegen (nicht-neutral)

Neben (20a) ist durchaus auch (20b) mit nicht-integrierender (oder 'isolierender') Akzentuierung im Neutralfall möglich (wenn auch nur bei etwas exaltiertem Sprechstil). (21b) ist dagegen als Neutralakzentuierung nicht möglich, sondern nur, wenn sowohl NP als auch P ein eigenes Fokusmerkmal tragen. Uhmann erklärt diesen Unterschied damit, daß in einer KoKonstituentenmenge, die integrierbar ist (wie in {NP,P}), in keinem Fall das letzte +-Merkmal auf der zu integrierenden Konstituente (X von (R5)) liegen darf. (Bei Uhmann: Der letzte Hauptakzent muß auf dem 'Fokusexponenten' liegen.) Diese Bedingung erlaubt es, einerseits die Anwendung von (R5) als grundsätzlich optional zu betrachten und damit Variationen wie (20a,b) Rechnung zu tragen, andererseits dennoch vorauszusagen, daß die Nicht-Anwendung von (R5) in bestimmten Fällen (wie (21b)) inakzeptabel ist. (Die Optionalität von (R5) haben wir schon in (r-15) ausgenutzt, wo wir für {N,DET} (R4) statt (R5) anwendeten.)

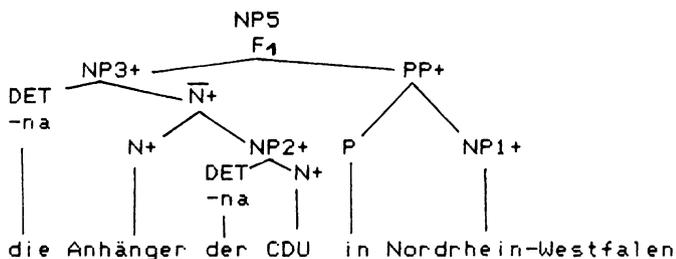
Schließlich läßt uns die in (R5) vorgesehene Möglichkeit, daß X = A oder ADV, erwarten, daß es auch Akzentintegration in umgekehrter Richtung gibt, daß also bei Neutralakzentuierung unter bestimmten Bedingungen auch der Kopf gegenüber seinem Begleiter hervorgehoben werden kann. Ein Beleg für diese Möglichkeit ist die folgende Neutralakzentuierung (übrigens durchaus im Einklang mit unserer Bedingung, daß Integration eine Argument-Beziehung zwischen K und X voraussetzt, s.o.):

(22) ²Gerda hat schlecht ¹geschlafen.

Komplexe aus Verb und Adverbial sind jedoch keineswegs immer integrierbar. So kann (R5) nicht auf Komplexe aus Verben und Temporal- oder Lokaladverbialen angewandt werden.

Leider müssen wir die integrierende Akzentuierung hier verlassen, um nicht zu weit vom eigentlichen Thema, der Fokusakzentuierung, abzukommen. Hinzugefügt sei nur noch, daß es in NP-wertigen Kokonstituentenmengen anscheinend nur im Bereich vor dem Kernsubstantiv zu Akzentintegration kommt, vielleicht deswegen, weil es danach keine Argument-Beziehungen im eigentlichen Sinn (wie zwischen Verb und Objekt) gibt. Entsprechend haben wir unter Neutralbetonungsbedingungen vom Kernsubstantiv bis zum Ende der NP, wie es durch (R4) vorausgesagt wird, eine gleichmäßige Betonung aller unter diesen Bedingungen akzentuierbaren Konstituenten, mit der nun schon bekannten Privilegierung des letzten Akzents:

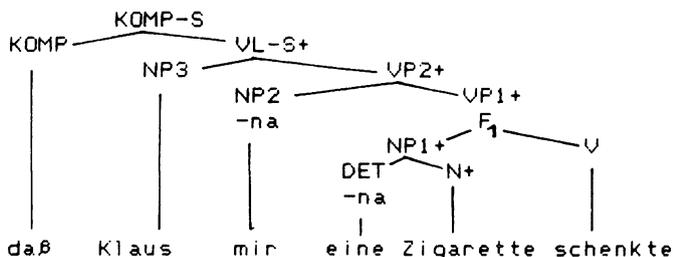
(r-23)



(a-23) die ²Anhänger der CDU ²in Nordrhein-Westfalen ¹

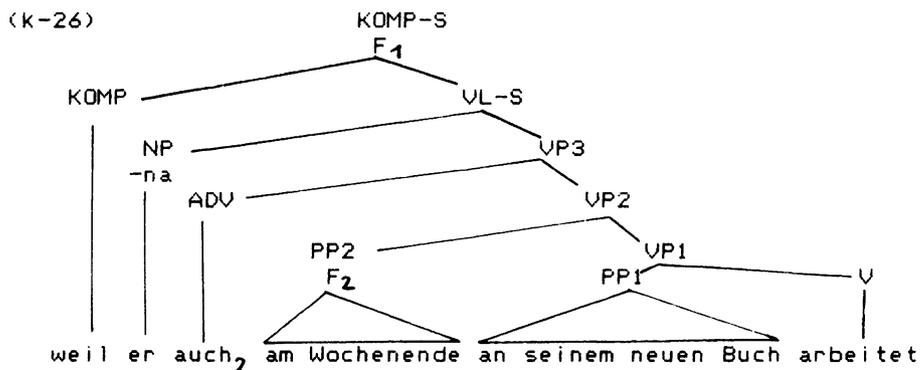
Zurück zum fokusmarkierenden Akzent: Oft liegen die F_i-Merkmale so, daß man für die Ableitung der Akzentuierung sowohl Neutral- als auch Fokusakzentregeln anwenden muß. Ein Beispiel ist (r-24):

(r-24)



hier dahingestellt bleiben.

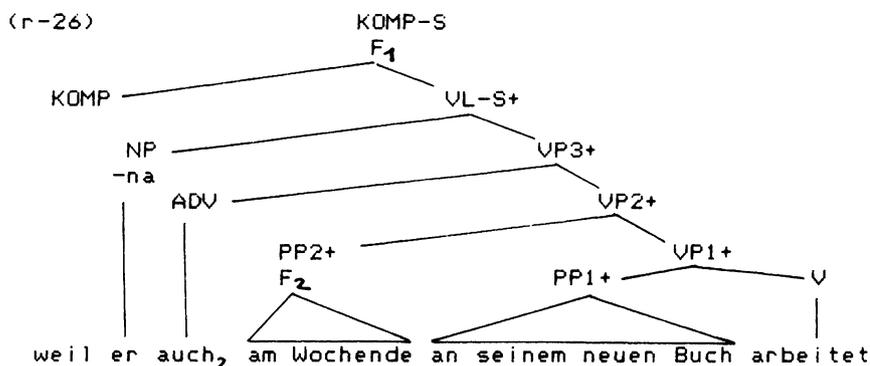
Die bisher formulierten Regeln erfassen nicht die AKzentuierung von nicht-fokusfreie Kokonstituentenmengen, deren Elemente gemeinsam von einem höher bewerteten Fokus dominiert werden, Mengen wie {KOMP,VL-S} in (k-26) also:



Bei (k-26) (Kontext: Warum sieht er am Montag immer so müde aus?) ist eine Betonung wie (n-26) natürlich:

(n-26) weil er auch am ¹Wochende an seinem neuen ¹Buch arbeitet

Dies entspricht der folgenden +-Markierung:



Eine Regel, mit der man (n-26) aus (k-26) erzeugen kann, ist (R6):

(R6) Wenn in einer Menge von Kokonstituenten K die Elemente k_1, \dots, k_n mit F_i ($1 \leq i$) markiert sind oder eine damit markierte Konstituente dominieren, wenn außerdem kein Element von K mit F_j ($j < i$) markiert ist, es aber einen damit markierten Knoten gibt, der alle Elemente von K dominiert, dann wird die +-Verteilung in K wie in der entsprechenden fokusfreien Kokonstituentenmenge einge-richtet, mit der Einschränkung, daß auf jeden Fall k_1, \dots, k_n mit "+" markiert werden.

Sehen wir uns an, wie damit (r-26) erzeugt wird: Alle Kokonstituentenmengen von (K-26) außer {PP1,U} (die von (R5) erfaßt wird) erfüllen die 'input'-Bedingung von (R6): Sie enthalten alle ein Element, das mit F₂ markiert ist oder eine damit markierte Konstituente dominiert, und ihre Elemente werden alle von ein und derselben F₁-Konstituente dominiert, ohne selbst F₁-Konstituenten zu sein oder zu dominieren. Wenn nun in {KOMP,VL-S} die +-Verteilung eingerichtet werden soll, als sei diese Menge fokussfrei, kann (R5) angewandt werden, womit VL-S ein "+" erhält. Die zusätzliche Bedingung, daß diejenigen Konstituenten, die eine F₂-Konstituente dominieren, auf jeden Fall "+" erhalten müssen, ist damit ebenfalls erfüllt, denn dies trifft in {KOMP,VL-S} ja gerade auf VL-S zu. In {NP,VP3} kann im fokussfreien Fall (R4) angewandt werden, was zu einem "+" auf VP3 führt, womit gleichzeitig wieder die Konstituente, die F₂ dominiert, markiert ist. Für {ADV,VP2} wollen wir Integrierbarkeit annehmen (mit VP2 als Argument von ADV), womit (R5) das "+" auf VP2 lenkt und damit wieder auf die Konstituente, die den Partikelfokus dominiert. In {PP2,VP1} schließlich muß im Neutralfall (R4) angewandt werden, was gleichzeitig dazu führt, daß der Partikelfokus, wie von (R6) gefordert, "+" erhält.

Nach (R6) (wie (R3) eine obligatorische Regel) setzt sich also bei Fokusverschachtelung nicht der höher bewertete Fokus gegen den niedriger bewerteten akzentuell durch (vgl. (R3)), sondern es werden beide Foki gleich stark akzentuiert. (Man könnte auch sagen, daß der niedriger bewertete Fokus durch sein Enthaltensein in einem höher bewerteten 'angehoben' wird.)

3.4 Deakzentuierung und kontextuelle Präsenz

Die FHG ist nicht die ~~einzige Ebene der Informationsstruktur, die einen Einfluß auf die Akzentuierung ausübt~~. Wir deuteten in 2.1 schon die Möglichkeit an, daß die Topik-Kommentar-Gliederung mit der Akzentuierung interagiert. Deutlicher ist aber der Einfluß eines anderen informationsstrukturellen Faktors, nämlich der der kontextuellen Präsenz. Als kontextuell präsent bezeichnen wir Konstituenten von geäußerten Sätzen, deren Denotat in der unmittelbaren, d.h. dem Kurzzeitgedächtnis und den Wahrnehmungsorganen der Kommunikationspartner direkt zugänglichen Umgebung der Äußerung bereits vorhanden und auffällig ist, sei es, weil es dort gerade erwähnt wurde, sei es, weil es dort für alle Beteiligten sichtbar, hörbar oder anders wahrnehmbar physisch präsent und 'im Vordergrund' ist (so daß auf es nicht eigens, z.B. durch Zeigegesten, aufmerksam gemacht werden muß). Kontextuell präsente Konstituenten werden normalerweise genau dann deakzentuiert, wenn sie nicht semantischer Fokus sind. Dazu einige Beispiele (der semantische Fokus wird durch tiefgestellte Halbklammern angedeutet, Indizierung wie beim syntaktischen Fokus):

(27) A: Was tat Petra, nachdem sie den Raum betreten hatte, in dem Gerd auf sie wartete?

B: Sie begrüßte¹ Gerd₁.

(28) A: Kennst Du Peter?

B: Nein, ich könnte² nur₂ die Schw¹ester von Peter₁.

(29) (Situation: B trifft A, als A verzweifelt und vergeblich versucht, seinen Schirm zu öffnen.)

A: Was soll ich nur machen?

B: Kauf² dir einen neuen¹ Schirm₁.

(30) A: Wer hat sich über Peters Erfolg am meisten gefreut?

B: Peter¹ (hat sich darüber am meisten gefreut).

(27) - (29) zeigen, daß kontextuell präsente Konstituenten (Gerd, Peter bzw. Schirm) auch dann deakzentuiert werden, wenn sie innerhalb des semantischen Fokus liegen. (30) zeigt, daß solche Konstituenten nicht deakzentuiert werden, wenn sie semantischer Fokus sind.

Bevor wir diskutieren, wie solchen Daten Rechnung getragen werden kann, muß einem naheliegenden Einwand entgegengetreten werden, nämlich dem, die Einführung des Begriffs der kontextuellen Präsenz sei überflüssig, da die fraglichen Deakzentuierungen mit dem bereits eingeführten Begriff 'Hintergrund' erklärt werden könnten. In den obigen Beispielen, so der Einwand, könne die Verteilung von Fokus

und Hintergrund ja so interpretiert werden, daß die deakzentuierten Konstituenten im Hintergrund liegen. Damit müssen sie nach den oben formulierten Regeln deakzentuiert werden.

Leider ist dieser Vorschlag, soweit er sich auf die semantische FHG bezieht, nicht haltbar. Er würde für die fraglichen Beispiele die folgenden semantischen FHG-Verteilungen erfordern:

(27') Sie begrüßte¹₁ Gerd.

(28') Ich kenne nur₂ die Schwester¹₁ von Peter.

(29') Käu² dir einen neuen¹₁ Schirm.

(28') würde nach den Regeln aus 1. implizieren, daß im Kontext (27) die Frage zur Debatte steht, welche auf Gerd gerichteten Handlungen Petra in der beschriebenen Situation vollzogen hat. Tatsächlich steht aber zur Debatte, was Petra gemacht hat, und zwar ohne eine Einschränkung auf Handlungen, die Gerd affizieren. (Deswegen hätte B in (27) z.B. auch antworten können: Sie setzte sich auf's Sofa.) Analog - und noch deutlicher - in (28): (28') würde unter anderem implizieren, daß mit der B-Äußerung in (28) die Proposition zum Ausdruck gebracht wird, daß Peters Schwester die einzige Verwandte oder Bekannte von Peter ist, die der Sprecher kennt, wobei offen gelassen wird, ob er Peter kennt. Das ist natürlich falsch: B bringt in (28) zum Ausdruck, daß er Peter nicht kennt und daß er von allen in Frage kommenden Alternativen zu Peter (die nicht unbedingt Verwandte oder Bekannte von ihm sein müssen) eben nur Peters Schwester kennt. (29') schließlich würde vollkommen kontraintuitiv besagen, daß es in (29) um die Frage geht, welchen Schirm A sich kaufen soll. (Tatsächlich steht ja nur zu Debatte, was A tun soll, um sein Schirm-Problem zu lösen.)

Bei Beibehaltung unserer FHG-Semantik scheint es also ausgeschlossen, die fraglichen Deakzentuierungen durch passende 'Verengungen' des semantischen Fokus zu erklären und dadurch den Begriff der kontextuellen Präsenz zu eliminieren. Man könnte natürlich versuchen, die FHG-Semantik geeignet zu ändern. Jeder solche Versuch würde jedoch m.E. auf eine Variante dessen hinauslaufen, was wir in 1.4 als traditionelle FHG-Auffassung bezeichnet haben, nämlich auf eine Theorie, die die Unterscheidung zwischen Fokus und Hintergrund inhaltlich auf die zwischen kontextuell präsenter und kontextuell nicht präsenter Information zurückführt. Argumente gegen eine solche Theorie wurden bereits in 1.4 angeführt (ausführlicher in Jacobs 1984). Ein Hauptproblem ist, daß das, was man aus unabhängigen Gründen als Hintergrund und Fokus bezeichnen möchte, nicht notwendigerweise kontextuell präsent bzw. kontextuell nicht präsent Material beinhaltet. In (30) ist kontextuell präsent Material im Fokus, und in (k-26) ist kontextuell nicht präsent Material im Hintergrund, nämlich in dem zum Fokus von auch komplementären Teil an seinem neuen Buch arbeitet. (Man beachte, daß man dies in jeder Theorie als den zum Partikelfokus informationell komplementären Teil, also als Hintergrund, betrachten muß, es sei denn, man verzichtet auf eine

adäquate Analyse der mit der Partikel verbundenen Quantifikation. Vgl. Jacobs 1983.)

Es bleibt die Möglichkeit, die angestrebte Fokusverengung als eine des syntaktischen Fokus zu betrachten, womit nach unseren Regeln die fraglichen Deakzentuierungen ebenfalls zu erfassen wären. Dagegen gäbe es kaum schwerwiegende Argumente. Insbesondere könnte man es nicht als Einwand gelten lassen, daß dadurch die Isomorphie von semantischem und syntaktischem Fokus zerstört würde: Eine strenge Isomorphie besteht ja ohnehin nicht, vgl. z.B. (6) in 2.1. Ein zunächst überzeugenderes Gegenargument wäre der Hinweis auf Daten wie (31):

(31) A: Warum heiratet Peter Gerda nicht?

B: Er ist mit Luise¹ verheiratet.² / *Mit Luise¹.

Die B-Antwort kann hier nicht lauten Mit Luise. Bei einer Ausparung des kontextuell präsenten Materials würde der syntaktische Fokus aber nur noch diese PP umfassen. Damit wäre aber wohl zu erwarten, daß das umliegende Material genauso getilgt werden kann wie nach der Frage Mit wem ist Peter verheiratet?, bei der ebenfalls nur die PP im syntaktischen Fokus liegt. (Wir setzen voraus, daß gleiche FHG bei auch sonst gleichen Sätzen gleiche Tilgungsmöglichkeiten impliziert.)

Diese falsche Voraussage würde sich aber tatsächlich nur dann ergeben, wenn wir die Domänen möglicher Ellipsen an der syntaktischen FHG festmachen (wie wir es versuchsweise in 2.1 taten). Wenn wir stattdessen hierzu die semantische FHG heranziehen, könnten wir gleichen syntaktischen Fokus in beiden Kontexten annehmen, ohne das unterschiedliche Tilgungsverhalten unerklärt lassen zu müssen. (Der semantische Fokus in (31B) ist der ganze Satz. In dem anderen genannten Kontext umfaßt er dagegen nur die PP.) So geben uns Ellipsedaten zwar einen weiteren Hinweis darauf, daß man kontextuell präsenten von im Hintergrund befindlichem Material unterscheiden muß, aber sie hindern uns nicht daran, kontextuell präsenten Material aus dem syntaktischen Fokus auszusparen.

Technisch könnte eine solche Ausparung durch entsprechende Änderungen von (R1) und (R3) in 2.1 eingeführt werden. Da dies recht aufwendig wäre, möchte ich im folgenden einen anderen, vielleicht einfacheren Weg der Behandlung der Deakzentuierung von kontextuell präsentem Material skizzieren, bei dem weder der semantische noch der syntaktische Fokus verengt wird. Unser Modell enthält ja bereits ein technisches Mittel zur Berücksichtigung von Deakzentuierungen, nämlich das Merkmal *-na*, vgl. 3.3. Dieses Mittel ist aus unabhängigen Gründen nötig und kann durch eine Theorie, die kontextuell präsenten Material aus dem Fokus verbannt, nicht eliminiert werden. So muß jemand als *-na* ausgezeichnet werden, obwohl es im Normalfall nicht kontextuell präsent sind:

(32) A: Was wolltest du dort?

B: Ich ²wollte jemand ¹besuchen.

Trotz klarer kontextueller Nicht-Präsenz und Enthaltensein im semantischen Fokus wird hier jemand deakzentuiert.

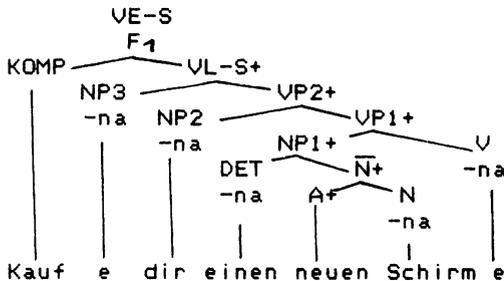
Einen einfachen Weg zur Beschreibung der Deakzentuierung von kontextuell präsentem Material mit Hilfe von -na weist die folgende Generalisierung:

(R7) Kontextuell präzente Konstituenten sind -na.

Wenn sich (R7) halten läßt, hat dies zunächst den Vorteil, daß sich die Liste von Konstituenten, die grundsätzlich -na sind (vgl. 3.3), teilweise aus (R7) ableiten läßt: Daß (nicht-demonstrative) Personalpronomina, Pro-Adverbien etc. unter Neutralbedingungen grundsätzlich nicht akzentuierbar sind, folgt nach (R7) aus ihrem notwendig kontextuell präsentem Status. Demonstrativa sind dagegen kontextuell nicht-präsent - die mit ihnen verbundene Zeigegeste impliziert das ja gerade - und entsprechend auch im Neutralfall akzentuierbar. (Bei den anderen Elementen der Liste spielen wohl Faktoren wie semantische 'Leere' und geringes phonologisches Gewicht eine Rolle.)

Mit (R7) können wir darüberhinaus auch die Deakzentuierung von nicht-notwendig kontextuell präsenten Konstituenten in den Griff bekommen: Wir erlauben einfach, daß das Merkmal -na fakultativ auch solchen Konstituenten zugewiesen wird und sorgen dafür, daß die jeweiligen Sätze im Sinne von (R7) interpretiert werden, d.h. daß in der Abteilung der Semantik (oder Pragmatik), die Sätzen passende Äußerungskontexte zuweist, nur solche Paare aus Sätzen und Kontexten zugelassen werden, die (R7) erfüllen. Damit, so scheint es, können wir die fraglichen Fälle mit den bereits formulierten Akzentuierungsregeln analysieren. Für (29B) würde sich z.B. die folgende Struktur ergeben (die voraussetzt, daß Imperative ein leeres Subjekt haben):

(r-29)



Dies ergibt genau die Betonung von (29B). Der Leser möge sich davon überzeugen, daß zur Erzeugung der +-Verteilung in (r-29) die bereits formulierten Regeln (R4) und (R5) ausreichen. Insbesondere entspricht die +-Verteilung in {A,N} der Regel (R4). ((R5) ist hier nicht anwendbar, weil N -na ist.)

Dieser Vorschlag beinhaltet leider gewisse technische Probleme. (So muß in bestimmten Fällen -na an höhere Konstituenten vererbt werden, die nicht zur Gänze kontextuell präsent sind. Außerdem muß die Akzentverteilung in KoKonstituentenmengen, die nur -na-Elemente enthalten, geklärt werden.) Darüberhinaus stellt sich die Frage, ob (R7) tatsächlich haltbar ist. Zweifel daran lassen z.B. Possessivpronomina aufkommen, die genau wie Personalpronomina auf im Kontext hinreichend auffallende Entitäten verweisen, doch anders als diese im Neutralfall (und im exaltierten Stil oder bei

Exklamativen) durchaus akzentuierbar sind:

(33) Mit ¹deinem ¹Einkommen kannst du dir das nicht ¹leisten.

Vielleicht liegt dies daran, daß Possessivpronomina neben der deiktischen auch noch eine nicht auf kontextuell gegebene Faktoren verweisende Bedeutungskomponente haben, nämlich die Possessivrelation.

Zur Klärung solcher Fragen benötigte man eine wesentlich genauere Bestimmung der kontextuellen Präsenz als unsere obige. Auf der Basis einer Präzisierung dieses Begriffs könnte man dann auch versuchen, gewisse 'puzzles' der Akzenttheorie zu klären, so die Frage, warum Komplexe aus 'externen' Subjekten und Verben im Gegensatz zu der in 3.3 formulierten Regel manchmal doch integrierbar zu sein scheinen:

(34) A: Was ist da los?

B: Ein ¹Hund ²bellt.

Dem Subjekt von (34B) fehlen die meisten der für nicht-agentivische 'interne' Subjekte typischen Eigenschaften (vgl. Grewendorf 1986). Dennoch zieht es, wie für solche Subjekte typisch, bei Gesamtsatzfokussierung die Hauptbetonung auf sich. Wenn man dennoch die Regel, daß 'externe' Subjekte mit dem Verb nicht integrierbar sind, beibehalten möchte, könnte man versuchen, hierin keinen Fall integrierender Akzentuierung, sondern einer Deakzentuierung des Verbs aufgrund kontextueller Präsenz zu sehen. In der Sprechsituation des Dialogs (34) kann man sich ja ein von A und B gemeinsam wahrgenommenes Geräusch vorstellen. Dies bewirkt zusammen mit der Erwähnung eines Hundes durch den ersten Teil der B-Außerung, daß eine Fortsetzung mit bellt mit hoher Wahrscheinlichkeit erwartbar wird. Wenn man diese Erwartbarkeit als eine Form kontextueller Präsenz deuten will, subsumieren sich solche Fälle unter (R7).

Ob man so aber auch Beispiele wie (35) erklären kann, ist fraglich:

(35) A: Woher hast du denn das tolle Armband?

B: Das hat mir ¹Gerda ²geliehen.

Eindeutig ist hier der Komplex Gerda geliehen im semantischen Fokus und Gerda agentivisches Subjekt. Wenn auch hier keine Integrierbarkeit von Subjekt und Verb vorliegen soll, müßte man nach oben Gesagtem geliehen als kontextuell präsent betrachten. Doch gibt es dafür unabhängige Argumente? Erwartbar im Sinne von (34) ist geliehen hier jedenfalls nicht.

Solche Beispiele machen erneut deutlich, daß ein Hauptproblem der Akzenttheorie die Prinzipien neutraler, d.h. nicht fokusmarkierender Akzentuierung sind. Zur Lösung dieses Problems konnte und wollte das Vorangehende nur wenig beitragen.

LITERATUR

- Abraham, W.,/S. de Meij (Hg.). 1986. Topic, focus, and configurationality. Amsterdam, Benjamins.
- Allerton, D.J., A. Cruttenden. 1979. Three reasons for accenting a definite subject. *Journal of Linguistics* 15. 49-53.
- Altman, H. 1981. Formen der Herausstellung im Deutschen. Tübingen, Niemeyer.
- Atlas, J.D., S. Levinson. 1981. It-clefts, informativeness and logical form. *Radical pragmatics*, hg. von P. Cole. New York, Academic Press. 1-61.
- Bayer, J. 1987. The syntax of scalar particles and so-called 'floating quantifiers'. Ms. Nijmegen.
- Behaghel, O. 1932. *Deutsche Syntax. Eine geschichtliche Darstellung*. Bd. 4. Heidelberg, Carl Winter.
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. *Subject and topic*, hg. von Ch. Li. New York, Academic Press.
- Cresswell, M.J. 1985. *Structured meanings. The semantics of propositional attitudes*. MIT Press.
- Cresswell, M.J., A. v. Stechow. 1982. De re belief generalized. *Linguistics and Philosophy* 5. 503-535.
- Czepluch, H. 1987. Lexikalische Argumentstruktur und syntaktische Projektionen: zur Beschreibung grammatischer Relationen. *Zeitschrift für Sprachwissenschaft* 6. 3-36.
- Fuchs, A. 1976. 'Normaler' und 'kontrastiver' Akzent. *Lingua* 38. 293-312.
- 1980. Accented subjects in 'all-new' utterances. *Wege zur Universalienforschung*, hg. von G. Brettschneider u. Chr. Lehmann. Tübingen, Narr.
- Givón, T. 1985. The pragmatics of word-order: predictability, importance and attention. Ms. Eugene.
- Grewendorf, G. 1986. Ergativität im Deutschen. *Vervielfältigtes Vorlesungs-Ms.* Frankfurt/M.
- Harries-Delisle, H. 1978. Contrastive emphasis and cleft-sentences. *Universals of human language*, Bd. 4, hg. von J.H. Greenberg. Stanford U.P.
- Hetzron, R. 1975. The presentative movement or why the ideal word order is VSOP. *Word order and word order change*, hg. von Ch. Li. Austin U.P.
- Höhle, T. 1982. Explikationen für 'normale Betonung' und 'normale Wortstellung'. *Satzglieder im Deutschen*, hg. von W. Abraham. Tübingen, Narr.
- 1986. *Verum-Fokus*. Ms. Tübingen.
- Jacobs, J. 1982. *Syntax und Semantik der Negation im Deutschen*. München, Fink.
- 1982a. Neutraler und nicht-neutraler Satzakkzent im Deutschen. *Silben, Segmente, Akzente*, hg. von T. Vennemann. Tübingen, Niemeyer.
- 1983. *Fokus und Skalen. Zur Syntax und Semantik der Gradpartikeln im Deutschen*. Tübingen, Niemeyer.
- 1984. Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91. 25-58.
- 1986. The syntax of focus and adverbials in German. In Abraham/de Meij 1986.
- 1987. *Negation*. Ms. München. Erscheint in: *Handbuch Semantik*, hg. von A. v. Stechow u. D. Wunderlich. Königstein/Ts., Athenäum.

- 1987a. Probleme der 'freien' Wortstellung im Deutschen. Ms. München.
- 1987b. On the semantics of modal particles. Ms. München. Erscheint in: Discourse particles, hg. von W. Abraham. Amsterdam, Benjamins.
- De Jong, J. 1980. On the treatment of focus phenomena in functional grammar. Perspectives on Functional Grammar, hg. von T. Hoekstra, H. van der Hulst u. M. Moortgat. Dordrecht, Foris. 89-116.
- É. Kiss, K. 1980. Structural relations in Hungarian, a 'free' word order language. Linguistic Inquiry 12. 185-213.
- Krifka, M. Zur semantischen und pragmatischen Motivation syntaktischer Regularitäten. München, Fink.
- 1984. Fokus, Topik, syntaktische Struktur und semantische Interpretation. Ms. München.
- Kuno, S. 1980. The scope of question and negation in some verb-final languages. Papers from the 16th regional meeting. Chicago, CLS.
- Lang, E. 1984. The semantics of coordination. Amsterdam, Benjamins.
- Lieb, H. 1983. Akzent und Negation im Deutschen. Linguistische Berichte 84 und 85. 1-32 bzw. 1-48.
- Lötscher, A. 1983. Satzakkzent und Funktionale Satzperspektive im Deutschen. Tübingen, Niemeyer.
- Reinhart, T. 1981. Pragmatics and linguistics: an analysis of sentence topics. Philosophica 27. 53-94.
- Reis, M. 1977. Präsuppositionen und Syntax. Tübingen, Niemeyer.
- 1987. Die Stellung der Verbargumente im Deutschen. Stilübungen zum Grammatik:Pragmatik-Verhältnis. Sprache und Pragmatik. Beiträge zum 5. Lunder Symposium, hg. von I. Rosengren. Stockholm, Almqvist & Wiksell. 139-177.
- Rochemont, M.S. 1986. Focus in Generative Grammar. Amsterdam, Benjamins.
- Rooth, M. Association with focus. Ph.D.-Dissertation. University of Massachusetts, Amherst.
- Selkirk, E.O. Phonology and syntax. The relation between sound and structure. MIT-Press.
- Sperber, D., D. Wilson. 1986. Relevance. Oxford, Blackwell.
- v. Stechow, A. 1982. Structured proposition. Arbeitspapier des SFB 99. Universität Konstanz.
- 1986. Alternative semantics for questions. Ms. Wellington. Erscheint in: Discourse particles, hg. von W. Abraham. Amsterdam, Benjamins.
- v. Stechow, A., S. Uhmann. 1984. On the focus-pitch accent relation. Groninger Arbeiten zur Germanistischen Linguistik 25. 228-263.
- 1986. Some remarks on focus projection. In Abraham/de Meij 1986.
- Uhmann, S. 1987. Fokussierung und Intonation. Diss. Universität Konstanz.
- Uszkoreit, H. 1984. Word order and constituent structure in German. Ph.D.-Diss. University of Texas, Austin.
- Zaefferer, D. 1984. Frageausdrücke und Fragen im Deutschen. München, Fink.

Leila Luukko-Vinchenzo (München)

0. 'Fragesatz' im Satzmodus-System des Finnischen

In den finnischen Grammatiken werden gewöhnlich vier Satztypen unterschieden: 1) väite-/deklaratiivilause 'Aussage-/Deklarativsatz'¹ 2) kysymys-/interrogatiivilause 'Frage-/Interrogativsatz' 3) käsky-/imperatiivilause 'Heische-/Imperativsatz' 4) huudahdus-/eksklamatiivilause 'Ausrufe-/Exclamativsatz'.

Hakulinen/Karlsson (1979:276) bemerken die Ungenauigkeit dieser Terminologie, mit der nicht genügend zwischen Form und (Rede-!)Funktion eines Ausdrucks differenziert werden kann. Als Beispiele geben sie an, daß ein Deklarativsatz eine Frage, ein Interrogativsatz einen Befehl etc. ausdrücken kann (wohl korrekter: der Sprecher kann mit einem Deklarativsatz eine Frage etc. ausdrücken!). Mit ihrer Kritik deuten die Verfasser auf die Probleme einer allzu direkten Zuordnung von Formtypen und kontextabhängigen Interaktionstypen; sie halten es aber - und dies im Hinblick auf die Vielzahl der letztgenannten sicher mit Recht - für mühsam und oft für unnötig, Form und Redefunktion terminologisch auseinanderzuhalten, und übernehmen die traditionellen linguistischen Mischbezeichnungen 'Ausagesatz', 'Fragesatz' etc.

Ich möchte Fragesätze des Finnischen im Rahmen eines ähnlichen Satzmodus-Systems beschreiben, wie es für das Deutsche H. Altmann entwickelt hat (vgl. Altmann (1984); (1987)). Von dieser kontrastiven "Brücke" erhoffe ich mir neue Einsichten in die finnischen Satzmodi, die bisher von Matihaldi (1979) am ausführlichsten behandelt worden sind. I. S. v. Altmann verstehe ich unter 'Satzmodus'

(...) die regelmäßige Zuordnung eines Satztyps (oder einer Gruppe von Satztypen) mit angebbaren formalen Eigenschaften zu einer bestimmten Art von Funktion (oder zu einer Gruppe von Funktionen) im sprachlichen Handeln, die ich Funktionstyp nennen will (...). "Satzmodus" bezeichnet also ein komplexes sprachliches Zeichen mit einem Form- und einem Funktionsaspekt. (Altmann 1987:22)

Als zu weit führend muß ich an dieser Stelle auf eine nähere Vorstellung dieses hier für das Finnische adaptierten Systems verzichten und mich auf die Benennung der darin enthaltenen fünf Haupt-Satzmodi beschränken:

- 1) Aussage-Satzmodus (vgl. Oppenrieder (1987))
- 2) Frage-Satzmodus (vgl. Luukko (1985); (1987))
- 3) Imperativ-Satzmodus (vgl. Winkler (1986))
- 4) Exclamativ-Satzmodus (vgl. Netter (1982))
- 5) Wunsch-Satzmodus (vgl. Scholz (1987))

Der Frage-Satzmodus läßt sich ähnlich wie der Exclamativ-Satzmodus ferner einteilen erstens in die Formtypen mit einem *m/k*-Ausdruck (im Deutschen entsprechend mit einem 'w-Ausdruck') wie *mikä* 'was', *millä tavalla* 'auf welche Weise', *kuka* 'wer', *kuinka* 'wie' etc. und zweitens in die Typen, die durch andere syntaktische Mittel, z.B. durch die Fragepartikel *-ko* '??', gekennzeichnet sind. Analog zu der deutschen Bezeichnung 'w-Fragesatz' verwende ich im Finnischen für den ersteren Formtyp 'm/k-Fragesatz' und für den letzteren 'Satz-Fragesatz'.

Diese beiden Haupttypen haben gemeinsam, daß der Sprecher mit ihnen ausschließlich aufgrund ihrer grammatischen Struktur seine Frageeinstellung zu erkennen geben kann (i.S.v. Lang (1981) und (1983)). Die ausgedrückte propositionale Grund-Einstellung kann für die *m/k*-Fragesätze mit 'wissen wollen, *m/k-* ('w-')' und für die Satz-Fragesätze mit 'wissen wollen, *-ko* ('ob')' beschrieben werden (vgl. Altmann (1987) und Bierwisch (1979)). In beiden Frage-Satzmodi befinden sich Unter-Typen (z. B. deliberativer Fragesatz: (1)) und Misch-Typen (z. B. assertiver Fragesatz: (2)), die nach ihren Formen und Funktionen in das Gesamt-System eingeordnet werden können und auf die bei den entsprechenden Beispielen weiter unten im Rahmen der hier vorhandenen Möglichkeiten eingegangen wird:

- | | |
|------------------------|-------------------------------|
| (1) Tuleekohan Matti? | - deliberativer Fragesatz mit |
| 'komm-?wohl Matti' | -han |
| 'Ob Matti wohl kommt?' | - (sich) fragen, ob (wohl) |
| (2) Kai Matti tulee? | - assertiver Fragesatz mit |
| 'wohl Matti komm-' | kai |
| 'Matti kommt?' | - sagen, daß + fragen, ob |

Im Hinblick auf die bisherigen Untersuchungen (vgl. z. B. Tarvainen (1985)), die zeigen, daß das Deutsche und das Finnische trotz ihrer typologischen Verschiedenheit größtenteils ähnliche grammatische Kategorien und Strukturen verwenden - wie beispielsweise die oben angesprochenen *m/k-* und *w-*Ausdrücke -, halte ich es für plausibel, die von Altmann (1987:30) vorgeschlagenen syntaktischen Mittel Reihenfolgebeziehungen, morphologische Markierung, kategoriale Füllung und die intonatorische Markierung daraufhin zu untersuchen, ob sie in ihren verschiedenen Kombinationen auch im Finnischen Satzmodi unterschiedlicher Art und Hierarchie bedingen. Hierzu werde ich im folgenden unter Einbeziehung der übrigen syntaktischen Merkmale versuchen, vor allem die Rolle der bisher insgesamt wenig untersuchten intonatorischen Merkmale bei der Aufstellung von Satzmodi, und insbesondere von Frage-Satzmodi, herauszuarbeiten.

1. Bisherige Untersuchungen zur Intonation finnischer Fragesätze

Das Finnische wird oft als eine Sprache bezeichnet, in der andere syntaktische Mittel für eine genügende Kennzeichnung der Satztypen sorgen würden, so daß der Intonation hier keine distinktive Funktion mehr zukäme (vgl. u.a. Östman (1977)). Es steht jedoch gerade die Intonation der Fragesätze im Mittelpunkt der meisten - zahlenmäßig insgesamt allerdings wenigen - satzintonatorischen Untersuchungen zum Finnischen.

Kallioinen (1968) schreibt der Intonation im Finnischen keine distinktive, sondern lediglich eine expressive Funktion zu. Die expressive Intonation ist lt. Kallioinen (ebd., 44) verbunden mit einem Nukleus von erhöhter Grundfrequenz, Intensität und Dauer. Dieser Nukleus ist u. a. in einer 'die Verwunderung ausdrückenden Pseudofrage' zu finden, wofür als Beispiel der kontextlos und auf Anweisung neutral bzw. verwundert gelesene Satz *hän on tullut* 'Er/Sie ist gekommen' aufgeführt wird.

Freihoff (1974) verneint ebenfalls jegliche distinktive Funktion der Intonation: 'Die Intonation kann keine entscheidende Bedeutung haben bei der Identifikation eines Fragesatzes' (ebd., 69). Ohne Beispiele zu geben, erweitert er seine Behauptung über das Finnische hinaus auch auf das Englische und das Deutsche (seine Muttersprache). Die Intonation hat Freihoffs Ansicht nach insgesamt nur eine abtönende - bspw. eine drohende, Unsicherheit oder Interesse anzeigende etc. - Funktion.

Zu einem anderen Ergebnis kommt Hirvonen (1970):

Intonation also has a distinctive function in Finnish. The intonation of general questions, particular questions, and imperatives differs markedly from the normal breath-group both perceptually and acoustically. Hirvonen (1970:39)

Hirvonen schränkt die Distinktivität der Intonation allerdings ein, indem er auf die anderen Mittel zur Kennzeichnung der o. g. "Appell-Satztypen" hinweist, nämlich auf das Fragewort oder die -partikel bzw. auf den imperativen Verbmodus. Er führt jedoch Beispiele auf, in denen das gemeinsame Merkmal der "Appell-Sätze", ein hoher F_0 -peak am Satzanfang, das einzige unterscheidende Merkmal bleiben würde ((3) kontradiktorischer Aussagesatz vs. (4) kolloquialer Fragesatz) (vgl. Hirvonen 1970:47):

- (3) Osaat sä uida.
'könn-du schw.'
'Du kannst (sehr wohl) schwimmen.'
- (4) Osaatsä uida?
'könn-du schw.'
'Kannst du schwimmen?'

Iivonen (1978) führt sechs diskurs- und konnotationsabhängige F_0 -Konturen auf, die er in spontanen Äußerungen festgestellt hat. Keine der aufgezeichneten Grobkonturen - von denen jeweils eine mit "high initial" bzw. mit "extra high initial" beschrieben wird - kann lt. Iivonen (ebd., 52) jedoch nur in Fragesätzen vorkommen. Iivonen bezeichnet gewisse, in Fragesätzen häufig auftretende Grundfrequenzmuster als optionale Mittel zur Kennzeichnung der interrogativen Funktion einer Äußerung und verneint die Existenz einer speziellen Frageintonation im Finnischen. Die Höhe des satzinitialen F_0 -Gipfels (vgl. Hirvonen) ist vielmehr abhängig von mehreren Faktoren (vgl. Iivonen 1978:50):

<u>higher F_0 peak</u>	<u>lower F_0 peak</u>
question	statement
indicative mood	conditional mood (-isi-)
contrastive stress	thematic stress
word with the suffix /-han/	no suffix
word with the suffix /-kin/	no suffix
noun	verb

In Iivonen (1983) wird die Höhe der satzinitialen F_0 -Kurve jedoch auch als vom Akzentmuster der Gesamtäußerung abhängig bezeichnet, so daß bspw. ein Kontrastakzent im späteren Verlauf der Äußerung den initialen F_0 -Gipfel herabsetzen kann. Ferner weisen längere sowie "laut ausgerufene" Äußerungen einen höheren F_0 -peak auf als kürzere und "normal" gesprochene Äußerungen.

Iivonen (1981) stellt eine Bestandsaufnahme verschiedener 'hörbarer Züge der Tonführung' (ebd., 102) in finnischen Fragesätzen dar. Er führt mehrere Beispiele - allerdings ohne Oszillomink-Aufzeichnungen o.ä. - mit finalem "rise" auf, bemerkt aber, daß dieser Typ nicht der üblichste in finnischen Fragesätzen ist.

In den hier kurz skizzierten Untersuchungen ist die verwendete linguistische Terminologie oft unpräzise (vgl. Kallioinens 'eine die Verwunderung ausdrückende Pseudofrage'). Ebenso sind fast alle analysierten Sätze kontextlos produziert worden, was unter Umständen die Natürlichkeit und später die Identifikation stark beeinträchtigen kann. Speziell Iivonens Betrachtung des äußerungsinitialen F_0 -Gipfels enthält auch viele syntaktische, morphologische und phonetische Aspekte, die einerseits über die reine Satzmodus-Problematik hinausreichen, die andererseits jedoch nur Frage- und Aussagesätze betreffen. Dem Untersuchungsbereich der vorliegenden Arbeit kommen die von Hirvonen erwähnten, auf intonatorischen Unterschieden beruhenden Satztypen-Oppositionen (vgl. Satz (3) und (4)) nahe, die von ihm aber nicht näher behandelt werden.

2. Zur Untersuchung der Intonation im finnischen Satzmodus-System

Eine befriedigende Antwort auf die oft gestellte Frage, ob der Intonation im Finnischen eine distinktive Funktion zukommt, kann m. E. zuverlässig nur über intonatorische Minimalpaare gefunden werden, i.e. über funktional verschiedene Ausdrücke, die sich formal nur durch ihre Intonation unterscheiden. Soweit möglich, soll die Realisierung des Hauptakzents in solchen Minimalpaaren stets an der gleichen Stelle erfolgen, was durch eine Situationsbeschreibung und - wenn zusätzlich nötig - durch Einbettung in einen eindeutigen sprachlichen Kontext gesteuert werden kann.

Hier wurden zwei Korpora dieser intonatorischen Minimalpaare mit 2 - 11 Oppositionen gebildet (etwa: Ergänzungsfragesatz (5) vs. *m/k*-Exclamativsatz (6) vs. Versicherungsfragesatz (7) vs. rhetorischer Fragesatz vs. Schlagzeile etc.). Das erste Korpus enthält hauptsächlich Sätze mit einem ausdrucksinitialen *m/k*-Wort und das zweite vornehmlich Sätze, in denen der Fokusbereich variiert wurde.

- (5) 'Situation: Der Sprecher kann die Dauer der Reise nicht schätzen und fragt im Reisebüro:
 Sprecher: "Kuinka kauan matka voi kestää?"
 wie lange Fahrt/könn-dauer-Reise
 "Wie lange kann die Reise dauern?" '
- (6) 'Situation: Dem Sprecher wird die Dauer der Fahrt ungemütlich:
 Sprecher: "Voi kuinka kauan matka voi kestää!"
 ach/oh/ wie lange Fahrt könn-dauer-weh
 "Oh weh, wie lange die Fahrt dauern kann!" '
- (7) 'Situation: Zwei Sprecher unterhalten sich:
 Sprecher 1: "Die Reise kann sogar 25 Jahre dauern."
 Sprecher 2: "Das kann nicht wahr sein!"
 "Kuinka kauan matka voi kestää?!?"
 wie lange Reise könn-dauer-
 "Die Reise kann wie lange dauern?!?" '

Zu den Aufnahmen:

Die Aufnahmen erfolgten im schalltoten Raum der Universität Turku. Die Sprecher waren 2 weibliche und 3 männliche Studenten, die die folgenden Voraussetzungen erfüllten: 23-29 Jahre alt, Geburtsort sowie Wohnort über längere Zeit im südlichen/süd-westlichen Finnland, ohne starken dialektalen Hintergrund sowie ohne häufigen Kontakt mit Fremdsprachen (keine Sprachstudenten o. ä.). Durch die zuletzt genannte Bedingung sollten die häufig angesprochenen intonatorischen Interferenzerscheinungen weitgehend eliminiert werden. Von den vier Sprechern des ersten Korpus konnten drei für das zweite Korpus wiedergewonnen werden, so daß insgesamt 5 Sprecher analysiert wurden.

Zum Material:

Das Material umfaßt insg. 94 auf DIN₂A6-Karten geschriebene Analysesätze, die randomisiert und möglichst natürlich² jeweils zweimal von vier Sprechern produziert wurden. Außerdem wurden die erlaubten sofortigen Selbst-Korrekturen und

die von der Aufnahmeleiterin erbetenen Wiederholungen am Ende der jeweiligen Aufnahme berücksichtigt. Das somit gesammelte Material besteht aus 895 Beiträgen³, zu denen Oszillomink-Aufzeichnungen mit Grundfrequenz-Verlauf, Amplitude und Oszillogramm angefertigt wurden.

Zu den durchgeführten Tests:

Die 895 Sätze, die dieser Satzmodus-Untersuchung zugrundeliegen, wurden aus ihren Kontexten herausgeschnitten und auf gesonderte Testbänder umgespielt, die sich durchschnittlich 8 Versuchspersonen im Rahmen der Tests zur Hauptakzentlage und zur Satzmodus-Identifikation anhörten. Beim Akzent-Test hatten sie pro Satz die Silbe zu kennzeichnen, die sie für am meisten betont hielten.⁴ Beim Identifikations-Test mußten sie die kontextlosen Sätze nach der Funktion (nicht nach der Form) erkennen als 1) Aussage 2) Befehl 3) Frage 4) Exclamativ oder 5) Wunsch, wobei ihnen vorab anhand von Beispielen die einzelnen Funktionen erklärt wurden. In einem dritten Test, dem Natürlichkeits-Test, wurden die Sätze in ihren Kontexten Versuchspersonen vorgespielt, die sich vorher mit der jeweiligen Karte bekanntmachen konnten, so wie diese dem Sprecher vorgelegt worden war. Nach Abhören des Redebeitrags konnten sie die Natürlichkeit des für diesen Test besonders gekennzeichneten Testsatzes in der bestimmten Situation und in dem bestimmten Kontext beurteilen als 1) "luonnollinen" 'natürlich' 2) "menettelee" 'noch natürlich' 3) "epäluonnollinen" 'unnatürlich'.

Bei der folgenden Besprechung der Ergebnisse werden - wenn nicht anders vermerkt - intonatorische Eigenschaften von Sätzen diskutiert, die von mind. 75% der Versuchspersonen richtig identifiziert und als "natürlich" oder "noch natürlich" beurteilt wurden. Dieses Verfahren soll zumindest teilweise die im Studio herrschende Aufnahme-Atmosphäre kompensieren, in der nur schwer die Ungezwungenheit einer alltäglichen Rede-Situation erreicht werden kann.

2.1. Intonation in *m/k*-Fragesätzen

Die Intonation in finnischen Fragesätzen mit einem obligatorischen Frageausdruck am Satzanfang beschreibe ich im folgenden hauptsächlich anhand des Satzes

Kuinka kauan matka voi kestää.
'wie lange Fahrt/könn-dauer-'
Reise

Dabei berücksichtige ich die akustischen Parameter Grundfrequenz, Amplitude und Dauer, welche zunächst für die Fragesatz-Typen 'Ergänzungsfragesatz', 'Versicherungsfragesatz' und 'rhetorischer *m/k*-Fragesatz' behandelt und abschließend mit dem nicht-erotetischen Satzmodus '*m/k*-Exclamativsatz' kontrastiert werden.

Dadurch soll verdeutlicht werden, daß das Satzmodus-System außer dem Grundtyp 'Ergänzungsfragesatz' andere formal ähnliche Typen von Fragesätzen sowie Nicht-Fragesätzen aufweist, die sich jedoch zumindest bezüglich ihres Merkmals 'Intonation' formal und bezüglich ihrer Funktion zumindest teilweise unterscheiden.

Wegen der individuellen Grundfrequenz-Bereiche, in denen sich die einzelnen Sprecherinnen und Sprecher bewegen, wird in den Ausführungen auf genaue Hz-Anga-

ben meistens verzichtet. In den vorliegenden Vergleichen der Satzmodi sind prozentuale Daten ohnehin adäquater, damit ein möglichst allgemein gültiges und von der lexikalischen Füllung unabhängiges Modell erzielt werden kann.

2.1.1. Intonation eines Ergänzungsfragesatzes

Der Grundtyp der *m/k*-Satzmodi, der Ergänzungsfragesatz (vgl. (5)), dient – wie oben besprochen – zum Ausdrücken der Sprechereinstellung 'wissen wollen, *m/k*-', wozu auf der formalen Seite u. a. das syntaktische Merkmal 'Intonation' benötigt wird. Diese wird im folgenden anhand mehrerer Einzelkennzeichen als für diesen Satzmodus typisch identifiziert.

- Die Hauptakzentsilbe befindet sich am Satzanfang; trotz der bei bei der Akzentbeurteilung vorkommenden Streuung, die wohl nicht zuletzt auf die Satzlänge zurückzuführen ist, stellte sich *kuin* als die am meisten akzentuierte Silbe heraus. Darauf fällt ausnahmslos der höchste F_0 -Gipfel und eine konvexe F_0 -Bewegung, der höchste Amplitudengipfel ist mit dieser Silbe aber nicht immer verbunden.
- Der Ansatzpunkt des F_0 -Verlaufs, der "Onset", ist mittelhoch im Verhältnis zu den "Onsets" in anderen *m/k*-Satztypen desselben Sprechers.
- Nach dem F_0 -Maximum auf der 1. Silbe (*kuin*) sinkt die F_0 auf dem Frageausdruck kontinuierlich weiter, so daß sich auf *kuinka kauan* die 1. konvexe Großbewegung befindet, in deren Bereich auch der höchste Amplitudengipfel fällt. Für den Umfang dieser F_0 -Bewegung spricht ihr Verhältnis zur F_0 innerhalb der Äußerung: diese beansprucht ca. 60-80% (sprecherabhängig) vom gesamten "Range".
- Die 2. – ebenfalls konvexe – Großbewegung liegt auf dem Subjektsausdruck *matka*, sie fällt jedoch deutlich kleiner aus als die erste.
- Am Ende der 2. Großbewegung erreichen die Sprecher in der Regel ihr noch meßbares F_0 -Minimum; danach verläuft die F_0 -Kurve relativ linear in der Nähe der unteren Deklinationslinie weiter; auf der letzten Silbe kommt es üblicherweise zu Laryngalisierungserscheinungen.
- Die durchschnittliche Dauer der Äußerung beträgt 1580 msec, wovon 23-29% auf die Hauptbewegung fallen.

2.1.2. Intonation eines Versicherungsfragesatzes

Der Versicherungsfragesatz (vgl. (7)) ist ein Mischtyp aus einem Aussagesatz und einem *m/k*-Fragesatz. Seine Verwendung hängt meist mit Nicht-Glauben oder Nicht-Verstehen zusammen; hier wird eine Reaktion des Interaktionspartners noch dringender als beim Gebrauch eines Ergänzungsfragesatzes erwartet.

Die Intonation eines Versicherungsfragesatzes weist Gemeinsamkeiten mit der Intonation eines Ergänzungsfragesatzes auf:

- Die konvexen Großbewegungen finden analog zu denen in den Ergänzungsfragesätzen statt.
- Die Hauptakzentsilbe liegt hier ebenfalls auf *kuin*.
- Das Amplitudenmaximum liegt innerhalb der ersten und zugleich größeren Bewegung, i. e. auf dem Frageausdruck.
- Die dem Subjektsausdruck folgende F_0 -Kurve läuft ähnlich wie in den Ergänzungsfragesätzen an der unteren Deklinationslinie.

Die auffallenden Unterschiede zur Intonation in dem Grundtyp sind hier unabhängig vom Sprecher:

- Das auf *kuin* liegende F_0 -Maximum fällt um 50-85 Hz höher aus, und weil das vom jeweiligen Sprecher erreichte, noch meßbare Minimum relativ konstant bleibt (für die weiblichen Versuchspersonen bei ca. 120 Hz und für die männlichen Versuchspersonen bei ca. 80 Hz), bedeutet dies eine entsprechende Erhöhung des gesamten F_0 -Umfanges der Äußerung.
- Der von der 1. Großbewegung benötigte "Range" beansprucht 80-90% des gesamten F_0 -Umfanges.
- Die Gesamtdauer der Äußerung verlängert sich um ca. 150 msec auf ca. 1730 msec, wovon 29-35% auf *kuinka kauan* fällt - somit nehmen beide Dauervarianten um ca. 10% zu.
- Die Amplitude im Bereich der größten F_0 -Bewegung nimmt ebenfalls zu.

2.1.3. Intonation eines rhetorischen *m/k*-Fragesatzes

Der rhetorische *m/k*-Fragesatz zeichnet sich auf der nicht-intonatorischen formalen Seite vom Ergänzungsfragesatz dadurch ab, daß er obligatorisch eine Modalpartikel enthalten muß, und zwar *muka* oder *nyt* (hier in der Bedeutung:) 'schon' oder deren Kombinationen mit anderen (enklitischen) Modalpartikeln bzw. untereinander. Mit einem rhetorischen Fragesatz wird eine bestimmte 'Füllung' der *m/k*-Leerstelle präferiert, aber keine Antwortobligation etabliert:

- (8) 'Situation: Zwei Sprecher unterhalten sich:
 Sprecher 1: "Wird das aber wieder eine lange und anstrengende Reise werden!"
 Sprecher 2: "Übertreibe doch nicht!
Kuinka kauan matka nyt muka voi kestää?
 'wie lange Reise denn schon könn-dauer-'
 Wie lange kann die Reise denn schon dauern?
 Länger als zwei Tage werden wir dort nicht
 bleiben." '

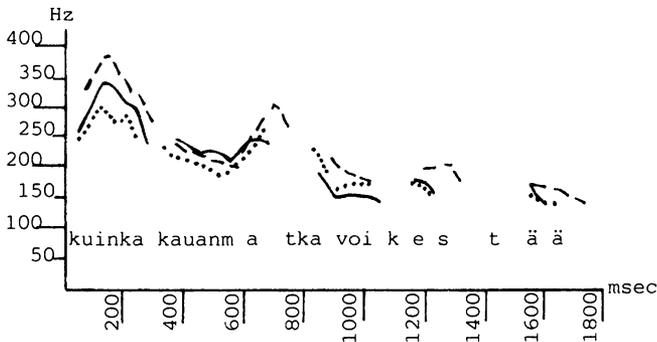
Auch bei diesem Fragesatz-Typ finden die - wiederum konvexen - Großbewegungen des Grundfrequenzverlaufs auf dem *m/k*- und dem Subjektsausdruck statt, wobei die erstere größer als die letztere ausfällt.

Die hauptsächlichsten intonatorischen Unterschiede zum Grundtyp Ergänzungsfragesatz sind:

- Die Hauptakzentsilbe verlagert sich vom Satzanfang gegen Ende der Äußerung hin. Sie wurde von den VPn oft auf *mu*- wahrgenommen, so daß es sich bei *muka* um eine betonbare Modalpartikel zu handeln scheint - eine auch für das Finnische seltene Erscheinung. Ferner wird bei diesem Satzmodus ersichtlich, daß die Hauptakzentsilbe nicht im Bereich der größten F_0 -Bewegung liegen muß, auch wenn dies meist der Fall ist.
- Der Onset liegt für jeden Sprecher hier um ca. 20-50 Hz niedriger als beim Ergänzungsfragesatz.
- Auch wenn die größte F_0 -Bewegung auf dem *m/k*-Ausdruck stattfindet, so fällt sie - abhängig vom Sprecher - um ca. 20-50% kleiner aus als beim Ergänzungsfragesatz.
- Bei ungefähr gleich bleibender unterer Deklinationslinie liegt das 1. F_0 -Maximum niedriger, so daß der Gesamt-Range im Durchschnitt um ca. 20% abnimmt.
- Die satzinitiale F_0 -Bewegung nimmt hier ca. 50-70% des gesamten F_0 -Umfangs in Anspruch.
- Die Dauer sowohl der Gesamtäußerung (die Modalpartikeln wurden zu diesem Zweck "herausgeschnitten") als auch des *m/k*-Ausdrucks verlängert sich sprecherabhängig. Es ist jedoch eine einheitliche Tendenz zu beobachten: je niedriger das satzinitiale F_0 -Maximum liegt, desto länger ist die Dauer des *m/k*-Ausdrucks.
- Der höchste Amplitudengipfel liegt regelmäßig weder auf dem Frageausdruck noch auf der Hauptakzentsilbe. Es ist auch in Bezug auf dessen Höhe keine einheitliche Tendenz im Verhältnis zum Ergänzungsfragesatz feststellbar.

Die folgende Abbildung zeigt die bereinigten F_0 -Kurven einer Sprecherin jeweils zum Ergänzungsfragesatz, Versicherungsfragesatz und zum rhetorischen *m/k*-Fragesatz (ohne Modalpartikel). Es sind die stimmhaften Segmente der Äußerungen aufgezeichnet, wobei die senkrechte Achse der Grundfrequenz in Hz und die waagerechte der Dauer in msec. entspricht:

Abbildung 1



Kuinka kauan matka (nyt muka) voi kestää
 'wie lange Reise schon könn-dauer-'

- = Ergänzungsfragesatz
- = Versicherungsfragesatz
- ... = rhetorischer *m/k*-Fragesatz

2.2. Intonation nicht-erotetischer *m/k*-Ausdrücke

Um feststellen zu können, ob sich mit den obigen Fragesätzen segmental identische Ausdrücke von diesen intonatorisch unterscheiden, wenn sie mit einer anderen als mit der Sprechereinstellung 'wissen wollen' verwendet werden, wurden diese Ausdrücke kontextuell so eingebettet, daß sie nicht als Fragen interpretiert werden konnten. Hier werden die intonatorischen Eigenschaften von *m/k*-Exclamativsätzen im Verhältnis zu denen der Fragesätze besprochen.

2.2.1. Intonation der *m/k*-Exclamativsätze

Mit der Grundeinstellung 'sich wundern, *m/k*' werden die *m/k*-Exclamativsätze (vgl. (6)) verwendet, in denen in der gesprochenen Umgangssprache häufig Interjektionen sowie enklitische und/oder nicht-enklitische Modalpartikeln vorkommen.

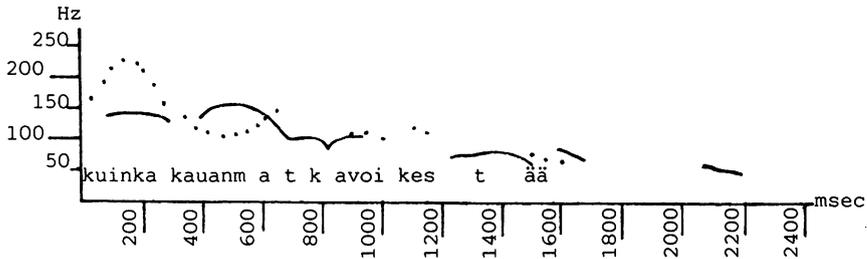
Die wichtigsten intonatorischen Unterschiede zu den oben besprochenen Satzmodi sind:

- Die Hauptakzentsilbe verlagert sich auf den wertenden Ausdruck *kauan*, auf der auch der größte F_0 -Gipfel und das Amplituden-Maximum liegen.
- Der Onset des F_0 -Verlaufs liegt niedriger als bei allen Fragesatz-Typen.
- Auf dem *m/k*-Ausdruck liegt die größte F_0 -Bewegung, die eine konvexe Form hat. Ihr Range fällt jedoch durchgehend kleiner aus als bei den Fragesätzen (Einzelbeispiel: Excl.: 50 Hz, rhet. Fr.: 120 Hz, Erg.fr.: 140 Hz, Vers.fr.: 220 Hz), was sich auch in dem insgesamt verkleinerten Gesamt-Range niederschlägt.

- Es verlängert sich sowohl die relative Dauer des m/k -Ausdrucks als auch die absolute Dauer der Gesamt-Äußerung im Verhältnis zu allen Fragesätzen.

Abbildung 2 verdeutlicht die Verhältnisse zwischen den F_0 -Verläufen eines Ergänzungsfragesatzes und eines m/k -Exclamativsatzes, gesprochen von einer männlichen Versuchsperson:

Abbildung 2



Kuinka kauan matka voi kestää
'wie lange Reise könn- dauer-'

... = Ergänzungsfragesatz
— = m/k -Exclamativsatz

2.3. Intonation in finnischen Satzfragesätzen

Der Nachweis für ähnliche Funktionen der Intonation wie im Bereich der m/k -Ausdrücke soll für die Satzfragesätze exemplarisch an dem Grundtyp 'Entscheidungsfragesatz', dem Sub-Typ 'deliberativer Satz-Fragesatz' und an dem nicht-erotetischen Satzmodus 'Satz-Exclamativsatz' geliefert werden. Der zur Demonstration herangezogene Satz lautet:

Omenistako(han/s) Matti (vasta) piti
'Apfel ? wohl/vielleicht Matti erst mög-'

2.3.1. Intonation eines Entscheidungsfragesatzes

Durch die Verwendung eines Entscheidungsfragesatzes kann der Sprecher die Grundeinstellung 'wissen wollen, ob' ausdrücken. Um ins Korpus möglichst viele funktionale Oppositionen einzugliedern, wurde für diesen Satzmodus als Beispielsatz u. a. ein Satz mit einem initialen fokussierten Objektsausdruck gewählt:

- (9) 'Situation: Der Sprecher zum anderen:
 Sprecher: "Omenistako Matti piti?"
 'Apfel ? Matti mög-'
 Waren es Äpfel, die der Matti so gern
 mochte?" '

Dieser Satztyp, mit dem eine Antwortverpflichtung verbunden ist, weist als hauptsächlichliche intonatorische Kennzeichen auf:

- Der Hauptakzent liegt (vermutlich je nach dem dialektalen Hintergrund des Sprechers) ausnahmslos entweder auf der 1. oder 2. Silbe.
- Der F_0 -Onset ist mittelhoch im Vergleich zu den anderen Oppositionen desselben Sprechers im Minimalpaar.
- Der F_0 -Gipfel liegt auf der 2. Silbe *me*-. Danach findet ein F_0 -Abfall bis zum Ende des *-kO*-Ausdrucks statt. Abgesehen von einer kleineren konvexen Bewegung auf dem Subjektsausdruck *Matti*, läuft die F_0 -Kurve danach bis zum Satzende relativ eben weiter.
- Die satzinitiale konvexe Bewegung auf *omenistako* ist die größte innerhalb des Gesamt-Satzes, und sie nimmt teilweise fast den Gesamt-Range in Anspruch, nämlich zwischen 86-97%.
- Die Gesamt-Dauer beträgt im Durchschnitt 1350 msec. Davon fallen 39-47% auf den *-kO*-Ausdruck.
- Der Amplitudengipfel fällt ausnahmslos auf das Segment *omenis*.

2.3.2. Intonation eines deliberativen Satz-Fragesatzes

Der deliberative Satz-Fragesatz, der auch als Selbst-Frage fungieren kann und nicht mit einer Antwortverpflichtung für den Hörer verbunden ist, dient zum Ausdrücken der Sprechereinstellung '(sich) fragen, ob (wohl)', und sein wichtigstes nicht-intonatorisches Kennzeichen im Unterschied zum Entscheidungsfragesatz ist die Modalpartikel *-hAn* 'wohl':

- (10) 'Situation: Der Sprecher denkt laut auf dem Weg zum
 Einkaufen:
 Sprecher: "Omenistakohan Matti piti?"
 'Apfel ? wohl Matti mög-'
 "Ob es wohl Äpfel waren, die
 der Matti so gern mochte?" '

Wie der Tabelle am Ende der Arbeit entnommen werden kann, teilt dieser Satztyp die meisten intonatorischen Eigenschaften mit dem Entscheidungsfragesatz. Die wesentlichen Unterschiede sind, abgesehen vom etwas kleineren Gesamt-Range, im Dauer-Bereich zu finden:

Bei einer um ca. 150 msec längeren Gesamt-Dauer des Fragesatzes bleibt der maximale Anteil des $-kO$ -Ausdrucks zwar gleich, insgesamt erhöhen sich aber also sowohl die Dauer der initialen Hauptbewegung auf *omenistako* (die Modalpartikel wurde zu diesem Zweck "herausgeschnitten") als auch die der darauf folgenden - im großen und ganzen - eben verlaufenden Segmente.

2.3.3. Intonation eines Satz-Exclamativsatzes

Ähnlich wie bei den *m/k*-Sätzen soll hier die Intonation eines nicht-erotetischen Satztyps verglichen werden mit der Intonation der oben besprochenen Fragesatz-Typen. Als Beispiel wurde (11) konstruiert:

- (11) 'Situation: Der Sprecher zum anderen:
 Sprecher: "Matti war ein Gourmand: Er aß gern
 allerlei Gemüse und Obst, aber
omenistakos Matti vasta piti!"
 'Apfel ? vielleicht Matti erst mög-'
 'aber Äpfel hat der vielleicht erst
 recht gern gemocht!'" '

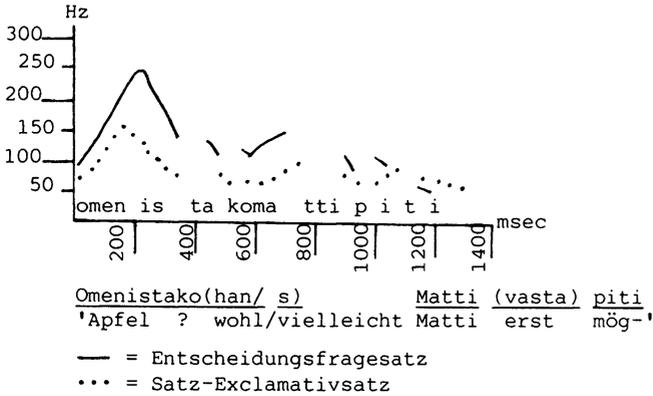
Diese Exclamativ-Variante teilt mit den Frage-Varianten die folgenden intonatorischen Eigenschaften: Die satzinitiale Lage der Hauptakzentsilbe, des Amplitudengipfels und des F_0 -Gipfels; die satzinitiale Lage und die konvexe Form der größten F_0 -Bewegung.

Die wichtigsten Unterschiede sind:

- Der F_0 -Onset liegt tiefer.
- Die größte Grundfrequenz-Bewegung fällt kleiner aus, aber umfaßt den gesamten Range-Bereich des Satzes. Daraus folgt, daß auch der Gesamt-Range kleiner als bei den Fragesätzen ist.
- Die Gesamt-Dauer des Satzes (ohne Modalpartikeln) ist länger als beim Entscheidungsfragesatz, aber etwas kürzer als beim deliberativen Fragesatz.

In Abbildung 3 werden die bereinigten Grundfrequenz-Verläufe eines Entscheidungsfragesatzes und eines Satz-Exclamativsatzes verglichen. Es handelt sich um Aufnahmen eines Sprechers:

Abbildung 3



3. Zusammenfassung

Es konnten bei der obigen Analyse satzmodus kennzeichnende Eigenschaften der untersuchten akustischen Parameter festgestellt werden. Die finnischen Satzmodi können hier in zwei Gruppen geteilt werden: Die eine Gruppe ist mit einer Antwortverpflichtung des Hörers verbunden, und bei der anderen Gruppe fehlt dieses Merkmal.

Von den intonatorischen Parametern sind es vor allem die Grundfrequenz und die Dauer, die es ermöglichen, zwischen zwei oder mehreren Fragetypen bzw. einem Nicht-Fragetyp zu unterscheiden. Dabei scheinen F_0 - und Dauer-Veränderungen sehr eng zusammenzuhängen und sich auch teilweise kompensieren zu können. Wie den Spalten "Gesamt-Dauer" und "-Range" der folgenden Tabelle entnommen werden kann, wird das erstere Merkmal meist länger, wenn das letztere kleiner wird. Mit der fehlenden Antwortverpflichtung korrelieren auch die kleineren F_0 -Bewegungen im Bereich des satzinitialen Ausdrucks, der in den vorliegenden Beispielen ein m/k - oder ein $-kO$ -Ausdruck war. Die Größe dieser F_0 -Bewegungen hat meist eine direkte Auswirkung auf den Gesamt- F_0 -Umfang des Satzes, so daß er bei den Nicht-Fragen im allgemeinen kleiner ist. Auch der Onset derjenigen Satzmodi scheint tiefer zu liegen, mit denen keine Reaktion des Hörers erwartet wird.

Insgesamt sind die Frage-Sätze also gekennzeichnet durch kürzere Dauer, höheren Onset, größere initiale F_0 -Bewegung und größeren Gesamt-Range.

Wie klein die Unterschiede in der intonatorischen Gestaltung sein dürfen, damit die finnischen Hörer sie noch wahrnehmen, muß in genaueren phonetischen Tests untersucht werden. Aufgrund dieser vornehmlich linguistisch gestalteten Intonationstests kann der Intonation im Finnischen jedoch zumindest im Satz-

modus-Bereich distinktive Funktion zugestanden werden. Wie die Intonation mit den anderen satzmodus-relevanten, formalen Kennzeichen, wie z. B. mit den zahlreichen - die Realisierung der beabsichtigten Lesart oft erleichternden - Modalpartikeln, im einzelnen korreliert wird, muß weiteren Untersuchungen vorbehalten bleiben.

Schlußwort

Die hier vorgestellte Untersuchung verdankt ihre Grundlage und Anregungen den entsprechenden Untersuchungen zum Deutschen am Münchner DFG-Projekt "Modus-Fokus-Intonation". Prof. Hans Altmann, dem Leiter des Projekts, danke ich für die vielen Ratschläge und Anregungen auch zum finnischen Satzmodus-System. Anton Batliner und Wilhelm Oppenrieder danke ich für ihre Beratung bei den verschiedenen phonetischen bzw. linguistischen Fragestellungen. Gedankt sei auch Prof. Kalevi Wiik für die großzügigen Arbeitsmöglichkeiten am phonetischen Institut in Turku sowie Markus Mattila für die technische Unterstützung bei den Aufnahmen und den Tests. Den Sprechern und den Versuchspersonen, ohne deren Mithilfe diese Untersuchung nicht möglich gewesen wäre, bin ich dankbar für ihre Geduld und ihr Interesse bei den verschiedenen Aufgaben.

ANMERKUNGEN

- 1 Alle Übersetzungen aus dem Finnischen stammen von der Verfasserin und sind mit einfachen Anführungszeichen versehen.
- 2 Die Sprecher hatten beliebig lange Zeit, sich in die Situation "einzuleben" und den Text ohne Aufnahme zu üben.
- 3 Am Anfang und am Ende der eigentlichen Aufnahme bekamen die Sprecher zusätzlich jeweils zwei Karten zum "An-" bzw. "Abgewöhnen" vorgelegt. Diese Aufnahmen wurden nicht weiter bearbeitet.
- 4 Zu einem Teil der Sätze wurde ein weiterer Test mit deutschen Versuchspersonen durchgeführt, die über keinerlei Finnisch-Kenntnisse verfügten. Auch sie hatten die Aufgabe, die Hauptakzentsilbe zu kennzeichnen. Auf die Ergebnisse dieses Tests kann hier jedoch nicht eingegangen werden.

Nr. Satztyp	Antwortobligat.	Lage Hauptakzent-silbe	Fo-Onset	Fo-Gipfel	g r ö ß t e F o - B e w e g u n g					Ges.-Dauer in msec	Gesamt-Range	Lage Amplituden-gipfel
					Lage	Form	Größe	Range in% vom Ges.-Range	Dauer in% von der Ges.-Dauer			
1. Ergänzungs- frages.	+	+satz- initial	mittel- hoch	+satz- initial	m/k- Aus- druck	konvex	mittel- groß	60-80%	23-29%	1580	mittel- groß	m/k- Aus- druck
2. Versiche- rungs- frage- satz	++	+satz- initial	sprecher- abhängig	+satz- initial höher als 1)	m/k- Aus- druck	konvex	größer als 1)	80-90%	29-35%	1730	größer als 1)	m/k- Aus- druck
3. rheto- rischer m/k- Frages.	-	-satz- initial	tiefer als 1)	+satz- initial kleiner als 1)	m/k- Aus- druck	konvex	kleiner als 1)	50-70%	30%	1720	kleiner als 1)	m/k- Aus- druck
4. m/k- Excla- mativs.	-	-satz- initial	tiefer als 1),3)	-satz- initial	m/k- Aus- druck	konvex	kleiner als 1)-3)	67-86%	36%	2170	kleiner als 1)-3)	m/k- Aus- druck
5. Entschei- dungs- frages.	+	+satz- initial	mittel- hoch	+satz- initial	-kO- Aus- druck	konvex	mittel- groß	86-97%	39-47%	1350	mittel- groß	-kO- Aus- druck
6. delibe- rativer Satz- Frages.	-	+satz- initial	tiefer als 5)	+satz- initial	-kO- Aus- druck	konvex	mittel- groß	83-100%	45-47%	1500	kleiner als 5)	-kO- Aus- druck
7. Satz- Excla- mativs.	-	+satz- initial	tiefer als 5),6)	+satz- initial	-kO- Aus- druck	konvex	kleiner als 5),6)	100%	46-49%	1450	kleiner als 5),6)	-kO- Aus- druck

Tabelle zu intonatorischen Merkmalen verschiedener Satzmodi

LITERATUR

- Altmann, H. 1984. Linguistische Aspekte der Intonation am Beispiel Satzmodus. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 19. 132-152
- Altmann, H. 1987. Zur Problematik der Konstitution von Satzmodi als Formtypen. in: Meibauer (Hg.) 1987. 22-56
- Bierwisch, M. 1979. Satztyp und kognitive Einstellung. Slovo a slovesnost XL. 194-199
- Freihoff, R. 1974. Kysymyslause ja intonaatio. Tampereen yliopiston puheopin laitoksen monisteita 1/75. Tampere. 69-71
- Hakulinen, A./Karlsson, F. 1979. Nykysuomen lauseoppia. Jyväskylää.
- Hirvonen, P. 1970. Finnish and English Communicative Intonation. Publications of the Phonetic Department of the University of Turku 8. Turku.
- Iivonen, A. 1978. Is There Interrogative Intonation in Finnish? Gårding, E. /Bruce, G./Bannert, R. (Hgg.) 1978. Nordic Prosody. Papers from a symposium. Lund. 43-53
- Iivonen, A. 1981. Suomen kysymyslauseiden prosodiikan määräytyvyydestä. Ikola, O. (ed.) 1981. Congressus Quintus Internationalis Fenno-Ugristarum, Turku 1980, Pars VI. Turku. 101-107
- Iivonen, A. 1983. On Explaining the Sentence Initial Pitch Height in Finnish. Mimeographed Series of the Department of Phonetics, University of Helsinki. Helsinki.
- Kallioinen, V. 1968. Suomen kysymyslauseen intonaatiosta. Virittäjä 72. 35-54
- Lang, E. 1981. Was heißt "eine Einstellung ausdrücken"? Rosengren, I. (Hg.) 1981. Sprache und Pragmatik. Lunder Symposium 1980. Lunder germanistische Forschungen 50. Malmö. 293-314
- Lang, E. 1983. Einstellungsausdrücke und ausgedrückte Einstellungen. Ruzicka, R./Motsch, W. (Hgg.) 1983. Untersuchungen zur Semantik. Studia grammatica XXII. Berlin. 305-341
- Luukko, L. 1985. Formen von Fragen und Funktionen von Fragesätzen. Eine deutsch-finnische kontrastive Studie unter besonderer Berücksichtigung der Intonation. Phil. Diss. München. (Erscheint voraussichtlich: Tübingen 1988.)
- Luukko-Vinchenzo, L. 1987. Entscheidungsfragesätze im Finnischen. Mit einem Exkurs ins Deutsche. in: Meibauer (Hg.) 1987. 125-138
- Matihaldi, H.-L. 1979. Nykysuomen modukset I. Kvalitatiivinen analyysi. (Acta Universitatis Ouluensis. Series B Humaniora No. 7. Philologica No. 2.) Oulu.
- Meibauer, J. (Hg.) 1987. Satzmodus zwischen Grammatik und Pragmatik. Referate anlässlich der 8. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Heidelberg 1986. Tübingen.
- Netter, K. 1982. Exklamativsätze mit Anfangsstellung des finiten Verbums im Deutschen. (Versuch einer formalen und funktionalen Analyse). M.A.-Arbeit. München.
- Östman, J.-O. 1977. On English *also*, Finnish *-kin*, and related matters. Palmberg, R./Ringbom, H. (Hgg.) 1977. Meddelanden för Åbo Akademi forskningsinstitut. Åbo Akademi Nr. 199. Turku. 165-196

- Oppenrieder, W. 1987. Aussagesätze im Deutschen. in: Meibauer (Hg.) 1987. 161-189
- Scholz, U. 1987. Wunschsätze im Deutschen - formale und funktionale Beschreibung. in: Meibauer (Hg.) 1987. 234-258
- Tarvainen, K. 1985. Kontrastive Syntax Deutsch-Finnisch. Heidelberg.
- Winkler, E. 1986. Der Satzmodus "Imperativsatz" im Deutschen und Finnischen: Form, Funktion und Aspekte der Verwendung. Satztypen mit Verberst- oder Verbzweitstellung. Phil. Diss. München.

INTONATION UND IDENTIFIKATION. KATEGORISIERUNGSTESTS ZUR KONTEXTFREIEN IDENTIFIKATION VON SATZMODI*

Wilhelm Oppenrieder (München)

0. Überblick

Die intonatorische Form von Äußerungen trägt in systematischer Weise zur Identifizierbarkeit bestimmter Bedeutungsaspekte bei. So gehört sie im Deutschen zu den formalen Mitteln, mit deren Hilfe der Satzmodus einer Äußerung gekennzeichnet wird. Die Rolle der Intonation bei der Satzmodusanzeige läßt sich insbesondere an Sätzen untersuchen, in denen die anderen Merkmale neutralisiert sind. Derartige Sätze lagen den Hörtests zugrunde, über die im folgenden Beitrag berichtet wird. Dabei mußten die Versuchspersonen kontextlos vorgespielte Äußerungen bestimmten Satzmodi, bzw. Funktionstypen, zuordnen. Obwohl die Äußerungen bis auf die intonatorische Form (und teilweise charakteristische Modalpartikeln) 'satzmodusambig' waren, konnte in den meisten Fällen der Funktionstyp sicher identifiziert werden.

1. Grundlagen und Testaufbau

Die in den Tests verwendeten Äußerungen basieren auf einer Anzahl von Sätzen, die auf dem Hintergrund des in Altmann (1984) und (1987) entwickelten Satzmodussystems konstruiert wurden und sogenannte 'intonatorische Minimalpaare' bilden. Satzmodi sind danach aufzufassen als die Zuordnung eines abstrakten Funktionstyps zu einem Formtyp, der durch bestimmte kategoriale, topologische, morphologische und intonatorische Merkmale, vgl. das Schema in Altmann (1987: 32), charakterisiert ist. Die Relevanz der intonatorischen Markierung zeigt sich insbesondere dann, wenn bei einer

* Für Unterstützung sei hiermit den anderen Mitgliedern des DFG-Projekts Modus-Fokus-Intonation gedankt.

Neutralisierung der nicht-intonatorischen Merkmale dennoch einer Struktur zwei verschiedene Funktionstypen zugeordnet sind. Da die Intonation in solchen intonatorischen Minimalpaaren funktional besonders hoch belastet ist, sollten sich die für die Satzmodus-kennzeichnung relevanten intonatorischen Parameter auch besonders gut untersuchen lassen.

Wir konstruierten daher eine Anzahl von Testsätzen, die intonatorische Minimalpaare bilden. Diese Sätze wurden in Kontexte eingebettet, die jeweils nur eine Satzmodus-Interpretation zulassen. Die Kontext-Testsatz-Kombinationen wurden 6 Versuchspersonen (3 weiblichen, 3 männlichen) zur Realisation vorgelegt (zum Verfahren vgl. Oppenrieder (1988)). Auf diese Weise erhielten wir 956 Testsatz-Realisationen.

In den hier vorgestellten **Kategorisierungstests** ging es uns darum zu untersuchen, inwieweit tatsächlich allein die intonatorische Form einer Äußerung die Identifizierung des Satzmodus gewährleisten kann. Dazu mußten die Minimalpaar-Äußerungen natürlich aus dem jeweiligen satzmoduslenkenden Kontext herausgeschnitten werden.

Die Versuchspersonen (VPen) in den Kategorisierungstests, Student/inn/en der Linguistik oder Phonetik, die mit dem zugrundeliegenden Satzmodussystem nicht vertraut waren, sollten entscheiden, zu welcher **funktionalen Kategorie** eine Realisation jeweils gehört: Sie sollten die vorgespielten Äußerungen also nach funktionalen Gesichtspunkten klassifizieren, während die zugrundeliegenden Testsätze so konstruiert waren, daß ihnen jeweils ein ganz bestimmter Komplex formaler Merkmale zukam - bzw. zukommen sollte, da die Realisation der intendierten intonatorischen Form natürlich davon abhing, inwieweit die VPen der Produktionstests den lenkenden Kontext richtig umsetzten.

Die Frage nach dem **Formtyp**, der uns ja vor allem interessierte, schloß sich schon wegen der fehlenden Kenntnis des Satzmodussystems bei den VPen aus. Bei derartig spezifischen Merkmalskombinationen, wie sie die Satzmodi kennzeichnen, hätten nur linguistisch speziell vorgebildete VPen an den Kategorisierungstests teilnehmen können; in diesem Fall würde dann aber zunächst einmal nur festgestellt, wie gut die VPen die lingu-

stischen Vorgaben auf konkrete Äußerungen anwenden können. Außerdem war nicht von vornherein klar, welche intonatorischen Merkmale tatsächlich einen bestimmten Satzmodus charakterisieren - vielmehr sollte dieses ja erst herausgefunden werden, unter anderem durch derartige Kategorisierungstests.

Der Ausweg, die VPen die interessierenden Formtypen ohne explizite linguistische Erläuterungen allein über die Einführung positiver und negativer Beispiele unterscheiden lernen zu lassen (vgl. Jaeger (1986)), fällt mit großer Sicherheit weg, da die entsprechenden Kategorien viel zu komplex sind.

Globale funktionale Kategorien sind dem 'naiven' Sprachbenutzer dagegen sehr viel leichter zugänglich; die Feststellung des Funktionstyps einer Äußerung, wenn auch nicht die explizite Etikettierung mit einer Funktionstypbezeichnung, ist schließlich eine Aufgabe, die jeder kompetente Hörer lösen können muß.

Allerdings sagen die Ergebnisse derartig konstruierter Tests dann auch zunächst nur etwas über die (für die Hörer wahrscheinlichste) Funktion der Testsatzrealisationen aus; daß die von uns angesetzte Menge der formalen Merkmale, insbesondere die spezifischen intonatorischen Eigenschaften der Äußerungen, die entsprechende Reaktion der VPen gelenkt hat, kann daraus nur erschlossen werden. Bei derartigen Schlüssen muß berücksichtigt werden, daß andere Faktoren, etwa eine bestimmte lexikalische Füllung, die funktionalen Urteile deutlich beeinflussen können, so daß es in manchen Fällen zu einer Uminterpretation des Funktionstyps, zu einer indirekten Interpretation, kommt. Neben der formtypabhängigen direkten funktionalen Interpretation kann also eine wahrscheinlichere indirekte Interpretation existieren, die jene überdeckt: *Kannst du mich am Buckel kratzen?* wird eher als Aufforderung, denn als Frage interpretiert werden. Der Rückschluß auf einen zugeordneten Formtyp ist in einem solchen Fall sehr viel problematischer.

Vorgegeben waren die fünf Kategorien **Aufforderung**, **Frage**, **Aussage**, **Ausruf/Exklamativ** und **Wunsch**, die durch eine kurze Beschreibung der jeweiligen Funktion (eine Beschreibung, die in Anlehnung an die im Satzmodussystem angesetzten propositionalen Grundeinstellungen, vgl. Altmann (1987), formuliert war) und

einige charakteristische Beispiele eingeführt wurden; wir setzten dabei voraus, daß bei den 'Mischtypen' des Satzmodussystems (Altmann 1987: 48f.), z.B. den assertiven Fragen (*Du kommst?*), eine Funktionstypinterpretation die dominante ist, bei den assertiven Fragen z.B. die als Frage (und nicht die als Aussage).

Den VPen wurden die Äußerungen jeweils einmal vorgespielt (dafür wurde die (Pseudo-)Randomisierung aus den Produktionstests beibehalten); sie hatten dann 4 Sekunden Zeit, sich für eine der Kategorien zu entscheiden; dazu mußten sie an einem Abstimmkästchen einen von 5 Knöpfen drücken. An den einzelnen Testsitzungen nahmen im Durchschnitt 12 VPen teil.

Die Kategorisierungstests waren Teil einer größeren Serie von Tests, die dazu dienten, Urteile von 'naiven' Hörern über bestimmte Aspekte der Testsatzrealisationen zu erhalten. Nach bestimmten Kriterien ausgewählte akustische Parameterwerte derjenigen Äußerungen, bei denen unter anderem die Ergebnisse der Kategorisierungstests dem intendierten Funktionstyp entsprachen, wurden zur Bildung sogenannter intonatorischer Prototypen herangezogen, dem intonatorischen Merkmalskomplex im Bündel der satzmodus kennzeichnenden formalen Merkmale (vgl. Oppenrieder (1988)).

2. Ergebnisüberblick

Die Testergebnisse, die auf den drei Blättern der Tabelle im Anhang zusammengefaßt sind, waren bei den meisten Testsätzen ziemlich konsistent und entsprachen den von uns angezielten Interpretationen. Sie zeigen also, daß tatsächlich die intonatorische Form der Realisationen für eine Einordnung in die fünf angegebenen Kategorien ausreichte. Abweichungen sind in jedem Fall erklärbar.

3. Erläuterungen zur Tabelle

Aufgeführt sind sämtliche untersuchten Satztypen, ab Nr. 16 werden einzelne Minimalpaare einander gegenübergestellt.

Unter der Satztypbezeichnung findet sich in Klammern jeweils ein Beispielsatz aus unserem Korpus, daneben steht die Anzahl der Realisationen des Typs.

Die fünf Spalten in der rechten Hälfte der Tabelle geben an, wieviel Prozent der Versuchspersonen im Durchschnitt ein Exemplar des jeweiligen Satztyps als Aufforderung (AUF), Frage (FRA), Aussage (AUS), Exklamativ (EXK) oder Wunsch (WUN) kategorisierten. Die Anzahl der Realisationen pro Satztyp schwankt ziemlich stark. Das liegt daran, daß manche Satztypen die Bildung mehrerer Minimalpaare zulassen (z.B. Nr. 1, der Verb-Erst-Fragesatz bzw. Entscheidungsfragesatz), während für andere Typen nur ein Minimalpaar gebildet werden konnte (z.B. für Nr. 3, die assertive Frage). Die mit durchgezogenen Strichen eingerahmten Werte verweisen auf die von uns angezielten Kategorisierungen, mit gestrichelten Linien eingerahmte Werte auf häufigere 'Fehl kategorisierungen', nämlich solche, die die 10%-Grenze überschreiten.

4. Einzelergebnisse

In den Zeilen 1 bis 4 der Tabelle sind die Ergebnisse für die Haupttypen der Fragesätze (diejenigen mit Verb-Erst- und Verb-Zweit-Stellung) aufgeführt, in Zeile 5 und 6 die für bestimmte Aufforderungssatztypen. Die Ergebnisse sind eindeutig. Das gilt von den Fragesatztypen mit Verb-Letzt-Stellung auch für die deliberativen *w*-Verb-Letzt-Fragesätze (Nr. 11) und die Wortfrage-Übernahmen (Nr. 12: dabei handelt es sich um Strukturen, mit denen der Sprecher einen unmittelbar vorhergehenden wortgleichen *w*-Fragesatz eines anderen Sprechers als 'Vorbereitung' auf eine eigene Antwort wiederaufnimmt).

Allerdings zeigen sich auch hier schon bestimmte typische 'Fehl kategorisierungen', bzw. von uns nicht erwartete Kategorisierungen, wenn auch noch in bescheidenem Ausmaß. Sie sind darauf zurückzuführen, daß die Kategorien nur funktional eingeführt wurden.

So erklären sich die 10.5% Wunschkategorisierungen für die Adhortativsätze in Zeile 6: Aufforderungen, die den Sprecher selbst mit einschließen, sind anscheinend auch 'schwächer', nämlich als Ausdruck eines bloßen Wunsches, interpretierbar (unter 'Ausdruck eines Wunsches' ist dann eine allgemeinere Einstellung zu verstehen als die propositionale Grundeinstellung, die im zugrundeliegenden Satzmodussystem dem speziellen Formtyp Wunschsatz zugeordnet wird).

Ebenfalls ein nicht vermeidbares Resultat der funktionalen Einführung der Kategorien sind Fehl kategorisierungen in solchen Fällen, wo eine indirekte Interpretation die wahrscheinlichere war. Die 5.1% Aufforderungs-Kategorisierungen beim Verb-Erst-Fragesatz in Zeile 5 stammen z.B. nahezu ausschließlich von dem in Zeile 15 der Tabelle angeführten Ausdruck *Gehst du nun bald nach Hause?*, der in den Produktionstests (aufgrund einer entsprechenden Kontextvorgabe) als eine 'unhöfliche Frage' realisiert werden sollte. Die Interpretation als indirekt ausgedrückte Aufforderung - bzw. wieder als etwas schwächerer Wunsch - an den Gegenüber liegt hier näher als die direkte Interpretation als Frage.

Ganz deutlich läßt sich das Durchschlagen einer indirekten Interpretation auch an einem der beiden Beispiele für den Aussagesatz erkennen, nämlich an *Du kommst.*, aufgeführt unter Nr. 14a in der Tabelle. Mit 60% überwiegt hier deutlich die Interpretation als Aufforderung gegenüber der Aussage-Interpretation mit 32%.

Generell war es für die VPen schwierig, die **Exklamativtypen** klar von den anderen Typen zu trennen. Daß gerade die funktionale Trennung von Aussage und Exklamativ problematisch ist, zeigen auch die Ergebnisse der Perzeptionstests mit gezielt manipulierten Stimuli (Batliner (1988a); vgl. auch die NMDS-Analyse zu einem anderen Minimalpaarkorpus in Batliner (1988b)).

Zum einen sind Exklamative in ihrer Funktion den hauptsächlich 'informierenden' Aussagen insofern ähnlich, als der Hörer natürlich Schlüsse über das Vorliegen eines Sachverhalts ziehen kann, auf den sich die Verwunderung des Sprechers bezieht.

In den Zeilen 7 bis 10 und in Zeile 16 und 18 der Tabelle sieht man relativ hohe Prozentwerte für die Aussage-Interpretation bei den verschiedenen Exklamativsatztypen.

Gerade bei Nr. 9, 10 und 18, wo in jedem Fall ein *w*-Ausdruck vorhanden ist, wird deutlich, daß die von uns nicht angezielte Kategorisierung als Aussage nicht über eine falsch interpretierte Form zustandekommen kann, sondern damit zusammenhängt, daß das Vorliegen eines bestimmten Sachverhalts, auf den sich die exklamativtypische Einstellung bezieht, erschlossen werden kann. Nr. 9

und 18 unterscheiden sich dabei durch die Stellung des wertenden Adjektivs, das in 18 mit im Vorfeld erscheint, in 9 zum Frageanzeiger wie in Distanzstellung steht (eine Erscheinung, die dem bekannten Quantoren-Floating ähnelt, daher 'F' für 'Floating').

Zum zweiten wurden öfters Nicht-Exklamativsatzrealisationen als Exklamative bzw. Ausrufe kategorisiert. Hier wirkt sich insbesondere aus, daß die Bezeichnungen 'Exklamativ' bzw. 'Ausruf' nicht als in ihrem Geltungsbereich genau umgrenzte technische Ausdrücke für eine abstrakte propositionale Grundeinstellung eingeführt wurden, sondern nur als verhältnismäßig vage Bezeichnungen für 'Expressivität'. Auch Äußerungen von Sätzen der nicht-exklamativischen Formtypen können natürlich das Merkmal der 'Expressivität' aufweisen, also z.B. mit besonderem 'Nachdruck' gesprochen werden, so daß eine Interpretation als eine Art Exklamativ, der eben auch schon von seiner Form her auf eine 'expressive' Funktion festgelegt ist, manchmal naheliegen kann. Als Beispiel für eine derartige von uns nicht angezielte Kategorisierung sei auf die 8.9% Exklamativ-Urteile bei der assertiven Frage *Du kommst?* in Zeile 3 verwiesen: Vor allem der starke Anstieg der Grundfrequenz in den Realisationen dieses Testsatzes ist für den Eindruck der 'Expressivität' verantwortlich. Diese Erklärung trifft wohl auch für die verhältnismäßig 'expressiven' *w-Verb-Letzt-Rückfragen* zu (Nr. 13: mit ihnen reagiert der Sprecher wie bei Wortfrage-Übernahmen auf wortgleiche Vorgängerfragen eines anderen Sprechers; Ziel ist jedoch die Bestätigung, daß diese Frage richtig verstanden wurde).

Bei den vier unter Nr. 7 bis 10 aufgeführten Exklamativsatz-Typen ist bemerkenswert, daß kaum Frage-Klassifizierungen vorkommen, obwohl doch die Verb-Erst-Stellung bzw. das Auftreten eines *w*-Ausdrucks die Interpretation als Frage scheinbar nahelegen. Hier ist die intonatorische Form der Realisierungen also ausreichend, um diese Fehlinterpretation abzublocken - wobei allerdings darauf hingewiesen werden muß, daß in etliche der Testsätze Modalpartikeln eingefügt waren, die sicherlich wesentlich zur Identifikation des gemeinten Typs beitragen. Die ('Fehl'-)Klassifizierungen als Aussagen sind in allen vier Fällen sehr zahlreich.

Bei Nr. 7, wo im Durchschnitt ein Viertel der Testpersonen die

jeweiligen Realisierungen als Aussagen einordnete, läßt sich aber auch argumentieren, daß diese sogenannten 'Satz-Exklamativsätze' mit anscheinend fester Verb-Erst-Stellung besser als spezielle Aussagesätze mit nicht realisiertem Vorfeld und starker 'expressiver' Komponente analysiert werden sollten: Unter Satz-Exklamativsätzen sind dabei Ausdrücke zu verstehen, bei denen sich die propositionale Einstellung auf das simple Bestehen eines Sachverhalts bezieht - Beispiel: *Kennt der (doch glatt) den Kaiser von China!* -; bei den Grad-Exklamativsätzen mit dem typischen 'Exklamativakzent' auf einem definiten Ausdruck bezieht sich die zugeordnete propositionale Einstellung dagegen auf den Grad der Ausprägung einer Eigenschaft oder die Größe einer Menge - Beispiel: *Kennt DER (vielleicht) viele Kaiser!*. Auf Nr. 8, den Grad-Verb-Erst-Exklamativsatz wird weiter unten noch etwas genauer eingegangen.

Die Einordnung der Grad-Verb-Zweit-Exklamativsätze und der Aussagesätze mit 'Kontrastakzent' fiel den Testpersonen offensichtlich nicht leicht, wie die Prozentzahlen unter den Nummern 16 und 17 der Tabelle zeigen. Der 'Kontrastakzent' in den Aussagesätzen war dabei aus Gründen der Minimalpaarbildung notwendig.

Zum einen wirkt sich bei den Klassifizierungsunsicherheiten die nicht ganz eindeutige funktionale Abgrenzung von Exklamativ und Aussage aus - zumal die Kontrastakzentuierung für eine erhöhte 'Expressivität' bei den entsprechenden Aussagesatz-Äußerungen sorgt.

Zum anderen wurden die Realisationen in den Kategorisierungstests ja kontextlos dargeboten, so daß die Realisierung einer kontextbedingten 'Kontrastvariante' eines Aussagesatzes weniger nahelag als die Realisierung eines Exklamativsatzes.

Schließlich ist aber vor allem von Bedeutung, daß schon im ursprünglichen System der Markierungsmittel für beide Satzmodi ein fallendes Tonmuster charakteristisch ist und als differenzierendes Merkmal lediglich ein Exklamativakzent angesetzt wurde. Anscheinend ist die Form dieses Akzenttyps nicht hinreichend verschieden von der eines ausgeprägten Fokusakzents, wie er bei Kontrastierung gewöhnlich auftritt, um die (kontextlos) wahrscheinlichere Interpretation als Exklamativ bei den Aussagesatz-Äußerungen abzublocken.

Allerdings sind auch bei den Exklamativsatz-Realisationen im Durchschnitt sehr viel mehr Aussage-Kategorisierungen zu registrieren als bei den oben besprochenen Exklamativsatz-Typen, die schon von der Verbstellung oder dem Vorhandensein eines *w*-Ausdrucks her formal den Aussagesätzen nicht gleichen. Obwohl die VPen nach funktionalen Kriterien urteilen, wirkt sich die starke formale Ähnlichkeit von Aussagesatz und Grad-Verb-Zweit-Exklamativsatz aus. Auffällig ist in diesem Zusammenhang, daß die Kategorisierungsurteile bei den Exklamativsatz-Realisationen ohne Modalpartikel *vielleicht* und bei den Aussagesatz-Realisationen in genau der gleichen Weise verteilt sind - etwas mehr als ein Drittel Aussage-Kategorisierungen, der Rest fast nur Exklamativ-Kategorisierungen.

Ähnlich häufige Fehl kategorisierungen gab es beim *w*-Verb-Zweit-Exklamativsatz - Nr. 18 der Tabelle. Dieser bildete zusammen mit einem *w*-Verb-Zweit-Fragesatz (Ergänzungsfragesatz), Nr. 19, ein Minimalpaar.

Die 20% Aussage-Kategorisierungen dürfen wohl (wieder) mit den Exklamativ-Kategorisierungen zu einer nicht ganz klar abgegrenzten 'exklamativischen Funktions-Familie' zusammengefaßt werden, so daß also nur gut zwei Drittel der Testpersonen die einzelnen Realisationen wie erwartet klassifizierten, während das restliche Drittel die Äußerungen als Fragen kategorisierte.

Hier ist zu beachten, daß *w*-Exklamativsätze, um die Verwechslung mit Ergänzungsfragesätzen zu verhindern, häufig Verb-Letztstellung aufweisen, unsere Testsätze also die wahrscheinlich sowieso schon seltenere Form exemplifizierten.

Bei den parallel gebauten *w*-Verb-Zweit-Fragesätzen (Ergänzungsfragesätzen) ist die 'Erkennensrate' höher, nämlich 82.5%. Die Kategorisierungsergebnisse waren hier charakteristischerweise vor allem dann konsistent und wie von uns angezielt, wenn die Fragesatz-Realisationen einen hohen F_0 -Offset (einen F_0 -Wert am Ende der Äußerung, der in der oberen Hälfte des in der Gesamtäußerung ausgeschöpften F_0 -Umfangs liegt) aufwiesen, also ein steigendes Tonmuster hatten, wobei das Tonmuster für *w*-Verb-Zweit-Fragesätze nicht von vornherein auf steigend festgelegt ist.

Bei einem tiefen F_0 -Offset (einem äußerungsfinalen F_0 -Wert, der in der unteren Hälfte des realisierten F_0 -Umfangs liegt, im prototypischen Fall in der Nähe des tiefsten für den Sprecher/die Sprecherin erreichbaren F_0 -wertes) und bei Verb-Zweit-Stellung ist eine Frageinterpretation allerdings immer noch wahrscheinlicher als eine Exklamativinterpretation. Das legen zumindest die Ergebnisse in Zeile 18 nahe: Die intonatorische Form - möglicherweise gibt es auch hier die spezielle Realisierungsform eines Exklamativakzents - kann die kontextlos wahrscheinlichere Interpretation als Frage nicht genügend abblocken. Andererseits gibt es auch bei den w -Verb-Zweit-Fragesätzen mit 17.5% noch relativ viele Fehlklassifikationen als Exklamativ bzw. Aussage.

Einige Grad-Verb-Erst-Exklamativsätze (Nr. 21) wurden so formuliert, daß sie in Wortfolge und Akzentverteilung identisch waren mit Entscheidungsfragesätzen sowie mit Imperativsätzen mit realisiertem Subjektspronomen der 2.Ps.Pl., wobei auf dieses Subjektspronomen ein 'Kontrastakzent' fiel (Nr. 20).

Wieder müßte vom System der Markierungsmittel her der Unterschied lediglich im Vorhandensein eines Exklamativakzents einerseits und eines bestimmten Fokusakzents beim Imperativsatz andererseits liegen.

Insgesamt wurden die Imperativsatz-Äußerungen nur zu zwei Dritteln wie erwartet klassifiziert. Bei den Realisationen ohne Modalpartikel waren es noch weniger, nämlich nur rund 60%. Der Rest entfällt auf Exklamativ-Kategorisierungen und die hier wieder exklamativverdächtigen Aussage-Kategorisierungen. Bei der Verwendung der Modalpartikel *doch* steigt der Anteil der Aufforderungs-Kategorisierungen auf drei Viertel.

Interessanterweise scheint diese Modalpartikel auch den Aufforderungscharakter der jeweiligen Äußerungen abmildern zu können. Das legen zumindest die 11.1% Wunsch-Kategorisierungen nahe. Die Modalpartikel *doch* ist im Formtyp Exklamativsatz inakzeptabel. Daß dennoch auch beim Vorhandensein dieser Partikel Exklamativ-Kategorisierungen vorkommen, dürfte wieder mit einer besonderen 'Expressivität' der Äußerung zusammenhängen, die die VPen zur Einordnung unter die Exklamative bzw. Ausrufe veranlaßte.

Auch der Exklamativsatz in diesem Minimalpaar ist, wie die

Zahlen unter Nr. 21 belegen, nicht eindeutig erkennbar. Selbst bei Verwendung des Ausdrucks *vielleicht*, der in einem Imperativsatz auf keinen Fall akzeptabel ist, gibt es noch einige Aufforderungs-Kategorisierungen.

Sehr anfällig für Exklamativ-Interpretationen sind die **Wunschsätze** (vgl. Nr. 22 der Tabelle). Faßt man bei 22bi Aussage- und Exklamativ-Kategorisierungen zusammen, so überwiegen die Exklamativ-Interpretationen sogar leicht. Allerdings muß dabei berücksichtigt werden, daß es sich in diesem Fall um einen sehr stark markierten Wunschsatz mit Kontrastakzent handelt, der zusammen mit einem **Grad-Verb-Erst-Exklamativsatz**, Nr. 23, ein Minimalpaar bildet. Andererseits wurden auch die Exklamativsatz-Realisationen als Wünsche fehlklassifiziert. Innerhalb des Minimalpaares sind die beiden Funktionstypen zumindest ohne Modalpartikelunterstützung nicht hinreichend unterscheidbar. Beim Testsatz 22a ohne das exklamativsatztypische definite Element - das andere Glied des Minimalpaares war hier ein Entscheidungsfragesatz - können dagegen deutlich weniger Exklamativ-Kategorisierungen festgestellt werden. Wünsche teilen natürlich mit den Exklamativen einige Merkmale, unter anderem das Moment der 'Expressivität'; die Fehleinschätzungen sind also erklärlich. Bei Verwendung einer Modalpartikel sinkt die Rate der Fehlklassifizierungen, aber gerade bei den sehr stark markierten Testsätzen (z.B. bei Verwendung des Konjunktiv II im Exklamativsatz unter Nr. 23b) kommt es bei der Einschätzung durch die VPen ziemlich häufig zu Fehl kategorisierungen, so wenn sehr viele der Äußerungen mit *vielleicht* noch als Wünsche eingeschätzt werden.

5. Schlußbemerkung

Bei den einzelnen Fragesatz- und Imperativsatztypen und teilweise auch bei den Aussagesätzen zeigen die Ergebnisse, daß die intonatorische Form von Äußerungen auch kontextlos eine ziemlich sichere Einordnung in ein Funktionssystem der Satzmodi zulassen kann. Probleme mit der funktionalen Zuordnung ergeben sich dagegen einerseits, wenn der Wortlaut einer Äußerung eine indirekte Interpretation nahelegt, und andererseits im Zusammenhang

mit den 'nicht-klassischen' Satzmodi Exklamativ(-Satz) und Wunsch(-Satz). Obwohl also 'Fehlklassifizierungen' vorkommen, sehen wir die Ergebnisse der Kategorisierungstests als aussagekräftige Stützung für die Ansicht an, daß intonatorische Merkmale wesentlich an der Kennzeichnung der Satzmodi beteiligt sind.

LITERATUR

- Altmann, H. 1984. Linguistische Aspekte der Intonation am Beispiel Satzmodus. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 19. 132-152.
- Altmann, H. 1987. Zur Problematik der Konstitution von Satzmodi als Formtypen. Meibauer, J. (Hg.) 1987. Satzmodus zwischen Grammatik und Pragmatik. Tübingen. 22-56.
- Batliner, A. 1988a. Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des F_0 -Gipfels. In diesem Band.
- Batliner, A. 1988b. Modus und Fokus als Dimensionen einer Nonmetrischen Multidimensionalen Skalierung. In diesem Band.
- Jaeger, J.J. 1986. Concept Formation as a Tool for Linguistic Research. Ohala, J.J./Jaeger, J.J. (eds.) 1986. Experimental Phonology. Orlando etc. 211-237.
- Oppenrieder, W. 1988. Intonatorische Kennzeichnung von Satzmodi. In diesem Band.

ANHANG: ERGEBNISSE DER KATEGORISIERUNGSTESTS

Nr	Satztyp (/Beispielsatz)	Zahl der Realisationen	AUF	FRA	AUS	EXK	WUN
			(in Prozent der Versuchspersonen)				
1	Entscheidungsfragesatz (<u>Hätte er (wohl) getroffen?</u>)	266	5.1	88.8	1.2	2.6	2.3
2	Ergänzungsfragesatz (<u>Wie ist der (denn) reich geworden?</u>)	69	0	91.6	3.2	5.2	0
3	assertive Frage (<u>Du kommst?</u>)	12	0.9	88.4	0	8.9	1.8
4	Versicherungsfragesatz (<u>Er sieht w a s?</u>)	24	0	95.1	0	4.9	0
5	Sie-Imperativsatz (<u>Schlafen Sie!</u>)	12	92.2	2.6	0.8	1.8	2.6
6	Adhortativsatz (<u>Gehen wir!</u>)	25	83.4	0	1.3	4.8	10.5
7	Satz-V-1-Exklamativsatz (<u>Rennt der (einfach) wieder!</u>)	71	0	2.5	22.4	75.1	0
8	Grad-V-1-Exklamativsatz (<u>Stellt ihr euch an!</u>)	126	3.4	6.1	19	63.3	8.2
9	W-V-2-F-Exklamativsatz (<u>Wie ist der (doch) reich geworden!</u>)	26	0.4	3.9	13.7	81.1	0.9
10	W-V-L-Exklamativsatz (<u>Wie der (erst) läuft!</u>)	44	0.2	3.3	15	81.5	0
11	deliberativer W-V-L-Fragesatz (<u>Wie der wohl läuft?</u>)	13	0	95.5	0.9	2.7	0.9

Nr	Satztyp (/Beispielsatz)	Zahl der Realisationen	AUF	FRA	AUS	EXK	WUN
			(in Prozent der Versuchspersonen)				
12	Wortfrage-Übernahme (Wie groß der ist?)	15	0.8	94.2	1.4	3.6	0
13	W-V-L-Rückfrage (Wie der läuft?)	12	0	68.4	7.8	22.8	1
14	Aussagesatz	24	29.1	3.4	63.2	3.4	0.9
14a	(Du kommst.)	12	60	4	32	2	2
14b	(Er sieht was.)	12	0	2.8	92.4	4.8	0
15	Entscheidungsfragesatz (aus Nr 1) (Gehst du nun bald nach Hause?)	16	48.8	31.4	0	2.3	17.9
16	Grad-V-2-Exklamativsatz (Der hat geflucht!/Die ist naiv!)	53	0	0.7	29.6	69.7	0
16a	(Realisationen ohne Modalpartikel)	28	0	0.8	38.1	61.1	0
16b	(mit Modalpartikel <u>vielleicht</u>)	25	0	0.4	20.9	78.7	0
17	Aussagesatz mit 'Kontrast' (D e r hat geflucht./D i e ist naiv.)	56	0.2	1.2	36.6	62	0
18	W-V-2-Exklamativsatz (Wie laut ist es hier!/Wie alt ist er geworden!)	25	0	31.3	20.3	48.4	0
19	Ergänzungsfragesatz (aus Nr 2) (Wie laut ist es hier?/Wie alt ist er geworden?)	31	0	82.5	7.2	10.3	0

Nr	Satztyp (/Beispielsatz)	Zahl der Realisationen	AUF	FRA	AUS	EXK	WUN
			(in Prozent der Versuchspersonen)				
20	Imperativsatz mit 'Kontrast' (Stellt i h r euch an!)	26	67.7	0.8	6.8	18.7	6.1
20a	(Realisationen ohne Modalpartikel)	13	60.3	0.8	12.3	25.7	0.9
20b	(mit Modalpartikel <u>doch</u>)	13	74.8	0.8	1.5	11.8	11.1
21	Grad-V-1-Exklamativsatz (aus Nr 8) (Stellt ihr euch an!)	26	16	3.7	16	62.3	2
21a	(Realisationen ohne Modalpartikel)	13	23.8	0	23	51.7	1.5
21b	(mit Modalpartikel <u>vielleicht</u>)	13	7.6	7.6	8.5	73.8	2.5
22	Wunschsatz	57	0	4	6.6	22.3	67.1
22ai	(<u>Hätte er getroffen!</u>)	15	0	9.3	7.8	18.5	64.4
22aii	(<u>Hätte er doch getroffen!</u>)	13	0	0	1	9.4	89.6
22bi	(<u>Wäre ich glücklich!</u>)	17	0	4.4	12.6	36.4	46.6
22bii	(<u>Wäre ich doch glücklich!</u>)	12	0	0.8	1.8	18.9	78.5
23	Grad-V-1-Exklamativsatz (aus Nr 8) (<u>Wäre ich glücklich!</u>)	27	0	13.4	10	40	36.6
23a	(Realisationen ohne Modalpartikel)	13	0	4.5	9.2	44.1	42.2
23b	(mit Modalpartikel <u>vielleicht</u>)	14	0	20.6	10.7	36.6	32.1

INTONATORISCHE KENNZEICHNUNG VON SATZMODI*

Wilhelm Oppenrieder (München)

0. Überblick

Zu den Merkmalen, die Satzmodi formal kennzeichnen, zählen auch intonatorische Merkmale. Von besonderer Bedeutung sind diese bei intonatorischen Minimalpaaren, bei denen die anderen formalen Merkmale keine Unterscheidung erlauben. Im Münchner DFG-Projekt Modus-Fokus-Intonation wurde ein Korpus von derartigen Minimalpaar-Testsätzen erstellt. Die Testsätze wurden 6 Versuchspersonen zur Realisation vorgelegt und diese Realisationen weiter untersucht: Auf der einen Seite dienten Hörtests zur Feststellung der Natürlichkeit und der kontextfreien Klassifizierbarkeit, auf der anderen Seite erfolgte die Extraktion kennzeichnender akustischer Parameterwerte. Mit Hilfe der Werte derjenigen Realisationen, die aufgrund der Hörtests als 'gute und typische Exemplare' des entsprechenden Satzmodus eingestuft werden können, wurden intonatorische Prototypen gebildet, die die zentrale und typische Realisierungsform des entsprechenden Satzmodus in unserem Korpus darstellen.

1. Das Satzmodussystem des Deutschen und die Rolle der intonatorischen Merkmale bei der Konstitution der Formtypen

Der Untersuchung liegt das von Hans Altmann entworfene Satzmodusmodell zugrunde (vgl. Altmann (1984), (1987)). Satzmodi werden in diesem Modell aufgefaßt als komplexe formale Strukturen (=Formtypen), denen eine abstrakte Bedeutung (=Funktionstyp) zugeordnet ist. Die formtypkonstituierenden Merkmale stammen aus

* Für Hilfe und Beistand sei hiermit den anderen Mitgliedern des DFG-Projekts Modus-Fokus-Intonation, sowie Eva Breindl, Maria Thurmair und Eva-Maria Willkop gedankt.

vier verschiedenen Merkmalsmengen. Bedingung für die Aufnahme eines Merkmalsbereichs oder eines Einzelmerkmals in den Katalog der satzmodusrelevanten Merkmale ist, daß ein solches Merkmal für die Unterscheidung von mindestens zwei verschiedenen Satzmodi essentiell ist.

(a) Die erste Menge bezieht sich auf **satzmodusrelevante Ausdruckskategorien**, also zumindest auf die An- oder Abwesenheit eines finiten/infiniten Verbs, eines *w*-Ausdrucks, eines die Endstellung des Finitums steuernden Satzeinleiters sowie die Zulässigkeit oder Obligatheit von Modalpartikeln.

(b) Die zweite Menge enthält **Stellungsmerkmale** von einigen dieser Ausdruckskategorien, nämlich die Stellung eines finiten Verbs (Erst-, Zweit- oder Letzt-Stellung) und eines *w*-Ausdrucks (vorfeldgebunden oder frei).

(c) In der dritten Menge befinden sich die **morphologischen 'Verbmodus'-Merkmale**.

(d) Die vierte schließlich enthält die für eine Satzmodusdifferenzierung relevanten **intonatorischen Merkmale** - intonatorisch in einem weiten Sinn verstanden, also nicht auf Grundfrequenz, bzw. Tonhöhe, beschränkt.

Daß diese vierte Merkmalsmenge angesetzt werden muß - wenn auch im Gegensatz zu den anderen drei Mengen nicht klar ist, was denn nun die relevanten Einzelmerkmale sind -, zeigt sich sofort, wenn man versucht, die einzelnen Formtypen durch eine charakteristische Merkmalsmenge voneinander abzugrenzen: Die nicht-intonatorischen Merkmale ermöglichen in zahlreichen Fällen keine Unterscheidung, obwohl offensichtlich differenziert werden muß, da diesen scheinbar 'ambigen' Strukturen zwei (in Einzelfällen auch drei) klar unterscheidbare Funktionstypen zugeordnet werden können; ganz eindeutig sorgt hier die intonatorische Form für die konsistente Unterscheidung dieser 'intonatorischen Minimalpaare'.

Beschränkt man sich auf die Strukturen mit Erst- und Zweit-Stellung des finiten Verbs, dann ergeben sich die folgenden

unvollständigen Merkmalsmengen - unvollständig sind sie, da u.a. die intonatorischen Merkmale fehlen:¹

- {[+Verb-Zweit], [-w-Ausdruck], [-Imperativmorphologie]}
(1) = Aussagesatz vs. assertive Frage (*Du kommst./?*)
- {[+Verb-Zweit], [-w-Ausdruck], [-Imperativmorphologie],
(['graduierbares Prädikat'])}
(2) = Verb-Zweit-Exklamativsatz vs. Aussagesatz (*Der hat (vielleicht) geflucht!/. ; Die ist (vielleicht) naiv!/.*)
- {[+Verb-Erst], [-(eindeutig) 'nicht-indikativische' Verbmorphologie], [+Subjektsausdruck]}
(3) = Entscheidungsfragesatz vs. Adhortativsatz (*Gehen wir (denn jetzt)?/(doch jetzt)!*)
(4) = Entscheidungsfragesatz vs. Sie-Imperativsatz (*Schlafen Sie?!*)
(5) = Entscheidungsfragesatz vs. Imperativsatz vs. Verb-Erst-Exklamativsatz (*Stellt ihr euch (etwa/doch/vielleicht) an?!*)
- {[+Verb-Zweit], [+w-Ausdruck (im Vorfeld)]}
(6) = Ergänzungsfragesatz vs. w-Verb-Zweit-Exklamativsatz (*Wie alt ist er geworden?! ; Wie laut ist es hier?! ; Wie ist der (denn/doch) reich geworden?!*)
- {[+Verb-Zweit], [+ (betonter) w-Ausdruck]}
(7) = Versicherungsfragesatz vs. 'und-Ergänzungsfragesatz' (*(Und) wer spielt?*)
- {[+Verb-Zweit], [+w-Ausdruck im Mittelfeld]}
(8) = Versicherungsfragesatz vs. Aussagesatz (*Er sieht was?/.*)
- {[+Verb-Erst], [+Konjunktiv-II-Markierung]}
= Entscheidungsfragesatz vs. Wunschsatz (vs. Verb-Erst-Exklamativsatz) (*Hätte er (wohl/doch) getroffen?! ; Wäre ich (denn/doch/vielleicht) glücklich?!/!*)

Auch bei selbständig verwendbaren Verb-Letzt-Strukturen genügen die Merkmale der ersten drei Mengen nicht für eine formale Differenzierung, vgl. das folgende Beispiel:

¹ Die verdeutlichenden Beispielsätze stammen aus unserem im Anschluß beschriebenen Korpus, das zum Teil auch (modal)partikelhaltige Strukturen umfaßte; allerdings werden hier nicht alle im Korpus auftretenden Minimalpaare behandelt. Die Nummerierung bezieht sich auf die in Abschnitt 5 und im Anhang vorgestellten 'intonatorischen Prototypen' zu den jeweiligen Satzmodi; bei den am Ende der Verb-Erst/Verb-Zweit-Strukturen aufgeführten Wunschsätzen fehlt der Verweis, da kein 'intonatorischer Prototyp' gebildet werden konnte.

- {[+Verb-Letzt], [+w-Ausdruck als Satzeinleiter]}

(9) = Wortfrage-Übernahme vs. w-Verb-Letzt-Exklamativsatz (*Wie groß der ist?/!*)

2. Die Minimalpaarmethode: Gewinnung eines Testsatzkorpus

Für die Auffindung und Untersuchung der für die Satzmoduskennzeichnung essentiellen intonatorischen Merkmale sind nun gerade derartige Strukturen besonders geeignet, bei denen allein die intonatorische Form für die Unterscheidung der Satzmodi sorgen kann, da alle anderen Merkmale neutralisiert sind; mit ihnen lassen sich sogenannte intonatorische Minimalpaare bilden (vgl. Altmann (1984)).

Entsprechend diesen Überlegungen wurden der Untersuchung die soeben aufgeführten Verb-Erst- und Verb-Zweit-, sowie einige Verb-Letzt-Strukturen zugrundegelegt, die ohne intonatorische Markierung 'satzmodusambig' sind und bei denen die intonatorische Form demzufolge funktional stark belastet ist.

Dazu wurden die Testsätze, die oben in Klammern aufgeführt sind, in (sprachliche und situationsbeschreibende) Kontexte eingebettet, die zum ersten natürlich den Satzmodus festlegen, zum zweiten aber auch die Hauptakzentsetzung (bzw. die Fokussierung) lenken sollten, so daß die in Wortwahl und Wortfolge identischen Strukturen auch nach Möglichkeit dieselben akzentuellen Hervorhebungen aufweisen und sich somit tatsächlich nur in der Satzmodusindizierung (und in manchen Fällen jeweils in einer Version auch durch eine charakteristische Modalpartikel!) voneinander unterscheiden.²

Das Bestreben, identische Akzentverhältnisse zu erzielen, führt allerdings dazu, daß in allen Paaren/Tripeln mit (Verb-Erst- und Verb-Zweit-)Exklamativsätzen die Nicht-Exklamativsatz-Exemplare eine markierte 'kontrastive' Akzentstruktur aufweisen: Da zu den von uns ursprünglich angenommenen Kennzeichen dieser Exklamativstrukturen ein 'Exklamativakzent' auf einem definiten Ausdruck - in den Beispielstrukturen immer einem Pronomen in

² Beim Minimalpaar 8 sind keine identischen Akzentstrukturen zu erzielen, da in Versicherungsfragesätzen der w-Ausdruck stets den Hauptakzent trägt, das formgleiche Indefinitpronomen *was* im Aussagesatz dagegen nie betont werden kann.

Subjektsfunktion - gehört (vgl. Altmann (1984:149)), mußte dieser Ausdruck natürlich auch in den Realisationen des jeweils anderen Satzmodus im Minimalpaar den Hauptakzent erhalten.

Die Materialgrundlage für die Untersuchungen bildeten die in Produktionstests gewonnenen Realisationen der oben zitierten Testsätze.

Für diese Tests wurden die insgesamt 71 Testsatz-Kontext-Kombinationen (ohne spezielle Kennzeichnung des Testsatzes) auf DIN A6-Kärtchen geschrieben; jede Kombination war zweimal vertreten. Die 142 Kärtchen wurden in (pseudo-)randomisierter Reihenfolge 6 'naiven' Versuchspersonen (3 weiblich, 3 männlich; Linguistikstudent/inn/en ohne Kenntnis des zugrundeliegenden Satzmodussystems) vorgelegt, die sich in die jeweils beschriebene Situation versetzen und die zu lesenden Sätze möglichst situationsangemessen und natürlich produzieren sollten.

Die Tests fanden im schalltoten Raum des Instituts für Phonetik der Universität München statt (Aufnahme auf Telefonen M15-Tonbandgerät, 19 cm/sec).

Die Versuchspersonen (VPen) konnten sich selbst korrigieren, wenn sie mit ihren Realisationen nicht zufrieden waren. Zusätzlich vermerkten die Versuchsleiter außerhalb des schalltoten Raums diejenigen Fälle, in denen ihrer Meinung nach die Testsätze nicht kontextgerecht und natürlich realisiert worden waren; die entsprechenden Karten wurden den Testpersonen am Ende des normalen Durchlaufs noch einmal vorgelegt.

Wir erhielten insgesamt 956 Realisationen, die weiter analysiert und verarbeitet wurden.

Zur Problematik der Verwendung gelesener Realisationen:

Selbstverständlich ist es nicht unproblematisch, einer Untersuchung intonatorischer Phänomene gelesene Realisationen von vorgegebenen Satzstrukturen zugrunde zu legen - selbst wenn trotz des zusätzlichen Einflusses der ungewöhnlichen Testsituation verhältnismäßig 'natürlich' gelesen wird. Im schlimmsten Fall charakterisieren die Realisationen eben nur die Intonation gelesener und nicht die spontaner Äußerungen.³

Der Rückgriff auf gelesene Realisationen konstruierter Testsätze ist jedoch unabdingbar, wenn mit Hilfe der Minimalpaarmethode die für die Satzmodusdifferenzierung essentiellen intonatorischen Merkmale herausgefunden werden sollen - abgesehen von dem simplen technischen Grund, daß Aufnahmen im schalltoten Raum mit genau eingestellter Apparatur die Gewähr für eine problemlose

³ Brown et al. (1980:17f.) weisen z.B. darauf hin, daß die Wiedergabe eines vorgegebenen und beim Lesen interpretierten Textes an den Produzenten ganz andere Anforderungen stellt als die spontane Verfertigung eines Textes.

Analyse des Sprachsignals bieten. Zum einen muß offensichtlich ein unvertretbar hoher Aufwand betrieben werden, um Realisationen seltenerer Satztypen, z.B. Wunschsätze oder w-Verb-Zweit-Exklamativsätze, 'in freier Wildbahn' zu finden. Zum anderen ist nicht zu erwarten, daß Exemplare von Minimalpaar-Satztypen mit nur einigermaßen vergleichbarer lexikalischer Füllung und Akzentstruktur aufzufinden sind; damit wird die Aufstellung eines Inventars satzmodusdifferenzierender intonatorischer Merkmale aber stark behindert, da eben mehr als nur die intonatorische Form variiert. Durch die Konstruktion wortfolgen- und (soweit möglich) fokusgleicher Testsätze wird demgegenüber dafür gesorgt, daß tatsächlich nur die intonatorische Form zwischen den beiden 'Satzmodus-Lesarten' unterscheidet (vgl. auch die Argumente für die Verwendung von gelesenen Realisationen sorgfältig konstruierter Testsätze in Cooper/Sorensen (1981:20)).

Zudem scheint die Annahme plausibel, daß in spontaner Rede im Prinzip dieselben intonatorischen Merkmale bei der Satzmoduskennzeichnung eingesetzt werden, wenn sich auch die jeweilige Ausprägung dieser Merkmale bei spontanen und vorgefertigten Äußerungen unterscheiden dürfte. So stellt z.B. Trof (1985) bei einem Vergleich der F_0 -Kontur spontan geäußerter Ergänzungsfragesätze mit der F_0 -Kontur der nach 3 Monaten von denselben Sprechern gelesenen Wiederholungen dieser Strukturen (allerdings mit der Vorgabe der Hauptakzentsilbe und der Tonhöhe am Äußerungsende, um Vergleichbarkeit zu gewährleisten) fest, es gebe zwischen den beiden Realisationsformen zwar einen quantitativen, aber keinen qualitativen Unterschied (Trof (1985:57)).

Vielleicht ist in spontanen Äußerungen ein eindeutiger Kontext von größerem Einfluß, so daß intonatorische Differenzierungen minimiert werden können; und sicherlich trägt auch die Tatsache, daß die satzmodusfähigen sprachlichen Komplexe in der spontanen Rede nicht so wohlgegliedert sind wie in den vorgegebenen Texten, dazu bei, daß nicht immer eine prototypische intonatorische Form realisiert wird. Dennoch halten wir es nicht für wahrscheinlich, daß in der spontanen Rede völlig andere intonatorische Merkmalskomplexe für die Satzmodusdifferenzierung sorgen oder daß sich gar die Rolle der Intonation im Gesamtkomplex der satzmoduskennzeichnenden Merkmale wesentlich verändert.

3. Untersuchungen am Testsatzkorpus: Hörtests und Extraktion von Parameterwerten

Die weiteren Untersuchungen am Korpus der 956 Testsatzrealisationen erfolgten in zwei verschiedenen Richtungen. Auf der einen Seite wurde in Hörtests mit durchschnittlich einem Dutzend Teilnehmern (Student/inn/en der Linguistik oder Phonetik ohne genaue Kenntnis des Satzmodussystems) die 'Natürlichkeit', d.h. Kontextangemessenheit, und die kontextfreie Identifizierbarkeit der einzelnen Realisationen überprüft sowie die Position der am stärksten betonten Silbe festgestellt. Auf der anderen Seite wurden aus den aufgezeichneten F_0 - und Intensitätsverläufen charakteristische Parameterwerte extrahiert.

3.1 Hörtests

Es handelt sich um drei Typen von Hörtests, nämlich Akzent-, Natürlichkeits- und Kategorisierungstests.

(a) Akzenttests

Bei den Akzenttests bekamen die VPen die Realisationen kontextfrei dargeboten; gleichzeitig lag ihnen eine schriftliche Fassung der Testsätze vor, auf der sie die ihrer Meinung nach am stärksten akzentuierte Silbe markieren mußten. Die Silbe einer Realisation, auf die die meisten Akzentuiertheitsurteile fallen, nennen wir die Hauptakzentsilbe oder A1-Silbe.

Die Feststellung der Position dieser Silbe ist äußerst wichtig, da sich auf ihr, bzw. in ihrer Umgebung, für die Satzmodus-kennzeichnung besonders relevante intonatorische Merkmale manifestieren (vgl. z.B. das Konzept der 'Schwerpunktsilbe' bei v. Essen (1964) und der 'Schwerpunkt-' bzw. 'Tonsilbe' bei Pheby (1980:51), Pheby (1981:855), sowie der 'tonic syllable' bei Halliday (1976:217)).

Zudem sollte mit diesen Tests auch überprüft werden, ob die Akzentsteuerung durch den Kontext gelungen war. Dies war bis auf eine Ausnahmegruppe tatsächlich der Fall. Die Ausnahmegruppe umfaßt die (Verb-Erst- und Verb-Zweit-)Exklamativsatz- und zu einem geringeren Teil auch die Wunschsatzrealisationen: Hier gibt

es offensichtlich sprecherspezifische Strategien der Akzentuierung, die den akzentlenkenden Einfluß des Kontexts überspielen: Während bei den Exklamativsatzrealisationen in vielen Fällen der Hauptakzent wie erwartet auf das definite (und 'thematische') Subjektspronomen fällt, finden sich auch Realisationen mit Hauptakzent auf dem (im Vergleich zum Subjekt eher 'rhematischen') Prädikatsausdruck, daneben aber auch Versionen mit 'Akzentsplitting', d.h. einer annähernden Gleichverteilung der Akzentuierungsurteile auf Subjekts- und Prädikatsausdruck. Akzentsplitting tritt auch teilweise bei den Wunschsatzrealisationen auf, wobei neben dem infiniten 'Prädikatskern' auch das initiale Finitum hervorgehoben wird. Wenn der Kontext bei der Akzentpositionierung derart zurücktritt, dann gilt für die entsprechenden Satztypen die übliche Fokus-Hintergrund-Strukturierung nicht. Das ist nicht ganz unerwartet, da Sprachhandlungen, die mit den beiden Satztypen Exklamativ- und Wunschsatz typischerweise ausgeführt werden, nicht adressatenbezogen sind, sondern quasi außerhalb der 'Sprachhandlungskette' stehen. Die weitgehende Kontextunabhängigkeit der Akzentstruktur von Exklamativsatzrealisationen trat auch in Folgeuntersuchungen auf (vgl. z.B. Batliner (1988a,b)), ist also gut abgesichert.

(b) Natürlichkeitstests

Bei den Natürlichkeitstests lag den VPen neben dem eigentlichen Testsatz auch die gesamte Kontextbeschreibung schriftlich vor; der Testsatz war jeweils markiert. Sie mußten beurteilen, wie natürlich die Realisation des entsprechenden Testsatzes relativ zum Kontext ist, wie kontextgerecht jener also realisiert worden war. Die Beurteilung erfolgte gemäß einer Ratingskala von 1 bis 5, auf der 1 für eine sehr natürliche, 2 für eine natürliche, 3 für eine noch natürliche, 4 für eine unnatürliche und 5 für eine sehr unnatürliche Realisation zu vergeben war.

Jeder Testsatz-Realisation kann also ein Natürlichkeitswert als arithmetisches Mittel aus den einzelnen Natürlichkeitsbewertungen der VPen zugeordnet werden.

(c) Kategorisierungstests

Bei den Kategorisierungstests lautete die Anweisung, die kontext-

frei dargebotenen Realisationen den fünf durch eine funktionale Beschreibung und durch einige typische Beispiele eingeführten Gruppen **Aufforderung, Frage, Aussage, Ausruf/Exklamativ** und **Wunsch** zuzuordnen (vgl. dazu Oppenrieder (1988)). Erst derartige Tests zur kontextfreien Identifikation (der 'typischen Funktion') des Satzmodus einer Äußerung zeigen, ob die intonatorische Form dem Hörer genügend Information für die 'Disambiguierung' von Strukturen liefert, die aufgrund aller anderen formalen Merkmale (mit Ausnahme der in manche Testsätze eingefügten Modalpartikeln) 'satzmodusambig' sind.

Die Ergebnisse der Natürlichkeits- und Kategorisierungstests benutzten wir als **Filter** für die Auswahl derjenigen Realisationen, die bei der Bestimmung der satzmodusrelevanten intonatorischen Merkmale der einzelnen Satzmodi berücksichtigt werden.

Durch diesen Filter kommen nur Exemplare, deren Natürlichkeitswert kleiner als 2.5 ist (ein Wert, der zwischen den Natürlichkeitsstufen 'natürlich' und 'noch natürlich' liegt) und die gleichzeitig im Kategorisierungstest zu mehr als 80% wie erwartet klassifiziert worden waren. Der Zweck ist, diejenigen Realisationen herauszufiltern, bei denen erstens die intonatorische Form allein für eine konsistente Satzmodusmarkierung ausreicht und die zweitens auch in ihrem Kontext als ohne weiteres akzeptable Realisationen gelten können.

Von den insgesamt 956 Realisationen passierten den Filter 353, also gut 36%.

3.2 Extraktion der akustischen Parameterwerte

In einem parallelen Verarbeitungsstrang wurden charakteristische akustische Parameterwerte extrahiert. Zu diesem Zweck erstellten wir Mingogramme vom **Zeitsignal**, vom **Grundfrequenzverlauf** (gemessen mit einem Frøkjær-Jensen-Pitchmeter) und vom **Intensitätsverlauf** (gemessen mit einem Frøkjær-Jensen-Intensitymeter) jeder Realisation. Anhand dieser Mingogramme wurden systematisch für alle 956 Realisationen die folgenden Grundfrequenz-, Intensitäts- und Dauerwerte extrahiert:

- (1) der Hz-Wert am Beginn (= 'Onsetwert') und am Ende (= 'Offsetwert') jeder Äußerung;
- (2) der Hz-Wert der beiden höchsten (lokalen) F_0 -Maxima und der beiden tiefsten F_0 -Minima (sofern tatsächlich in der Äußerung zwei oder mehr dieser beiden Arten von Extremwerten vorkommen);
- (3) die Hz-Werte der 'Eckpunkte' der beiden (sofern zwei oder mehr vorhanden sind) im Umfang größten ' F_0 -Bewegungen'; als solche 'Bewegungen' zählen neben einem einfachen Anstieg und Fall (Eckpunkte A und B) auch komplexere Konfigurationen mit drei Eckpunkten, wobei im mittleren Punkt B eine deutliche Änderung in der Richtung des F_0 -Verlaufs auftritt (bis auf wenige Ausnahmen ein Umschlag von Anstieg in Fall oder umgekehrt); über stimmlose Abschnitte, die zwischen den Eckpunkten liegen, wurde interpoliert;
die Ablesegenauigkeit für alle diese F_0 -Werte betrug 5 Hz; für jeden Wert wurde festgehalten, auf die wievielte Silbe der Äußerung er fällt;
- (4) die beiden Silben mit den höchsten dB-Werten;
- (5) die Dauer (des stimmhaften Teils) der in den Akzenttests festgestellten A1- oder Hauptakzentsilbe;
- (6) die Dauer der Gesamtäußerung (vom F_0 -Onset bis zum F_0 -Offset).

Jeder Realisation wird also u.a. eine genau charakterisierte Menge von akustischen Parameterwerten zugeordnet, die eine Vergleichbarkeit der einzelnen Realisationen ermöglicht.

Hinter dem schematischen Vorgehen bei der Auswertung liegt zum einen die Absicht, möglichst wenig an Interpretation in die Auswahl der akustischen Kennwerte einfließen zu lassen.⁴

Zum anderen sollte die Auswertungsmethode zumindest dem Geist, wenn auch nicht dem laxen Buchstaben nach ohne größere Veränderungen in einen Algorithmus umformulierbar sein, der die Basis für eine automatische Extraktion der satzmodusrelevanten akustischen Parameterwerte abgibt - was die Bearbeitung großer Korpora natürlich sehr erleichtern würde - und der damit auch für eine automatische Satzmoduserkennung genutzt werden kann (etwa

⁴ Insbesondere die Regeln für die Festlegung der relevanten F_0 -Bewegungen lassen allerdings noch einen beträchtlichen 'Interpretationsspielraum' zu; einen solchen Spielraum gibt es natürlich auch beim Ablesen der konkreten Werte von den Mingogrammen, angefangen vom Problem, die Kontexteffekte im F_0 -Bereich einigermaßen auszuschalten (vgl. Thorsen (1979: 60ff.)), bis zu der Einschätzung von Ausreißern bei den vom Oscillomink aus geschriebenen F_0 -Werten. An diesem Punkt bewährt sich wiederum die Minimalpaarmethode, da beim Vergleich von Minimalpaaräußerungen Veränderungen des Grundfrequenzverlaufs durch intrinsic pitch oder Kontexteffekte nicht ins Gewicht fallen.

mittels diskriminanzanalytischer Verfahren; vgl. dazu Batliner (1987) und Batliner (1988c), sowie das Verfahren in Nöth et al. (1987)).

3.3 Was sind die für die Satzmoduskennzeichnung relevanten intonatorischen Parameter? - Die Rolle der Tonhöhe in linguistischen Intonationstheorien

Die Auswahl der akustischen Kennwerte spiegelt die allgemeine Ansicht wider, daß im Gesamtkomplex der linguistisch relevanten intonatorischen Eigenschaften einer Äußerung die Grundfrequenz eine dominante Rolle spielt. Für die intonatorische Satzmodus-kennzeichnung von überragender Bedeutung ist danach die Form des Grundfrequenzverlaufs (Tonhöhenverlaufs) 'in der Gegend' der Nukleussilbe und die Grundfrequenz (Tonhöhe) am Äußerungsende.⁵

Für v. Essen sind die Schwerpunktsilbe und die möglicherweise noch folgenden Nachlaufsilben Träger der zeichenhaften Melodeme (v. Essen (1964:15f.)); Bierwischs Regelapparat sorgt dafür, daß die (für den Satzmodus) entscheidende Tonhöhenveränderung bei der Nukleussilbe stattfindet und die erreichte Tonhöhe sich bis zum Ende des Satzes fortsetzt (Bierwisch (1971:146ff.)); bei Halliday sind die (zum Ausdruck verschiedener 'speech functions', u.a. des Satzmodus, dienenden) 'tones' durch den Tonhöhenverlauf in der 'tonic syllable' und den daran anschließenden Silben bis zum Ende der 'tone group' gekennzeichnet (Halliday (1976:221)); ebenso wird nach Pheby (1980) und (1981) das Tonmuster in der Schwerpunktsilbe und den folgenden Silben bis zum Ende der Tongruppe realisiert. Auch in Arbeiten im Rahmen des 'Tonsequenzen-Ansatzes' sind die tonale Charakteristik der Fokussilbe und die Art des 'Grenztons' kennzeichnend für den Satzmodus (vgl. die Beschreibung der Echofragen-Intonation in Wunderlich (1986:51ff.) und einige der Analysen in Ladd (1983b)).

Neben dem Offset- wird auch der Onsetwert notiert, da wir es zum einen für möglich halten, daß auch der letztere zur Satzmodus-kennzeichnung beiträgt; zum anderen läßt sich mit Hilfe der (tiefen) Onset- und Offsetwerte eine sprecherspezifische 'Baseline' berechnen, ähnlich der, die in den Tonsequenzen-Theorien als Referenzlinie für die auf ihr aufgereihten 'Tonhöhen-Akzente'

⁵ Wir schließen natürlich nicht aus, daß der Satzmodus auch durch globalere Mittel signalisiert werden kann oder daß die der Nukleussilbe vorangehenden Akzentsilben eine satzmoduscharakteristische Kontur aufweisen.

eine bedeutende Rolle spielt (vgl. Pierrehumbert (1980), Ladd (1983a)).

Die von uns erfaßten F_0 -Bewegungen stellen das Pendant zur Tonbewegung auf und nach der Nukleussilbe in den erwähnten Theorien dar; obwohl diese F_0 -Bewegungen ohne den Blick auf die Akzentstruktur der jeweiligen Äußerung erfaßt wurden, fällt die in den Akzenttests festgestellte Hauptakzentsilbe ausnahmslos in eine der beiden Bewegungen. (Allerdings sind die meisten Testsätze sehr kurz, so daß in vielen Fällen ein Zusammenfall schon allein deshalb zu erwarten ist, weil die beiden F_0 -Bewegungen den gesamten F_0 -Verlauf weitgehend abdecken.)

Die Extremwerte (Maxima und Minima) sind wichtig für die Berechnung des Grundfrequenz-Umfangs (=Range), der in einer Äußerung ausgeschöpft wird; zudem können diese Extremwerte und ihre Position natürlich auch mit den Ergebnissen der Akzenttests in Beziehung gesetzt werden.

Wir nehmen an, daß die Grundfrequenzwerte (Tonhöhenwerte) an bestimmten relevanten Punkten einer Äußerung eine überragende Rolle bei der Satzmoduskennzeichnung spielen. Dennoch halten wir es für wahrscheinlich, daß manche Satzmodi durch diesen Faktor allein noch nicht hinreichend von den entsprechenden anderen Satzmodi im Minimalpaar abgegrenzt werden können. Kandidaten sind hier insbesondere die verschiedenen Spielarten des 'nicht-klassischen' Satzmodus Exklamativ, die im ursprünglichen Merkmalssystem durch das Merkmal eines Exklamativakzentes - wie auch immer dieser realisiert sei - von den formalen Nachbarmodi abgegrenzt wurden (vgl. Altmann (1984:149f.)). Wir erwarteten, daß der Exklamativsatz durch irgendwelche 'Extensionsphänomene' gekennzeichnet sein würde. Im F_0 -Bereich bedeutet das einen besonders großen Range; wir vermuteten jedoch, daß auch die Intensität und vor allem die Dauer (der Gesamtäußerung oder der betonten Silben) in Relation zu den Verhältnissen bei Realisationen des jeweils anderen Minimalpaar-Satzmodus eine Rolle spielt. Daher wurden auch die erwähnten Dauerwerte extrahiert und die Intensitätsmaxima lokalisiert.⁶

⁶ Möglicherweise haben sogar spektrale Eigenschaften einen Einfluß auf die Satzmoduskennzeichnung. Der Faktor wurde von uns aber bisher noch nicht berücksichtigt.

Bei der Auswahl der akustischen Kennwerte für die Analyse der Testsatzrealisationen ist zu bedenken, daß es nur darauf ankam, diejenigen Werte zu erfassen, die bei der Satzmoduskennzeichnung aller Wahrscheinlichkeit nach eine wesentliche Rolle spielen.

Eine exakte Erfassung des gesamten Tonverlaufs, wie es etwa bei Brown et al. (1980:52ff.) oder beim IPO-Modell (dort mit Hilfe einer Verkettung perceptiv relevanter stilisierter F_0 -Veränderungen, die auch in einem generativen Algorithmus verwendet und getestet werden können, vgl. 't Hart/Collier (1975) und eine erste Anwendung aufs Deutsche in Adriaens (1984)) geschieht, lag also nicht in unserer Absicht: Die Aufgabe wäre angesichts der 956 Realisationen kaum zu bewältigen gewesen; zudem sind für die Untersuchung der satzmodusdifferenzierenden intonatorischen Eigenschaften doch zunächst wieder nur die Werte an bestimmten 'auffälligen' Stellen relevant.

Eine gezielte Auswahl der zu extrahierenden Daten wird auch in den Arbeiten der Gruppe um Cooper vorgenommen, wo ähnlich wie bei uns die gelesenen Realisationen von systematisch konstruierten Testsätzen analysiert werden. Allerdings werden dort die Werte bei bestimmten vorher ausgewählten 'Schlüsselwörtern' extrahiert, um die Realisation einer Fokusstruktur (vgl. Cooper et al. (1985); Eady et al. (1986)) oder die Kennzeichnung bestimmter syntaktischer Grenzen (Cooper/Sorensen (1981:101 ff.)) zu untersuchen, während wir (bis auf die Dauermessung der A1-Silbe) ohne Berücksichtigung der 'Trägerstruktur' charakteristische Grundfrequenz- und Intensitätswerte extrahieren.

3.4 Bemerkungen zum Tonsequenzenansatz

Im Rahmen eines Tonsequenzen-Ansatzes (vgl. z.B. Pierrehumbert (1980); Ladd (1983b); Gussenhoven (1985)) werden Grundfrequenzverläufe phonologisch als Verkettungen von 'Tönen' (nicht zu verwechseln mit den lexikalisch distinktiven Tönen in Tonsprachen!) repräsentiert. Dabei muß die eine Art dieser Töne ('Akzenttöne') mit metrisch prominenten Silben assoziiert werden - also Silben, die Fokusexponenten einer fokussierten Einheit sind oder wenigstens den Wortakzent tragen; die andere ('Grenztöne') charakterisiert dagegen unabhängig von der metri-

schen Struktur die linke und rechte Grenze einer 'Intonationsphrase'. Tonale Grundeinheiten sind (üblicherweise) ein H(och)- und ein T(ief)-Ton, die aber auch, je nach spezieller Theorie, zu bi- oder tritonalen (z.B. in Gussenhoven (1985)) Komplexen zusammengefaßt werden können, innerhalb derer ein Ton als derjenige ausgezeichnet ist, der mit einer metrisch prominenten Silbe assoziiert werden muß.

In einigen Versionen des Ansatzes (vgl. Ladd (1983a,b); Gussenhoven (1985)) werden die Töne zusätzlich durch bestimmte Merkmale charakterisiert, die unterschiedlichen Realisationsvarianten entsprechen.

Der Vorteil des Ansatzes liegt darin, daß er mit wenigen Beschreibungseinheiten auskommt, für deren Verkettung und 'phonetische Realisation' genaue Regeln formuliert werden können. Daß wir uns bei der Beschreibung unserer Daten dennoch nicht auf ein Tonsequenzen-Modell stützen, hängt mit folgenden Punkten zusammen:

Wir gehen davon aus, daß zumindest die Exklamativ-Modi nicht allein durch tonale Besonderheiten von den Nachbar-Modi unterschieden werden können. Insbesondere nehmen wir an, daß die Dauer der Gesamtäußerung und/oder der A1-Silbe und die intonatorische Feinstruktur dieser Silbe, nämlich die Lage eines F_0 -Gipfels in der Silbe und relativ zu einem Intensitätsgipfel, eine Rolle spielen (s.u.). Sollten sich diese Eigenschaften tatsächlich als beschreibungsrelevant erweisen, dann müßten in einem tonalen Ansatz zumindest zusätzliche Merkmale eingeführt werden, die aber dann nicht mehr rein tonaler Natur sind.

Grundsätzlich halten wir es für sinnvoller, nicht impressionistisch irgendwelche vorher unter dem Gesichtspunkt einer möglichst kleinen Menge von Beschreibungseinheiten ausgewählten 'Töne' zuzuweisen, sondern die relevanten F_0 -Veränderungen selbst sehr genau zu kennzeichnen. Das Vorgehen gleicht in diesem Punkt eher dem am IPO praktizierten, wo die (stilisierten) elementaren Einheiten des F_0 -Verlaufs exakt beschrieben werden, etwa durch die Angabe der Position einer Änderung des F_0 -Verlaufs in einer Silbe (vgl. 't Hart/Collier (1975:240ff.); Boves et al. (1984:22)). Erst so wird es möglich, einigermaßen präzise Regeln zu formulieren, wie ein gegebener F_0 -Verlauf mit Hilfe des

Inventars von typischen F_0 -Abschnitten beschrieben werden kann (vgl. die Regeln in Boves et al. (1984:38ff.)). Die typischen F_0 -Abschnitte oder Kombinationen aus ihnen können im Anschluß an die exakte Charakterisierung natürlich durchaus als 'Töne' etikettiert werden, insofern ist eine tonale Beschreibung der Realisationen unseres Korpus' ohne weiteres möglich - wenn auch wahrscheinlich nicht ausreichend.

Bei den üblichen tonalen Ansätzen ist nur die Art und Zuweisung der Grenztöne einsichtig. Im Bereich der Akzenttöne bleibt es dagegen ziemlich unklar, warum genau diese Einzel- und Kombinationstöne zur Beschreibung ausgewählt werden und auf welche Weise sie im konkreten Fall einer Silbenkette zugewiesen werden. Offensichtlich soll der von Kontexteffekten und intrinsischen Segmenteigenschaften 'gereinigte' F_0 -Verlauf erfaßt werden, bei einigen Autoren (z.B. Pierrehumbert (1980)) ziemlich exakt, bei anderen ohne die 'nicht-systemischen' Variationen (vgl. Gussenhoven (1985:121)); auf der anderen Seite müssen die 'Töne' aber sicherlich auch perceptiv relevant sein (oder sich wenigstens in irgendeinem Aspekt des F_0 -Verlaufs 'hörbar machen', wie z.B. der abstrakte L-Ton, der in der Theorie von Pierrehumbert (1980) für das 'Downstepping', also die Tonhöhenabsenkung eines folgenden H-Akzenttones sorgt) und entweder selbst Zeichencharakter besitzen oder als Teil in ein komplexeres Zeichen eingehen. Irgendwelche expliziten signal-, perceptions- oder bedeutungsbezogenen Kriterien für die Auswahl der tonalen Beschreibungseinheiten und die Zuweisungsregeln zu einem gegebenen F_0 -Verlauf werden nirgends formuliert. Die in diesem Ansatz entwickelten Theorien zielen allerdings auch eher auf die Generierung tonaler Strukturen als auf die Analyse gegebener F_0 -Verläufe. Da es uns aber vor allem darauf ankam, die akustischen Eigenschaften der Realisationen unseres Korpus' (zumindest in den für die Satzmoduszeichnung relevanten Abschnitten) möglichst genau und zunächst ohne irgendwelche Interpretationen, z.B. tonaler Art, zu erfassen, legen wir unserer Korpusanalyse keine tonale Theorie zugrunde: Zum einen soll mehr als nur der F_0 -Verlauf berücksichtigt werden, zum anderen soll eine willkürliche Zuweisung von 'Tönen' aufgrund unklarer Kriterien vermieden werden.

4. Das Modell der intonatorischen Prototypen

Die für jede Testsatzrealisation extrahierten Parameterwerte wurden auf Rohdatenfiles übertragen und zusammen mit aus ihnen berechneten weiteren Kennwerten einer deskriptiven statistischen Analyse mit Hilfe des Statistikpakets SPSSPC+ unterzogen. Die Hz-Werte wurden zur besseren Vergleichbarkeit der Sprecher/innen umgerechnet in Halbtonwerte über der Basis 1 Hz und über einer sprecherspezifischen (ebenen) 'Basislinie', deren Höhe durch den tiefsten Offsetwert (nach Abzug der untersten 5% der Offsetwerte, um vereinzelte Ausreißer auszuschalten) bestimmt wurde (vgl. Batliner (1988c)).

Mit Hilfe der Daten, die sich aus der statistischen Analyse ergaben (z.B. die Mittelwerte und Standardabweichungen bei den Grundfrequenz- und Dauerwerten, vgl. die im folgenden aufgezählten Parameter), bildeten wir sogenannte 'intonatorische Prototypen', die den einzelnen Satzmodi als ein Teil des formalen Merkmalsbündels zugeordnet sind. Diese intonatorischen Prototypen sind ganz wie die anderen Merkmalskomplexe aus denjenigen Merkmalen aufgebaut, die sich an wenigstens einer Stelle als wesentlich für die intonatorische Unterscheidung zweier Satzmodi erwiesen haben. Der Aufstellung der Prototypen wurden (bis auf besonders gekennzeichnete Ausnahmen) die Kennwerte von Realisationen zugrundegelegt, die den oben erwähnten Filter passierten, also in der Natürlichkeit mindestens zwischen 'natürlich' und 'noch natürlich' lagen und gleichzeitig von mehr als 80% der VPen kontextfrei wie von uns angezielt kategorisiert wurden.

Für die Prototypen-Bildung wurden bisher die folgenden Parameter systematisch berücksichtigt:

- (1) Im Bereich der Grundfrequenz der Offsetwert. Wir teilen dabei das Kontinuum der Offsetwerte in zwei Abschnitte: Ein Wert, der im Bereich der oberen Hälfte des in der jeweiligen Äußerung realisierten Gesamtumfangs der Grundfrequenz liegt, gilt als hoher Offsetwert, ansonsten ist der Offsetwert tief.
- (2) Weiter wurde die Form des F_0 -Verlaufs im Bereich der A1-Silbe, also der Hauptakzentsilbe aus den Akzenttests, berücksichtigt. Diese Form, hier 'Kontur' genannt, basiert auf den F_0 -Bewegungen. Die Kontur wurde grob

unterteilt in einen **konvexen, konkaven, einfach fallenden** oder **einfach steigenden** Verlauf. Sie kann sich über mehrere Silben erstrecken, enthält aber auf jeden Fall die A1-Silbe. Neben den genannten vier Grobtypen wird auch ein genaueres Notationssystem verwendet. In einem Drei-Punkte-System sind die relative Höhe (keine von der jeweiligen Realisation unabhängigen absoluten Stufen!) der beiden Endpunkte und eines eventuell vorhandenen Umschlagpunktes festgehalten: '1' liegt tiefer als '2', dieses wiederum tiefer als '3' (beträgt der Unterschied weniger als einen Halbton, so gehen wir von einer gleichen relativen Tonhöhe, d.h. von einer perzeptiv irrelevanten Differenz aus), so daß z.B. ein 231-Verlauf ein konvexer Verlauf ist, innerhalb dessen der F_0 -Abfall größer ist als der F_0 -Anstieg; der Tonumfang des Anstiegs und Abfalls kann gesondert berechnet werden, dagegen läßt unsere Art der Datenextraktion (fehlende zeitliche Lokalisation der F_0 -Werte) keinen Schluß auf den (durchschnittlichen) Neigungswinkel der betreffenden F_0 -Veränderungen zu.

Die beiden Merkmale Offsetwert und Kontur genügen in vielen Fällen für die Unterscheidung zweier Satzmodi; vgl. die Beschreibung der einzelnen Prototypen weiter unten in Abschnitt 5. Das ist nicht weiter überraschend und wird ja auch von den weiter oben erwähnten Theorien angenommen.

Neu ist jedoch, daß die Ergebnisse zum einen an einem großen Korpus gewonnen wurden, das auf Testsätzen beruht, die systematisch aufgrund eines detailliert ausgearbeiteten Satzmodus-Systems konstruiert wurden, das den intonatorischen Merkmalen einen bestimmten Platz im Gesamtkomplex der formtypkonstituierenden Merkmale zuweist, und daß die Ergebnisse zum anderen auf der Grundlage einer schematischen - im Prinzip algorithmisierbaren - Auswertungsmethode erzielt wurden.

Obwohl Offsetwert und Kontur die zentralen Merkmale zur intonatorischen Kennzeichnung verschiedener Satzmodi sind, gewährleisten sie doch bei manchen Minimalpaar-Satzmodi (Paare mit Exklamativsätzen s.u.) noch keine genügende Abgrenzung. Als Zusatzmerkmale wurden daher von uns berücksichtigt:

- (3) Die durchschnittliche Gesamtdauer (D_{Gesamt}) der jeweiligen Äußerungen eines Testsatzes und die durchschnittliche relative Dauer (D_{rel}) der A1-Silbe, mit $D_{\text{rel}} = \text{A1-Dauer} / (D_{\text{Gesamt}} / \text{Silbenzahl})$.
Bei der Berücksichtigung der Dauer als Kennwert muß beachtet werden, daß (ohne die Berechnung einer 'normalen Silbendauer') nur zwischen Minimalpaar-Strukturen mit der gleichen Wortfolge verglichen werden kann; zusätzlich

müssen auch die zu vergleichenden A1-Silben dieselbe Position in der Silbenkette haben.

- (4) Ein weiteres peripheres Zusatzmerkmal sind der F_0 -Gesamtrange der Äußerung und der F_0 -Range der Kontur. Insbesondere bei Satzmodi, die in ihrer Grundfunktion eine 'expressive Komponente' aufweisen, wie das bei den verschiedenen Exklamativsatztypen der Fall ist, ist mit der Rangeextension als Zusatzmerkmal zu rechnen.

Möglicherweise sind noch weitere Parameter für die intonatorische Satzmodusdifferenzierung von Bedeutung (s.u. zur 'Feinstruktur' der A1-Silbe); die Merkmalsmenge ist also durchaus noch erweiterbar.

Die weiter unten vorgestellten 'intonatorischen Prototypen' berücksichtigen die extrahierten Parameterwerte der 'guten' (d.h. natürlichen und kontextfrei kategorisierbaren) Testsatzrealisationen.

Um Prototypen handelt es sich, da Abweichungen von diesen zentralen Realisationsformen bis zu einem gewissen Ausmaß toleriert werden. Der Toleranzbereich selbst hängt wieder davon ab, wie wichtig ein Merkmal für die Satzmoduskennzeichnung ist.

Wir nehmen an, daß im Bereich der intonatorischen Kennzeichnung sprachlicher Funktionen in dreifacher Hinsicht zwischen 'zentral' und 'peripher' unterschieden werden muß.

Allerdings ist dem Bereich der formalen Kennzeichnung bereits innerhalb des Systems der Satzmodi selbst die Unterscheidung von periphereren und zentraleren Satzmodi vorgelagert. So stellen Sadock/Zwicky (1985:160) fest, daß in den meisten Sprachen drei grundlegende Satztypen mit ähnlichen Funktionen vorhanden sind, nämlich 'deklarative', 'interrogative' und 'imperativische' Satztypen. Diesen 'major types' stellen sie die 'minor types' gegenüber, die nicht so weit verbreitet sind und typischerweise nicht durch eindeutige Mittel gekennzeichnet werden (ebenda:156). Zu den letzteren zählen sie die 'exclamations'. Auch im Deutschen sind die Exklamativsatztypen, zusammen mit den Wunschsätzen, gegenüber den 'klassischen' Aussage-, Frage- und Imperativsatztypen peripherer und markierter. Entsprechend umstritten war und ist zum Teil immer noch, ob es sie überhaupt 'gibt', ob sie also

für eine Beschreibung der Regularitäten des Deutschen überhaupt benötigt werden (vgl. den Überblick zu den Grammatiken des Deutschen in Näf (1987:140f.)).

Die Rolle der intonatorischen Merkmale bei der Kennzeichnung einzelner Satzmodi ist in dreifacher Hinsicht zentraler oder peripherer:

(a) Auf das Satzmodussystem bezogen ist zum ersten der Beitrag der intonatorischen Form im Gesamtkomplex der formtypkonstituierenden Merkmale je nach Satzmodus mehr oder weniger zentral. Ausschlaggebend ist auf dieser Ebene, inwieweit die nichtintonatorischen Merkmale selbst für eine ausreichende Kennzeichnung sorgen.

So sind z.B. die Imperativsatztypen, bei denen ein Subjektspronomen fehlen kann, also die 'Imperativsätze der 2. Person', durch das Fehlen eines Subjektsausdrucks und/oder die Erststellung des Finitums und/oder die Verbmorphologie (vgl. für das Zusammenspiel dieser Faktoren Altmann (1987:35ff.)) hinreichend gekennzeichnet; abgesehen von den marginalen Rückfragen auf diese Typen von Imperativsätzen, wo gerade wieder die intonatorische Form für die Kennzeichnung sorgt. Bei einer derartigen Merkmalskonstellation spielt die Intonation nur eine sehr periphere Rolle bei der Satzmodusmarkierung.

Fast genauso eindeutig zeigen *w*-Ausdrücke im Vorfeld eines Verb-Zweit-Satzes, daß es sich um einen Ergänzungsfragesatz handelt, so daß sie mit der unmarkierten 'terminalen Intonation', d.h. einem tiefen F_0 -Offset und einer konvexen oder fallenden Kontur versehen werden dürfen. Andererseits können Fragesätze typischerweise (mindestens die 'Letztglieder' eines Alternativfragesatzes mit festgelegter Zahl der möglichen Antworten bilden hier allerdings eine Ausnahme) mit 'steigender Intonation', also einem hohen F_0 -Offset und konkaver oder steigender Kontur (s.u.) produziert werden. Die für die Satzmodusunterscheidung nicht benötigte intonatorische Form kann dann für die Kennzeichnung feinerer 'Untertypen' (z.B. solcher mit unterschiedlicher Antworterverwartung) innerhalb eines Satzmodus eingesetzt werden. Die Variation ist allerdings nur soweit frei, als es zu keinen

Überschneidungen mit anderen Formtypen kommt: Ein hoher F_0 -Offset mit einer konkaven Kontur auf dem akzentuierten w -Ausdruck im Vorfeld kann nicht mehr als Ergänzungsfragesatz interpretiert werden, sondern nur noch als Versicherungsfragesatz.

Das andere Extrem bilden die Exklamativsatztypen, die offensichtlich ganz wesentlich auf eine intonatorische Kennzeichnung angewiesen sind. Dabei spielt eine Rolle, daß diese innerhalb des Satzmodussystems selbst einen relativ peripheren Platz einnehmen. Bezüglich der nicht-intonatorischen Merkmale werden die Exklamativsätze von den 'Haupt-Satzmodi' Verb-Erst-Fragesatz, Aussagesatz und Ergänzungsfragesatz quasi vollständig überdeckt. Die in den Haupt-Formtypen eher periphere intonatorische Form gewinnt bei den im Satzmodussystem markierteren Exklamativsatztypen zentrale Bedeutung; umgekehrt kann ein ansonsten so eisenfestes Merkmal wie die Stellung des Finitums variieren (bei Exklamativtypen ohne w -Ausdruck zwischen Erst- und Zweit-Position, bei solchen mit w -Ausdruck zwischen Zweit- und Letzt-Position). Zwar liefern im konkreten Fall die Semantik der verwendeten Ausdrücke (z.B. die Graduierbarkeit des Prädikats), Modalpartikeln und Interjektionen ('Prä-Exklamative' wie *Mein Gott!*, *Mann!*, *Teufel auch!* usw.) oder auch der Verwendungskontext 'Hinweise' auf die Interpretation einer Struktur, aber alle diese 'Hinweise' können fehlen oder zumindest uneindeutig sein, so daß tatsächlich nur die intonatorische Form für die Identifikation des Satztyps sorgt. Allerdings ist die intonatorische Form - vor allem beim Verb-Zweit-Exklamativ (im Vergleich zum Aussagesatz, zwei zudem auch funktional nahe verwandte Satzmodi) - ein relativ 'schwaches' Merkmal: Es gibt nur einen verhältnismäßig kleinen Kernbereich von eindeutigen Fällen und einen sehr breiten Übergangsbereich von immer unklarereren Fällen (vgl. die Bemerkungen in Näf (1987) zu den Verb-Zweit-Exklamativsätzen, sowie Batliner (1988a)).

(b) Nicht nur die intonatorische Form insgesamt kann je nach Formtyp eine zentralere oder peripherere Rolle im Gesamtkomplex der Merkmale spielen, sondern auch innerhalb des intonatorischen Teilkomplexes gibt es zentralere und peripherere Merkmale. In der Mehrzahl der Fälle sind die F_0 -Offsethöhe und die Kontur im

Bereich der A1-Silbe ein stabiles und ausreichendes Markierungsmittel für die Abgrenzung der Realisationen von Minimalpaar-Satzmodi (s.u. Abschnitt 5). Dies gilt nicht mehr, wenn Exklamativsätze beteiligt sind. Von besonderem Interesse sind dabei die Verb-Zweit-Exklamativsätze: Bei diesen fallen nicht nur die nicht-intonatorischen Merkmale mit denjenigen der Aussagesätze zusammen, sondern auch die im Normalfall zentralen oder 'dominanten' auf die F_0 bezogenen Größen Offsethöhe und Kontur sind neutralisiert. Infolgedessen gewinnen bei ihrer Kennzeichnung die ansonsten peripheren Merkmale der Gesamt- oder der A1-Dauer oder die Feinstruktur der A1-Silbe an Gewicht. Allerdings deuten sowohl die Ergebnisse der Kategorisierungstests als auch die statistische Auswertung der akustischen Kennwerte darauf hin, daß sie verhältnismäßig unsichere Kennzeichnungsmittel sind. Anscheinend gibt es bei ihnen nicht den stabilen zentralen Prototyp, wie er bei F_0 -Offsethöhe und Kontur vorhanden ist, so daß es zu dem oben erwähnten breiten 'Unsicherheitsbereich' bei den Exklamativstrukturen kommt.

(c) Damit ist der dritte relevante Bereich angesprochen, bei dem ein prototypischer Kern allmählich in eine Peripherie ausläuft. Hier handelt es sich um die mehr oder weniger typischen **Ausprägungen** der im 'intonatorischen Prototyp' zusammengefaßten **Einzelmerkmale**.

Für die F_0 -Offsethöhe und die Kontur im Bereich der A1-Silbe ist offensichtlich ein stabiler zentraler Bereich vorhanden. Bei den Realisationen unseres Korpus' zeigte sich dies sowohl in den Kategorisierungstests als auch bei der statistischen Auswertung der extrahierten akustischen Daten. Äußerungen, deren Formtyp intonatorisch durch diese beiden Merkmale vom anderen Formtyp im Minimalpaar unterschieden werden kann, werden sehr zuverlässig kategorisiert; gleichzeitig zeigen die akustischen Daten, daß die verschiedenen Merkmalsausprägungen deutlich unterschiedene Zentralbereiche mit nur unwesentlichen Überlappungen aufweisen.

Im Bereich der Offsetwerte zeigt sich z.B. bei jeder der sechs VPen eine bimodale Verteilung mit einem schwach besetzten Übergangsbereich; bei den Onsetwerten ist die Verteilung dagegen unimodal, d.h. es liegen keine zwei oder mehr deutlich voneinan-

der unterschiedenen zentralen Ausprägungen vor. Die Werte gruppieren sich beim F_0 -Offset also um zwei zentrale Punkte, die den prototypischen tiefen und hohen Offset repräsentieren. Der Streubereich der Werte um diesen idealen Punkt herum ist beim tiefen Offset wesentlich geringer als beim hohen. Das läßt sich so interpretieren, daß beim hohen Offset wesentlich mehr an 'expressiver Variation' möglich ist als beim tiefen, bei dem es tatsächlich nur um das Erreichen der Baseline geht.⁷

Noch deutlicher wird die sehr zuverlässige Indizierungsfunktion der Offsethöhe, wenn die für jede Äußerung aufgrund der Halbierung des F_0 -Range getroffene Etikettierung des Offsetwertes als tief oder hoch (s.o.) mit dem Satzmodus der Äußerung in Beziehung gesetzt wird. Die ('guten') Nicht-Fragesatz-Realisationen in unserem Korpus hatten zu 100% einen - auf die angegebene Weise berechneten - tiefen Offset, die Fragesatz-Realisationen zu 100% einen hohen Offset, mit der Ausnahme der Ergänzungsfragesätze, bei denen der Offset zu 80% tief war. (Zu beachten ist natürlich, daß es sich bei den aufgrund der Minimalpaar-Methode konstruierten Testsätzen nicht immer um quasi neutrale Strukturen handelt; das gilt insbesondere für die Verb-Erst-Fragesätze, s.u.)

Ähnlich eindeutig ist die Ausprägung der Kontur den einzelnen Formtypen zugeordnet - zumindest solange man sich auf die Grobkonturen konvex/fallend vs. konkav/steigend beschränkt.

Die beiden Merkmale der Offsethöhe und der Kontur laden zu einer Deutung als binäre Merkmale ein. Wir glauben jedoch, daß die prototypische Strukturierung dieser Merkmale nicht unterschlagen werden darf (wenn der Offsetwert z.B. Ober- oder Untergrenze des in einer Äußerung verwirklichten Gesamttrange festlegt, kann er eindeutiger als 'tief' oder 'hoch' gewertet werden, als wenn er zwischen dem unteren oder oberen F_0 -Extremwert und der Rangemitte liegt). Sie stehen nämlich selbst am einen Ende eines Kontinuums, in dem sich auch die oben erwähnten weniger eindeutigen Markierungsmittel wie die Zeit-

7 Ähnlich stellt Pierrehumbert (1980:68f.) fest, daß die emphasebedingte Absenkung eines T(ief)-Tones bald an Grenzen stößt, während die Erhöhung eines H(och)-Tones viel geringeren Einschränkungen unterworfen ist.

struktur und die Feinstruktur der A1-Silbe befinden. Bei diesen gibt es keine so deutlich voneinander abgegrenzten zentralen Ausprägungen, so daß die Deutung als binäres (oder ternäres usw.) Merkmal problematisch ist. Einen Schnitt zwischen den 'indizierungsstarken' und den 'indizierungsschwachen' Merkmalen anzusetzen, erscheint nicht gerechtfertigt.

Bevor die von uns ermittelten intonatorischen Prototypen vorgestellt werden, muß noch ein Problem zumindest angesprochen werden, wenn es auch außerhalb unseres Untersuchungsbereichs lag. Besonders relevant für die intonatorische Satzmoduszeichnung ist der F_0 -Verlauf am Ende einer satzmodusfähigen Äußerung. Wo sich dieses Ende befindet, wird zunächst durch die (nicht-intonatorische) syntaktische Struktur einer Konstruktion festgelegt. Allerdings ist durchaus damit zu rechnen, daß auch mit Hilfe des Verlaufs der Grundfrequenz (Tonhöhe) oder anderer intonatorischer Mittel angezeigt werden kann, ob eine satzmodusfähige Äußerung abgeschlossen ist.

Das Problem besteht darin, daß eine satzmodusfähige Äußerung aus mehr als einer intonatorischen Phrase bestehen kann. Es geht also darum, zwischen letzten und nicht-letzten Phrasen zu unterscheiden, da eben nicht schon trivialerweise klar ist, zu welcher Phrase der satzmoduszeichnende rechte Rand gehört. Von besonderem Interesse ist die Unterscheidung zwischen dem, was von manchen Intonationsforschern (z.B. v. Essen (1964:15)) als progredienter Melodieverlauf bezeichnet wird, und den satzmodusrelevanten steigenden und fallenden Verläufen.

Vielleicht endet eine nicht-letzte intonatorische Phrase typischerweise in einer mittleren Offsethöhe nach einer verhältnismäßig wenig ausgeprägten F_0 -Kontur auf der letzten 'Hauptakzentsilbe' der Phrase. Eine eindeutige Abgrenzung gegenüber weniger prototypischen Ausprägungen der satzmoduszeichnenden Merkmale ist jedoch trotzdem nicht gegeben. Auf die mögliche Relevanz eines weiteren Merkmals deuten die in Isačenko/Schädlich (1971) dargestellten Experimente hin. Die beiden Autoren unterscheiden prä- und postiktische Tonbrüche, also Grundfrequenz- bzw. Tonhöhen-Sprünge, die grob gesprochen am Beginn bzw. Ende einer betonten Silbe erfolgen. Ein präiktischer Anstieg in der letzten

betonten Silbe entspricht dabei einem progredienten Tonverlauf, ein postiktischer einem fragemarkierenden (Isačenko/Schädlich (1971:46ff.)). Das genaue 'Timing' der Grundfrequenz- bzw. Tonhöhen-Veränderungen in der letzten 'Hauptakzentsilbe' einer intonatorischen Phrase kann also durchaus mit zur Charakterisierung einer solchen Phrase als satzmodusrelevante letzte beitragen.

Ob die intonatorische Kennzeichnung allein ausreicht, scheint aber eher zweifelhaft, da sich die feinen 'Timing'-Unterschiede auch bei der eigentlichen Satzmoduskennzeichnung als nicht sehr zuverlässig erweisen. Es ist also zu erwarten, daß die intonatorische Form auch bei der Auszeichnung satzmodusrelevanter intonatorischer Phrasen auf das Zusammenspiel mit den im engeren Sinn syntaktischen Strukturen angewiesen ist.

5. Intonatorische Prototypen

Im folgenden werden die intonatorischen Prototypen der Minimalpaarglieder jeweils einander gegenübergestellt - soweit nicht bei einem der beiden Glieder zu wenig Realisationen durch den Filter kamen, um eine zuverlässige Aussage über den Prototyp zu gestatten (die Numerierung verweist auf die entsprechenden Spalten im Anhang).

(1) Das erste Paar bilden der **Aussagesatz** und die **assertive Frage**. Beide sind durch die Offsethöhe und die Kontur im Bereich der A1-Silbe eindeutig voneinander unterschieden: Zu den Aussagesätzen (vgl. auch die Charakterisierung der Intonation von Aussagesätzen in Bannert (1983), (1985)) gehören ein tiefer Offset (100%) und eine konvexe Kontur (zu 70% die 'Feinkontur' 231), zu den assertiven Fragen ein hoher Offset (100%) und eine konkave Kontur (zu 70% 213).

(2) Als zweites werden **Aussagesätze** und **Verb-Zweit-Exklamativsätze** verglichen. Die Aussagesätze enthalten dabei aufgrund der Akzentuierungsparallelität einen Kontrastakzent auf dem Subjektsausdruck (vgl. die Erläuterungen weiter oben bei der Einführung der Minimalpaarmethode); sie sind also stark markiert.

Die Folge war, daß keine einzige der 'Kontrastaussage'-Realisationen von mehr als 80% der VPen im Kategorisierungstest als Aussage identifiziert wurde, daß also alle durch unseren Filter ausgesondert wurden - es gab 62% Kategorisierungen als 'Ausruf'. Diese Ergebnisse sind erklärlich, wenn man berücksichtigt, daß bei den Kategorisierungstests kontextfrei identifiziert werden mußte, eine Kontrastierung aber nur vor dem Hintergrund eines bestimmten Kontextes sinnvoll ist. Zusätzlich ist zu bedenken, daß die von uns vorgegebene Kategorie 'Ausruf' nicht exakt festgelegt wurde, so daß bei manchen VPen eine Tendenz bestanden haben könnte, alle in irgendeiner Dimension markierten Äußerungen als Ausrufe zu etikettieren, wenn keine der anderen Kategorien eindeutig besser paßte. Jedenfalls zeigt das Ergebnis der Hörtests, daß eine eventuell vorhandene eindeutige formale Auszeichnung von Kontrastakzenten gegenüber Exklamativakzenten nicht ausreicht, um bei einer kontextfreien Identifizierung die an sich naheliegendere Interpretation als Exklamativ - bzw. Ausruf - zu unterdrücken.

Um trotz des Fehlens von 'zulässigen' Kontrastaussage-Äußerungen auch hier miteinander vergleichbare intonatorische Prototypen konstruieren zu können, wurden diejenigen Realisationen berücksichtigt, die wenigsten einen kleineren Natürlichkeitswert als 2.5 haben, die also der einen Filterbedingung entsprachen. Beide Minimalpaar-Glieder, Aussagesatz mit Kontrastakzent und Verb-Zweit-Exklamativsatz, sind durch einen tiefen Offset und eine konvexe Kontur gekennzeichnet. Einzig auffällig ist, daß bei den Exklamativsatz-Realisationen relativ häufig ein 132-Verlauf auftritt (40%), also ein Verlauf, bei dem der F_0 -Gipfel gegenüber den anderen beiden konvexen Verläufen (231 und 121) 'später' erreicht wird (vgl. dazu auch Batliner (1988a)). Unsere Ergebnisse legen jedoch den Schluß nahe, daß dieses Merkmal zu instabil ist, um die beiden Satzmodi in der Mehrzahl der Fälle hinreichend genau unterscheiden zu können. Auch die Unterschiede im durchschnittlichen Range sind zu wenig ausgeprägt, um eine genügend eindeutige Trennung der Satzmodi zu gewährleisten - sowohl der Gesamt- als auch der Kontur-Range ist bei der Kontrastaussage etwas größer (Gesamtrange 13 vs. 10 Halbtöne und Konturrange 10 vs. 8 Halbtöne - die Kontrastaussage ist jeweils

zuerst genannt; ein großer Range auf der kontrastierten Konstituente wird in Bannert (1985:300) als Kennzeichen für den Kontrast angegeben).

Dagegen könnten die Dauerunterschiede von Bedeutung sein. Die vorhandenen Werte (es sind relativ wenige, so daß die folgenden Bemerkungen einer weiteren experimentellen Absicherung bedürfen) lassen sich so interpretieren, daß der Exklamativsatz durch größere Dauer ausgezeichnet ist. Zwei Strategien der Längung sind jedoch möglich: Bei den weiblichen VPen haben die Exklamativsatz-Realisationen eine deutlich längere Gesamtdauer als die Aussagesatz-Realisationen, während bei den Männern die Gesamtdauer ungefähr gleich ist; dafür ist bei den männlichen VPen die relative Dauer der A1-Silbe bei den Exklamativsätzen länger - bei den weiblichen VPen war hier kein Vergleich möglich, weil die A1-Silbe bei den Exklamativ- und Aussagesatz-Realisationen häufig nicht identisch war (vgl. das oben zu den Ergebnissen der Akzenttests Gesagte).

Die genauen Werte verteilen sich folgendermaßen: Bei den weiblichen VPen war die Exklamativsatz-Realisation von *Die ist naiv* durchschnittlich 1100 msec lang, die Aussagesatz-Realisation 870 msec; mit eingefügtem *vielleicht* war das Verhältnis 1240 zu 980 msec. Beidemale sind die Exklamativsatz-Versionen im Durchschnitt um 25% länger. Bei den männlichen VPen war das Verhältnis der Dauer des betonten *der* zur Durchschnittssilbendauer in *Der hat vielleicht geflucht* .75 im Aussagesatz zu .92 im Exklamativsatz, also eine 'Erhöhung' der relativen Dauer der A1-Silbe im Exklamativsatz um mehr als 20%.

Bei den restlichen Testsätzen reicht die Anzahl der auf gleichen Silben betonten Exklamativsatz- und Aussagesatz-Äußerungen für einen sinnvollen Vergleich nicht aus.

Die vorhandenen Daten zeigen also, daß die 'Kontrastaussage' und der Verb-Zweit-Exklamativsatz durch die ansonsten zentralen Merkmale Offsethöhe und Kontur nicht hinreichend voneinander unterschieden werden können. Offensichtlich müssen andere Mittel eingesetzt werden, die aber nicht sehr 'indizierungsstark' sind, soweit sich das aus den Kategorisierungstests erschließen läßt. Die ursprüngliche Hypothese, daß Exklamativstrukturen durch 'Extensionsphänomene' gekennzeichnet werden, scheint sich für das

Dauermerkmal zu bestätigen, nicht unbedingt jedoch für den F_0 -Range. Hier ist jedoch zum einen zu berücksichtigen, daß der Range nur gegenüber den Kontrastaussagen etwas kleiner ausfällt, ansonsten aber den des 'normalen' Aussagesatzes im ersten Minimalpaar übersteigt, und daß zweitens die 'Exklamativität' offensichtlich nicht an ganz bestimmten Stellen signalisiert werden muß: Nicht auf einer Fokussilbe, da anscheinend keine übliche Fokussierungsstruktur vorhanden ist (vgl. die Bemerkungen oben), und möglicherweise nicht einmal im eigentlichen Exklamativsatz, wenn diesem ein 'Prä-Exklamativ' vorangeht - das war bei einem der Testsätze der Fall: *Mann! Der hat (vielleicht) geflucht!* Auch durch die Feinstruktur der A1-Silbe kann möglicherweise Exklamativität angezeigt werden. Ein später F_0 -Gipfel in dieser Silbe scheint eher zu einer 'exklamativen' Interpretation der Gesamtäußerung zu führen als ein früher (vgl. Kohler (1987)); dem entspricht wohl auch in etwa das in manchen tonalen Ansätzen verwendete Merkmal [delayed peak], vgl. Ladd (1983a:42ff.) und Ladd (1983b:730ff.), sowie Gussenhoven (1985); interessanterweise gibt es bei diesen anscheinend kein Merkmalsgegenstück zu Kohlers frühem Gipfel). Auf den Zusammenhang von Exklamativstruktur und spätem Gipfel weist übrigens auch die oben erwähnte Tatsache hin, daß die 132-Kontur gehäuft auftrat. Die vorläufigen Ergebnisse eines zweiten Auswertungsdurchgangs, bei dem die genaue Lokalisation des F_0 - und Intensitäts-Gipfels in der A1-Silbe festgestellt wurde, zeigen, daß bei drei der sechs Sprecher/innen das A1-Intensitätsmaximum der Exklamativsatz-Realisationen vor dem A1- F_0 -Maximum liegt; für die Aussagesatz-Realisationen gilt das nicht. Bei den restlichen drei VPen ist allerdings keine einheitliche Tendenz festzustellen. Auch diese Ergebnisse sind mit der Annahme eines späten F_0 -Gipfels in der 'Hauptakzentsilbe' von Exklamativsätzen vereinbar, machen aber gleichzeitig deutlich, daß es sich wohl nicht um ein stabiles Merkmal handelt. Tests mit manipuliertem Material (Testsatz: *Der Leo säuft*), bei dem die F_0 -Gipfelposition gezielt variiert wurde, legen ebenfalls nahe, daß diese kein zentrales Merkmal ist (vgl. Batliner (1988a)).

(3/4) Eine Gegenüberstellung der imperativischen Satzmodi **Adhortativsatz** und **Sie-Imperativsatz** und der entsprechenden **Entschi-**

dungsfragesätze führt zu folgenden Prototypen: Die Imperativsatz-Realisationen haben stets einen tiefen Offset und eine konvexe oder einfach fallende Kontur, die Entscheidungsfragesatz-Realisationen dagegen immer einen hohen Offset und eine konkave oder einfach steigende Kontur.

Alle durch den Filter gelangten Entscheidungsfragesatz-Realisationen haben einen hohen Offset und eine konkave oder einfach steigende Kontur. Das heißt natürlich nicht, daß es nur diesen einen Prototyp für Entscheidungsfragesätze gibt. Ganz offensichtlich muß man auch einen Prototyp mit tiefem Offset und konvexer Kontur annehmen (vgl. Klein (1982); Studdert-Kennedy/Hadding (1973)). Daß dieser Typ bei unseren 'guten' Realisationen nicht auftritt, liegt an der Konstruktion der 'satzmodusambigen' Testsätze, die nicht die beliebige Verwendung eines der beiden Prototypen zulassen, sondern nur die des allein bei Fragesätzen möglichen.⁸

(5) Sehr interessant im Hinblick auf einen Beschreibungsansatz, der neben dem F_0 -Verlauf einer Äußerung auch die Zeitstruktur und die Feinstruktur der 'Hauptakzentsilbe(n)' berücksichtigt, ist das Minimaltripel aus Entscheidungsfragesatz, Imperativsatz mit 'Kontrast' und Verb-Erst-Exklamativsatz das auf der Wortfolge *Stellt ihr euch an* basiert. Der Entscheidungsfragesatz ist von den beiden anderen deutlich durch den hohen Offset und die konkave Kontur unterschieden.

Beim Minimalpaar Imperativsatz mit 'Kontrast' und Verb-Erst-Exklamativsatz ist dagegen nicht so offensichtlich, welcher Signalparameter für die ziemlich gute kontextfreie Identifizierbarkeit verantwortlich ist. Immerhin gelangten 46% der Imperativ-

⁸ Selbst der in den Experimenten von Studdert-Kennedy/Hadding als Frageindikator entdeckte extrem hohe F_0 -Gipfel reicht nämlich anscheinend für die Abgrenzung gegenüber dem Nachbarsatzmodus nicht aus (die beiden Prototypen sind 'zu nahe' beieinander und überdecken sich damit weitgehend): Von einem der Projektmitarbeiter existieren ebenfalls Realisationen der Testsätze, u.a. solche von Entscheidungsfragesätzen mit der uneindeutigen intonatorischen Form; und obwohl diese eine ganz typische konvexe Kontur mit steilem und großem F_0 -Anstieg und -Abfall enthalten, werden sie in den Kategorisierungstests extrem schlecht erkannt.

satz-Realisationen und 31% der Exklamativsatz-Realisationen durch den Filter.

Der Offset ist beidemale tief, die Kontur konvex. Wie beim Verb-Zweit-Exklamativsatz sind zwar auch hier relativ viele 132-Verläufe (40%) bei den Exklamativsatz-Äußerungen vertreten, aber in den restlichen Fällen ist eine Abgrenzung gegenüber dem Imperativsatz durch die Kontur selbst nicht gewährleistet. Im realisierten Range (Gesamtrange: beidemale 12 Halbtöne; Kontur-range: beidemale 9 Halbtöne) und in den Dauerverhältnissen (Gesamtdauer der modalpartikellosen Versionen: bei den Imperativsatzversionen 900 msec, bei den Exklamativversionen 950 msec; relative Dauer der A1-Silbe: beidemale .7) unterscheiden sich die beiden Gruppen auch nicht.

Bei einer Feinanalyse der A1-Silbe zeigt sich jedoch, daß die genaue Lokalisation des F_0 - und des Intensitäts-Gipfels in dieser Silbe für eine Differenzierung relevant sein kann. Beim Imperativsatz mit kontrastiertem Subjektspronomen liegt typischerweise der F_0 -Gipfel verhältnismäßig früh in der Silbe und auf keinen Fall hinter dem Intensitäts-Gipfel. Beim Exklamativsatz fallen beide annähernd zusammen, bzw. der F_0 -Gipfel liegt etwas hinter dem Intensitäts-Gipfel und auf jeden Fall verhältnismäßig spät in der Silbe. Als weiteres Merkmal kommt hinzu, daß die Intensität beim Exklamativsatz wesentlich schneller abfällt als beim Imperativsatz. Allerdings ist dieses Muster nur bei drei der sechs VPen einigermaßen konsistent anzutreffen. Es 'paßt' aber auf jeden Fall in seiner Tendenz zu der beim Verb-Zweit-Exklamativsatz beobachteten intonatorischen Form (auch hier ist eine Absicherung durch Experimente mit gezielt manipuliertem Material nötig).

Zu beachten ist jedenfalls, daß in unseren (kontextfreien) Kategorisierungstests der Funktionstyp relativ gut identifiziert wurde, obwohl sich die VPen auch in diesem Fall auf sehr feine intonatorische Unterschiede stützen müssen.

(6) Den **Ergänzungsfragesätzen** sind im Korpus **w-Verb-Zweit-Exklamativsätze** gegenübergestellt. Allerdings gelangten von den Realisationen des letzteren Satztyps nur so wenig Exemplare durch den Filter, daß eine vernünftige Prototyp-Bildung mit den wenigen zur Verfügung stehenden Daten nicht möglich war.

Der Ergänzungsfragesatz ist der einzige Typ, bei dem die Offsethöhe der 'guten' Exemplare variiert - sie ist zu 80% tief; ebenso variiert hier auch die Kontur.

(7) Der '**und-Ergänzungsfragesatz**' (bei dem durch die Kontextvorgabe alle Ausdrücke außer dem Frageausdruck als gegebener 'Hintergrund' festgelegt sind) und der **Versicherungsfragesatz** unterscheiden sich bei den Realisationen unseres Korpus' - außer durch das einleitende *und* - durch die Kontur, während der Offset beidemale hoch ist. Das betonte *wer* des Versicherungsfragesatzes weist gegenüber der konkaven Kontur des *und-Ergänzungsfragesatzes* eine umfangreichere F_0 -Bewegung auf, und zwar ein konvexe.

(8) **Versicherungsfragesätze** werden auch **Aussagesätzen** mit der Kurzform *was* des Indefinitpronomens gegenübergestellt; auf die prinzipiell ungleichartige Akzentstruktur wurde bereits hingewiesen. Die Realisationen des Versicherungsfragesatzes haben wieder zu 100% einen hohen Offset und diesmal eine konkave Kontur (vgl. auch die Charakterisierung in Wunderlich (1986:51ff.)); da das Fragewort einsilbig ist und am Ende der Äußerung steht, sind offenbar die Bedingungen für eine Realisierung der beim anderen Versicherungsfragesatz anzutreffenden konvexen Kontur nicht gegeben.

Beim Aussagesatz ist der Offset wieder tief und die Kontur konvex oder einfach fallend.

Von den **Wunschsätzen** und den wortfolgengleichen **Verb-Erst-Exklamativsätzen** kamen zu wenige durch den Kategorisierungs- und Natürlichkeitsfilter, so daß eine Prototypbildung nicht möglich war.

(9) Als letztes Minimalpaar werden zwei Verb-Letzt-Strukturen einander gegenübergestellt, die **Wortfrage-Übernahme** und der **w-Verb-Letzt-Exklamativsatz**. Die Frageexemplare sind durch einen hohen Offset und eine konkave Kontur gekennzeichnet, die Exklamativsatz-Realisationen dagegen durch tiefen Offset und konvexe Kontur. Interessanterweise ist auch hier der Exklamativsatz durch eine wesentlich größere relative Dauer der A_1 -Silbe ausgezeichnet

net, sie ist im Durchschnitt um 25% länger als bei den Wortfrage-Übernahmen (.8 bei der Frageversion vs. 1.0 bei der Exklamativversion).

Die intonatorischen Prototypen können also im wesentlichen auf den beiden Merkmalen der Offsethöhe und der Form des F_0 -Verlaufs im Bereich der A1-Silbe aufgebaut werden. Die beiden Merkmale sind aber zumindest im Bereich der Exklamativstrukturen noch nicht hinreichend für eine eindeutige Kennzeichnung. Hier werden ansonsten periphere Merkmale wichtig, wie die Zeitstruktur oder die Feinstruktur der A1-Silbe.

Die Menge der satzmodusrelevanten intonatorischen Merkmale liegt noch nicht fest. Es ist mit der Aufnahme von weiteren Merkmalen zu rechnen, obwohl Offsethöhe und Kontur sicherlich trotz aller Revisionen die im Normalfall dominanten Merkmale bleiben werden. Von Bedeutung könnten z.B. die verschiedenen Deklinationslinien sein, also die 'Toplines' durch nach bestimmten Kriterien (z.B. annähernd gleiche Prominenz der Silbenträger) ausgewählte F_0 -Maxima der Äußerungen oder die 'Bottomlines' durch entsprechende Minima. Im Dänischen ist offensichtlich die Neigung der durch Akzentsilben gezogenen Bottomline satzmodusunterscheidend - ein abfallender Verlauf dieser Linie kennzeichnet Aussagesätze, ein ebener intonatorisch markierte Fragesätze (vgl. Thorsen (1979:74f.); Thorsen (1983:192)). Der Deklinationsverlauf scheint aber auch in weiteren Sprachen mit verschiedenen Satzmodi korreliert zu sein (vgl. Vaissière (1983:57)). Im Deutschen wären satzmodustypische Deklinationslinien gegenüber den dominanten Merkmalen Offsethöhe und Kontur allerdings sicherlich von weniger zentraler Bedeutung. Eine Untersuchung der Deklinationslinien schien uns am gegenwärtigen Korpus wegen der Kürze der Äußerungen nicht sehr lohnend, zudem ermöglicht die angewandte Methode der Datenextraktion keine exakte zeitliche Lokalisierung der relevanten F_0 -Extremwerte. Dagegen wurden an einem weiteren Korpus, das auch verhältnismäßig lange Testsatz-Realisationen enthält, je Äußerung zahlreiche zeitlich genau lokalisierte F_0 -Extremwerte extrahiert. Eine Berechnung entsprechender Deklinationslinien ergab, daß tatsächlich Fragesatz-Realisationen durch steigende Deklinationsverläufe, Aussage- und Imperativsatzrealisationen

dagegen durch abfallende Verläufe gekennzeichnet werden; in wenigen Fällen sorgt dabei die Deklination allein für die Satzmoduskennzeichnung (vgl. Oppenrieder (1987)).

6. Schlußbemerkung

Die Aufstellung intonatorischer Prototypen auf der Grundlage eines großen Korpus' von Realisationen gezielt konstruierter Testsätze ist ein erster Schritt bei der Untersuchung der satzmoduskennzeichnenden intonatorischen Merkmale. Ihrer 'Entdeckung' muß die Bestätigung ihrer Relevanz und die exaktere Ausarbeitung der typischen Ausprägungen der einzelnen Merkmale folgen. Dies geschieht durch die Konstruktion und das Austesten von manipulierten Stimuli, die ganz gezielt nach der Vorgabe der in einer ersten Annäherung gewonnenen intonatorischen Prototypen variiert werden (vgl. für solche weiterführende Untersuchungen im Bereich von Aussage- und Verb-Zweit-Exklamativsatz Batliner (1988a)). Erst auf dieser Grundlage ist es dann auch möglich, den Beitrag eines Parameters beim Aufbau des intonatorischen Merkmalskomplexes genau abzuschätzen (und eventuell sogar so exakt zu gewichten, wie es z.B. bei Oden/Massaro (1978) mit den Einzelmerkmalen geschieht, die bei der Wahrnehmung eines Sprachlautes in einen Merkmalskomplex integriert werden).

LITERATUR

- Adriaens, L.M.H. 1984. A Preliminary Description of German Intonation. IPO Annual Progress Report 19. 36-41.
- Altmann, H. 1984. Linguistische Aspekte der Intonation am Beispiel Satzmodus. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 19. 132-152.
- Altmann, H. 1987. Zur Problematik der Konstitution von Satzmodi als Formtypen. Meibauer, J. (Hg.) 1987. Satzmodus zwischen Grammatik und Pragmatik. Tübingen. 22-56.
- Bannert, R. 1983. Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Linguistik 49. 9-34.

- Bannert, R. 1985. Fokus, Kontrast und Phrasenintonation im Deutschen. Zeitschrift für Dialektologie und Linguistik 52. 289-305.
- Batliner, A. 1987. Der Einsatz der Diskriminanzanalyse zur Prädiktion des Satzmodus. Tillmann, H.G./Willée, G. (Hgg.) 1987. Analyse und Synthese gesprochener Sprache. Hildesheim etc. 125-132.
- Batliner, A. 1988a. Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des F₀-Gipfels. In diesem Band.
- Batliner, A. 1988b. Modus und Fokus als Dimensionen einer Nonmetrischen Multidimensionalen Skalierung. In diesem Band.
- Batliner, A. 1988c. Produktion und Prädiktion. Die Rolle intonatorischer und anderer Merkmale bei der Bestimmung des Satzmodus. In diesem Band.
- Bierwisch, M. 1971. Regeln für die Intonation deutscher Sätze. Studia Grammatica VII. Untersuchungen über Akzent und Intonation im Deutschen. 21971. Berlin. 99-201.
- Boves, L./ten Have, B.L./Vieregge, W.H. 1984. Automatic Transcription of Intonation in Dutch. Gibbon, D./Richter, H. (eds.) 1984. Intonation, Accent and Rhythm. Berlin/New York. 20-45.
- Brown, G./Currie, K.L./Kenworthy, J. 1980. Questions of Intonation. London.
- Cooper, W.E./Sorensen, J.M. 1981. Fundamental Frequency in Sentence Production. New York/Heidelberg/Berlin.
- Cooper, W.E./Eady, S.E./Mueller, P.R. 1985. Acoustical Aspects of Contrastive Stress in Question-Answer Contexts. Journal of the Acoustical Society of America 77. 2142-2156.
- Eady, S.J./Cooper, W.E./Klouda, G.V./Mueller, P.R./Lotts, D.W. 1986. Acoustical Characteristics of Sentential Focus: Narrow vs. Broad and Single vs. Dual Focus Environments. Language and Speech 29. 233-251.
- v. Essen, O. 21964. Grundzüge der hochdeutschen Satzintonation. Ratingen/Düsseldorf.
- Gussenhoven, C. 1985. Intonation: a Whole Autosegmental Language. Hulst, H.v.d./Smith, N. (eds.) 1985. Advances in Nonlinear Phonology. Dordrecht/Cinnaminson. 117-131.
- Halliday, M.A.K. 1976. Intonation and Meaning. Kress, G.R. (ed.) 1976. Halliday: System and Function in Language. Selected Papers. London. 214-234.
- 't Hart, J./Collier, R. 1975. Integrating Different Levels of Intonation Analysis. Journal of Phonetics 3. 235-255.
- Isačenko, A.V./Schädlich, H.J. 1971. Untersuchungen über die deutsche Satzintonation. Studia Grammatica VII. Untersuchungen über Akzent und Intonation im Deutschen. 21971. Berlin. 7-67.
- Klein, W. 1982. Einige Bemerkungen zur Frageintonation. Deutsche Sprache 10. 289-310.

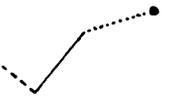
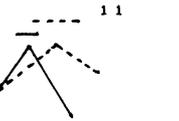
- Kohler, K.J. 1987. Funktionen von F_0 -Gipfeln im Deutschen. Tillmann, H.G./Willée, G. (Hgg.) 1987. Analyse und Synthese gesprochener Sprache. Hildesheim etc. 133-139.
- Ladd, D.R. 1983a. Peak Features and Overall Slope. Cutler, A./Ladd, D.R. (eds.) 1983. Prosody: Models and Measurements. Berlin etc. 39-52.
- Ladd, D.R. 1983b. Phonological Features of Intonational Peaks. Language 59. 721-759.
- Näf, A. 1987. Gibt es Exklamativsätze? Meibauer, J. (Hg.) 1987. Satzmodus zwischen Grammatik und Pragmatik. Tübingen. 140-160.
- Nöth, E./Batliner, A./Lang, R./Oppenrieder, W. 1987. Automatische Grundfrequenzanalysen und Satzmodusdifferenzierung. Tillmann, H.G./Willée, G. (Hgg.) 1987. Analyse und Synthese gesprochener Sprache. Hildesheim etc. 59-66.
- Oden, G.C./Massaro, D.W. 1978. Integration of Featural Information in Speech Perception. Psychological Review 85. 172-191.
- Oppenrieder, W. 1987. Deklination und Satzmodus. Ms. München.
- Oppenrieder, W. 1988. Intonation und Identifikation. Kategorisierungstests zur kontextfreien Identifikation von Satzmodi. In diesem Band.
- Pheby, J. 1980. Intonation und Grammatik im Deutschen. Berlin.
- Pheby, J. 1981. Phonologie: Intonation. Heidolph, K.E./Flämig, W./Motsch, W. 1981. Grundzüge einer deutschen Grammatik. Berlin. 839-897.
- Pierrehumbert, J.B. 1980. The Phonology and Phonetics of English Intonation. Phil. Diss. MIT.
- Sadock, J.M./Zwicky, A.M. 1985. Speech Act Distinctions in Syntax. Shopen, T. (ed.) 1985. Language Typology and Syntactic Description. Volume I: Clause Structure. Cambridge etc. 155-196.
- Studdert-Kennedy, M./Hadding, K. 1973. Auditory and Linguistic Processes in the Perception of Intonation Contours. Language and Speech 16. 293-313.
- Thorsen, N. 1979. Interpreting Raw Fundamental-Frequency Tracings of Danish. Phonetica 36. 57-78.
- Thorsen, N. 1983. Standard Danish Sentence Intonation - Phonetic Data and Their Representation. Folia Linguistica 17. 187-220.
- Tropf, H.S. 1985. Zur Intonation spontan gesprochener und laut gelesener W-Fragen. Kürschner, W./Vogt, R. (Hgg.) 1985. Grammatik, Semantik, Textlinguistik. Akten des 19. Linguistischen Kolloquiums Vechta 1984, Band 1. Tübingen. 49-60.
- Vaissière, J. 1983. Language-Independent Prosodic Features. Cutler, A./Ladd, D.R. (eds.) 1983. Prosody: Models and Measurements. Berlin etc. 53-66.
- Wunderlich, D. 1986. Echofragen. Studium Linguistik 20. 44-62.

ANHANG: Liste der intonatorischen Prototypen⁹

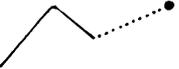
	Satztyp/ Beispielsatz	Realisie- rungen: igs. F.e.	Off- set	Kon- tur	prototypi- scher F ₀ -Verlauf
1a	Aussagesatz <i>Du kommst.</i>	12 40%	tief 100%	231 70%	
1b	assertive Frage <i>Du kommst?</i>	12 93%	hoch 100%	213 70%	
2a	Verb-Zweit- Exklamativsatz <i>Der hat (vielleicht) geflucht!/Die ist (vielleicht) naiv!</i>	53 38%	tief 100%	231 35% 132 40%	
2b	Aussagesatz mit 'Kontrast' <i>Der hat (vielleicht) geflucht./Die ist (vielleicht) naiv.</i>	56 55% ¹⁰	tief 100%	231 75%	
3a	Adhortativsatz <i>Gehen wir (doch jetzt)!</i>	24 58%	tief 100%	231 45% 121 30%	

9 Nach dem Formtyp und den verwendeten Testsätzen wird zuerst die Zahl der Realisationen insgesamt (igs.) angegeben. Die folgende Prozentangabe bezieht sich auf die Realisationen, die den Natürlichkeits- und Kategorisierungs-Filter erfüllten (F.e.). Die weiteren Prozentangaben basieren ausschließlich auf diesen 'guten' Exemplaren. Als prototypischer F₀-Verlauf ist die F₀-Bewegung im Bereich der A1-Silbe und die Offsethöhe eingezeichnet.

10 55% Prozent der Realisationen hatten einen durchschnittlichen Natürlichkeitswert < 2.5; beide Filterbedingungen wurden von keiner Realisation erfüllt.

3b	Entscheidungs- fragesatz <i>Gehen wir (denn jetzt)?</i>	25	48%	hoch 100%	konkav 60%	
4a	Sie-Imperativsatz <i>Schlafen Sie!</i>	12	50%	tief 100%	konvex 50%	
4b	Entscheidungs- fragesatz <i>Schlafen Sie?</i>	14	64%	hoch 100%	konkav 75%	
5a	Imperativsatz mit 'Kontrast' <i>Stellt ihr euch (doch) an!</i>	26	46%	tief 100%	231 65% 121 25%	
5b	Verb-Erst- Exklamativsatz <i>Stellt ihr euch (vielleicht) an!</i>	26	31%	tief 100%	132 40% 231 30%	
5c	Entscheidungs- fragesatz <i>Stellt ihr euch (etwa) an?</i>	27	48%	hoch 100%	213 65%	
6	Ergänzungs- fragesatz <i>Wie alt ist er geworden?/ Wie laut ist es hier?/ Wie ist der (denn) reich geworden?</i>	56	57%	tief 80%	132 50%	

- 11 Der in 5a und 5b (mit durchgezogenen Linien) dargestellte F_0 -Verlauf bezieht sich nur auf die A1-Silbe; gestrichelt ist der Intensitätsverlauf auf dieser Silbe hinzugefügt. Die horizontalen Striche geben den Varianzbereich an, innerhalb dessen die F_0 - und Intensitätsgipfel liegen können.

7a	'und-Ergänzungs- fragesatz' <i>Und wer spielt?</i>	13	62%	hoch 100%	213 100%	
7b	Versicherungs- fragesatz <i>WER spielt?</i>	12	100%	hoch 100%	132 60%	
8a	Versicherungs- fragesatz <i>Er sieht WAS?</i>	12	75%	hoch 100%	213 90%	
8b	Aussagesatz <i>Er sieht was.</i>	12	50%	tief 100%	231 50%	
9a	Wortfrage- Übernahme <i>Wie groß der ist?</i>	15	67%	hoch 100%	213 80%	
9b	w-Verb-Letzt- Exklamativsatz <i>Wie groß der ist!</i>	13	46%	tief 100%	231 50%	

PRODUKTION UND PRÄDIKTION. DIE ROLLE INTONATORISCHER UND ANDERER MERKMALE BEI DER BESTIMMUNG DES SATZMODUS.

Anton Batliner (München)

1. EINLEITUNG

Der vorliegende Beitrag ist eine erweiterte und modifizierte Version von Batliner (1987a). Er entstand im Rahmen des DFG-Projekts "Modus-Fokus-Intonation", das eine genauere Charakterisierung der intonatorischen Merkmale von Satzmodus und Fokus zum Ziel hat. Grundlage ist das von Hans Altmann entwickelte Satzmodussystem (vgl. Altmann 1987). Das Testsatzkorpus, die Extraktion der intonatorischen Parameterwerte sowie die daran anschließenden Hörtests werden im nächsten Abschnitt nur kurz charakterisiert, da sie genauer in Oppenrieder (1988) beschrieben sind.

2. SATZMODUSSYSTEM UND TESTSATZKORPUS

Unter einem **Satzmodus** wird die eindeutige Zuordnung eines **Funktionstyps**, z.B. der Ergänzungsfrage, und eines genau festgelegten **Formtyps** verstanden. Die einzelnen Formtypen können als **strukturierte Merkmalbündel** beschrieben werden, die sich in mindestens einem Merkmal voneinander unterscheiden. Nur solche Merkmale werden als satzmodusrelevant angesehen, die zumindest einmal für die Unterscheidung zweier Formtypen unverzichtbar sind. Die Merkmale stammen aus vier Mengen. Sie betreffen

- (a) das Vorhandensein von Ausdrücken bestimmter **Kategorien**, beim Formtyp W-Fragesatz z.B. das Vorhandensein irgendeines W-Ausdrucks;
- (b) die **Stellungseigenschaften** dieser satzmodusrelevanten Ausdrücke, z.B. die Zweitstellung des finiten Verbs;
- (c) die indikativische, konjunktivische oder imperativische **Markierung des finiten Verbs**
- (d) die **intonatorische Markierung**.

Wenn intonatorische Merkmale, die anders als die Merkmale aus den Bereichen (a) bis (c) in jeder Äußerung vorhanden sind, satzmodusrelevant sein sollen, müssen sie mindestens an einer Stelle für die Differenzierung zweier ansonsten

identischer Formtypen sorgen können. Genau an den Satzmodi, die sich formal nicht immer durch Merkmale aus den Bereichen (a) bis (c) unterscheiden lassen, können demnach die satzmodusrelevanten intonatorischen Parameter(-werte) aufgefunden und untersucht werden. Es wurden für alle jene Formtypen, die nur intonatorisch unterscheidbar sind, also für intonatorische Minimalpaare, Testsätze gebildet und in Kontexte eingebettet, die jeweils nur mit einem der Formtypen verträglich sind. Die Testsätze wurden zusammen mit ihrem modussteuernden Kontext sechs Versuchspersonen (drei weiblichen, drei männlichen) vorgelegt und von diesen mindestens zweimal realisiert; dabei ergaben sich 956 Äußerungen. Von jeder Äußerung wurde ein Mingogramm mit Zeitsignal (Oszillogramm), Fo-Kurve und Intensitätsverlauf erstellt, an dem die relevanten intonatorischen Parameterwerte wie Fo-Onset, Fo-Offset, Fo-Maximum, Fo-Minimum, Maximum der Intensität, Dauer der am stärksten akzentuierten Silbe etc. abgelesen wurden; diese Werte bildeten die Grundlage für die Berechnung weiterer Werte mit dem Statistikpaket SPSSPC+, wie des Tonumfangs ('Range') jeder Äußerung, der halbtonttransformierten Werte etc. Alle Realisationen wurden Hörtests unterzogen, bei denen im Schnitt 12 Versuchspersonen die Testsatzrealisationen u.a. (1) auf ihre Natürlichkeit im Kontext hin beurteilen und sie (2) kontextfrei Moduskategorien zuweisen sollten. Ziel war dabei zuerst eine **deskriptive** Statistik der intonatorischen Parameter, die gegebenenfalls prüfstatisch überprüft wird. Ergänzend zu diesem Verfahren wurde aber auch ein geeignetes **klassifizierendes** Verfahren, die sog. Diskriminanzanalyse, auf das Korpus angewandt. Es sollte damit untersucht werden, inwiefern automatische Verfahren bei Untersuchungen zur Intonation sinnvoll eingesetzt werden können. In unserem Fall handelt es sich zunächst um teilweise automatische Verfahren, da die Parameterwerte von Hand extrahiert wurden. Die Relevanz solcher Verfahren für die automatische Spracherkennung versteht sich von selbst; im Rahmen der Grundlagenforschung bietet sie eine sinnvolle Korrektur für das notwendigerweise beschränkte Material, mit dem Linguisten und Phonetiker üblicherweise arbeiten. Mit ihrer Hilfe wäre es möglich, große Korpora weitgehend automatisch zu bearbeiten und damit die Beschreibung zunehmend zu verbessern.

3. BESCHREIBUNG DES VERFAHRENS

Die Diskriminanzanalyse ist ein Verfahren, bei dem, basierend auf relevanten Variablen (in unserem Fall Fo-Onset, Fo-Offset etc.), die einzelnen Fälle (die Äußerungen unseres Korpus') distinkten Kategorien (den intendierten Satzmodi) zugewiesen werden. Das Verfahren funktioniert ähnlich wie die multiple Regression: Es werden lineare Kombinationen der unabhängigen 'Prädiktor'-Varia-

blen gebildet, die möglichst optimal zwischen den Kategorien unterscheiden können. Jede Prädiktorvariable erhält einen Gewichtungskoeffizienten, der anhand der vorgegebenen Daten so geschätzt wird, daß die resultierende Diskriminanzfunktion zwischen den Gruppen so stark wie möglich differiert; anders gesagt, die Variabilität **zwischen** den Gruppen soll im Verhältnis zur Variabilität **innerhalb** der Gruppen möglichst groß sein. Prädiktorvariablen sollten normalverteilte kontinuierliche Werte aufweisen, d.h. Fo-Werte und davon abgeleitete Werte sind auf alle Fälle gute Kandidaten. (Das Verfahren funktioniert aber auch oft recht gut bei dichotomen Variablen.) Es kann das gesamte Material Grundlage der Analyse sein und dann klassifiziert werden, es kann aber auch nur ein Ausschnitt aus dem Material (die 'Lernstichprobe') analysiert und Grundlage der Klassifikation eines anderen Ausschnitts (der 'Prüfstichprobe') werden. Bevor wir im nächsten Abschnitt unser spezielles Vorgehen beschreiben, wollen wir auf einige allgemeine Punkte eingehen:

Erwartungswert: Maßstab für die Güte der Prädiktion ist der Prozentwert der Fälle, die mit den jeweils angenommenen Prädiktorvariablen richtig klassifiziert werden, d.h. in unserem Fall, einem der im Korpus vorhandenen fünf Haupt-Modi **Frage (-satz)**, **Aussage (-satz)**, **Exklamativ (-satz)**, **Imperativ (-satz)** oder **Wunsch (-satz)** richtig zugeordnet werden. Dieser Wert muß immer in Relation gesetzt werden zu einer zufällig richtigen Zuordnung (dem 'Erwartungswert'), deren Wahrscheinlichkeit sich ergibt, wenn man 100 durch die Anzahl der Gruppen dividiert. Bei zwei Gruppen würde eine zufällige Verteilung im Schnitt 50%, bei fünf Gruppen 20% richtige Zuordnungen ergeben.

Auftretenswahrscheinlichkeit: Die Diskriminanzanalyse setzt normalerweise die Wahrscheinlichkeit, mit der Fälle, über die keine Information vorliegt, einer bestimmten Gruppe zugeordnet werden, für alle Gruppen gleich an; d.h. bei fünf möglichen Gruppen erhält jede eine Wahrscheinlichkeit von 20%. Wenn die Verteilung der Gruppen in der Stichprobe der Verteilung in der Grundgesamtheit (der 'Population') entspricht und von der Gleichverteilung abweicht, so kann die Auftretenswahrscheinlichkeit der einzelnen Gruppen vorgegeben und damit eine bessere Prädiktion erzielt werden, da die Auftretenswahrscheinlichkeit in die Berechnung der Prädiktion mit eingeht. Was die Verteilung der Satzmodi in der Population aller in einem bestimmten Zeitraum im deutschen Sprachraum realisierten Äußerungen betrifft, so läßt sich darüber nur spekulieren. Es kann aber als sicher angenommen werden, daß diskursspezifische Stichproben Ungleichverteilungen aufweisen. So werden beim Militär relativ gesehen mehr Imperativsätze realisiert werden, im Reisebüro oder bei einer Mensch-Maschine-Kommunikation mehr Fragesätze usw. So gesehen, ist die Ungleichverteilung in unserem Korpus, die sich aus der speziellen Form der Minimalpaarauswahl ergibt, gar nicht so ungewöhnlich. Eine Berücksichtigung dieses Umstands führt zwar zu besseren Ergebnissen; wir werden aber eine Gleichverteilung annehmen, da die Zugrundelegung aller möglichen Minimalpaare in unserem Korpus zur Folge hat, daß die stark überwiegenden Fragesätze immer auf Kosten der anderen Satzmodi besser klassifiziert werden.

Prädiktorvariablen: Bei jeweils nur einer Variablen als Prädiktor ist eine Interpretation der Ergebnisse noch relativ einfach. Sie wird schwieriger bei mehreren Variablen, da die Berücksichtigung einer zusätzlichen Variablen, von der man schon weiß, daß sie für eine bestimmte Distinktion relevant ist, nicht automatisch eine bessere Klassifikation ergibt, wenn alle Fälle und alle Gruppen klassifiziert werden. Eine Optimierung der Variablenauswahl ist natürlich erstrebenswert, bei unserem Testsatzkorpus allerdings noch nicht zu verwirklichen. Das sei an einem Beispiel erläutert: Eine Dehnung der Hauptakzentsilbe oder der ganzen Äußerung ist sicher eines der Merkmale, mit denen man einen Exklamativsatz kennzeichnen kann. Wenn man nun die Länge der Hauptakzentsilbe oder der Äußerung als Prädiktorvariable einsetzen möchte, so muß man natürlich wissen, um welche Segmente es sich handelt (Langvokal oder Kurzvokal etc.). Wenn darüberhinaus genug Exemplare mit der gleichen segmentalen Struktur im Korpus vorhanden sind, so kann man entweder nur die gleichen Strukturen miteinander vergleichen oder zumindest die relative Länge der Hauptakzentsilbe sinnvoll bestimmen (vgl. zu diesem Vorgehen Batliner 1987b). Da unser Korpus aber nicht auf diese Fragestellung hin konzipiert ist, gibt es in ihm zuwenig Fälle, bei denen die (relative) Länge als Prädiktorvariable sinnvoll angenommen werden könnte. Wir werden also nicht mit Variablen aus dem Zeitbereich, sondern nur mit Variablen aus dem Frequenzbereich arbeiten. Wir sehen auch davon ab, die einzelnen Diskriminanzfunktionen zu interpretieren (Welche Satzmodi werden in welchem Ausmaß durch welche Prädiktorvariable bestimmt?), da eine solche Interpretation in diesem Stadium noch recht fehleranfällig wäre.

4. STICHPROBENAUSWAHL UND TRANSFORMATION DER PRÄDIKTORVARIABLEN

Wir legen im folgenden immer die vier Fo-Variablen Onset (Fo-Wert am Beginn der Äußerung), Offset (Fo-Wert am Ende der Äußerung), Maximum (absolut höchster Fo-Wert der Äußerung) und Minimum (absolut tiefster Wert der Äußerung) den Berechnungen zugrunde. (D.h. daß die Variablen auch zufällig gleiche Werte annehmen und manchmal - z.B. bei Fragen mit einem sehr hohen Offset, der gleichzeitig das Maximum bildet - zusammenfallen können.) Diese Variablen haben den Vorteil, daß sie eindeutig bestimmbar sind, nicht auf der Zeitachse positioniert werden müssen und keine zusätzliche Information voraussetzen. (Eine solche zusätzliche Information ist z.B. beim Konturverlauf auf der Hauptakzentsilbe nötig, da diese Silbe ja erst bestimmt werden muß - z.B. durch Hörtests, so wie es in Oppenrieder 1988 beschrieben wird.) Die Variablen können sowohl mit analogen Geräten und nachfolgender Messung per Hand, wie in unserem Fall, als auch mit automatischen Algorithmen ermittelt werden. Den meßtechnischen Nachteil, daß sie ab und an nicht bestimmt werden können - etwa wenn der Sprecher am Ende der Äußerung laryngalisiert-, teilen sie mit anderen Variablen. (Bei unserer Stichprobe konnten wir in ca 5% der Fälle

einen der Fo-Werte nicht bestimmen. Solche Fälle werden von der Diskriminanzanalyse zwar nicht analysiert, wohl aber klassifiziert.) Vorteilhaft für eine Bearbeitung mit statistischen Verfahren ist, daß es sich um eindeutig intervallskalierte Daten handelt. Man beachte, daß die Variablen zwar gesondert in die Berechnungen eingehen, daß in ihnen aber andere, bekannte Zusammenhänge bzw. Beschreibungsgrößen 'versteckt' sind, von denen man annimmt, daß sie satzmodusrelevant sind: der globale **Deklinationsverlauf** (Differenz von Onset und Offset), der Tonumfang (**Range**, die Differenz zwischen Maximum und Minimum) und der finale hohe Tonverlauf (Offsethöhe). Durch die weiter unten beschriebenen Transformationen erhält man auch Angaben über die **Auslenkung** (monotoner vs. bewegter Tonverlauf) der Äußerung bzw. ihre **Positionierung** im Frequenzbereich bezüglich der sprecherspezifischen Basislinie.

Wir wollen mit unseren Analysen untersuchen, welche Verbesserung der Prädiktion sich durch die im folgenden beschriebenen Maßnahmen erreichen läßt:

Maßnahme 1 (Teilstichproben):

Zum einen sollen alle 956 Realisationen, zum anderen nur die 353 'Prototypen' zugrundegelegt werden. Prototypen nennen wir die Realisationen, die von den Versuchspersonen beim Hörtest im Kontext als natürlich, d.h. auf einer Skala von 1 bis 5 besser als 2.5, eingestuft wurden und die sie kontextfrei den intendierten Modi mit einer Trefferquote von mehr als 80% zuweisen konnten. (Damit sollte ein Filter gebildet werden, mit dem wir weniger akzeptable Realisationen aussondern; vgl. im einzelnen Oppenrieder 1988). Wir nehmen an, daß die Prototypen nicht nur von den Versuchspersonen richtig klassifiziert, sondern auch von der Diskriminanzanalyse besser prädiziert werden können. (Dieser Gesichtspunkt ist für die automatische Spracherkennung relevant, da man ja von einem automatischen Verfahren billigerweise nicht verlangen kann, wozu der Mensch nicht fähig ist: also etwa eine kontextfreie Kategorisierung von Nicht-Prototypen, die im Hörtest nicht dem richtigen Modus zugewiesen wurden.)

Maßnahme 2 (Analysierte vs. klassifizierte Stichprobe):

Wie oben erwähnt, kann man der Diskriminanzanalyse eine unterschiedlich große Auswahl aus der Gesamtstichprobe zur Analyse bzw. Klassifikation vorgeben. Erst dann, wenn man nicht die gesamte Stichprobe analysiert und anschließend klassifiziert, wird 'echt' prädiziert, also die Projektivität des Verfahrens ge-

testet. Es ist z.B. üblich, nur jeden zweiten Fall zu analysieren und dann die andere Hälfte zu klassifizieren. Da es sich bei unserer Stichprobe um sechs verschiedene Sprecher handelt und die Frage der Sprecherabhängigkeit bzw. -unabhängigkeit der Merkmale interessiert, bietet sich ein anderes Vorgehen an:

(1) Es wird reihum ein Sprecher analysiert und dann der Rest der Stichprobe, also fünf Sprecher, klassifiziert ($n-5$). Damit kann abgeschätzt werden, wie gut eine Prädiktion ist, die auf einer sprecherabhängigen Analyse beruht.

(2) Es werden reihum fünf Sprecher analysiert und als Grundlage für die Klassifikation des restlichen Sprechers genommen ($n-1$). Damit wird eine Sprecherunabhängigkeit der Klassifikation simuliert.

(3) Es werden alle sechs Sprecher analysiert und klassifiziert (n); damit läßt sich eine obere Grenze der Klassifikationsgüte beschreiben (Vgl. zum Stellenwert dieser Auswahl die Bemerkungen unten in der Diskussion).

Die Annahme, daß von $n-5$ über $n-1$ nach n eine Verbesserung eintritt, ist an sich trivial - es wäre verwunderlich, wenn sie nicht bestätigt würde. Interessant ist aber das Ausmaß der Verbesserung.

Maßnahme 3 (Transformation der Hz-Werte):

Es gibt verschiedene Ansichten darüber, ob Fo-Daten grundsätzlich auf einer Hz- oder einer Halbtonskala bzw. eher als absolute Werte oder bezogen auf einen Vergleichswert (gehörs-)adäquat repräsentiert sind. Bei einem Vergleich von Männer- und Frauenstimmen dürfte man sich aber immer für eine Transformation der Hz-Werte entscheiden. Um herauszufinden, wie sich die beste Prädiktion erzielen läßt, werden wir die Fo-Daten auf sechs unterschiedliche Weisen den Analysen zugrundelegen:

(1) Untransformierte Hz-Werte (Hz)

(2) Hz-Werte transformiert zum sprecherspezifischen Basiswert (Hz_{basis})

Der sprecherspezifische Basiswert ergibt sich aus dem tiefsten, vom jeweiligen Sprecher erreichten Offsetwert. Dieser Wert wird von jedem der vier Fo-Werte abgezogen.

(3) Hz-Werte, transformiert zum äußerungsspezifischen Mittelwert (Hz_{mittel})

Bei Hz_{basis} muß natürlich auf ein Wissen rekurriert werden, das man nicht anhand der jeweils betrachteten Äußerung erhält. Wir nehmen deshalb als äußerungsspezifischen Vergleichswert den Mittelwert der Äußerung an, der sich einfach aus dem Mittelwert der vier zur Verfügung stehenden Werte Onset, Offset, Maximum und Minimum ergibt. (Eine solche Berechnung simuliert sowohl die Situation, in der sich ein Hörer befindet, der eine einzige Äußerung eines ihm

fremden Sprechers beurteilen soll, als auch die Bedingungen eines sprecherunabhängigen automatischen Spracherkennungssystems; vgl. Nöth et al. 1987.)

(4) Halbtonwerte zur Basis 1 (Ht)

Die 'gehörsadäquate' Transformation in Halbtonwerte ergibt sich aus der 12. Wurzel aus zwei multipliziert mit dem natürlichen Logarithmus des Hz-Wertes - anders ausgedrückt, aus der Formel $17.31 \times LN(Hz)$.

(5) Ht-Werte transformiert zum sprecherspezifischen Basiswert (Ht_{basis})

Die Berechnung ist analog der zu HZ_{basis}.

(6) Ht-Werte, transformiert zum äußerungsspezifischen Mittelwert (Ht_{mittel})

Die Berechnung ist analog der zu HZ_{mittel}.

Maßnahme 4 (Moduskonstellationen):

Für Maßnahme 1-3 haben wir nur intonatorische Merkmale, eben die vier gewählten Fo-Werte, berücksichtigt. Natürlich wäre es möglich, auch syntaktische Merkmale wie Verb-Erst- vs. Verb-Zweit-Stellung, Vorhandensein eines W-Wortes etc. als Prädiktorvariablen anzusetzen. Der Status dieser Merkmale ist aber grundsätzlich anders: Es steht z.B. von vornherein fest, daß eine Aussage nicht mit einem W-Wort beginnen kann, d.h. wir haben es hier mit einem Merkmal zu tun, das bei unseren Merkmalkombinationen a priori distinktiven Charakter hat. Bei den intonatorischen Parametern wird dagegen die Diskriminanzanalyse als Instrument eingesetzt, mit dem ihre Relevanz für die Modusunterscheidungen erst entdeckt werden soll. Auch wird es sich bei intonatorischen Parametern nicht unbedingt um binäre, sondern oft um graduelle Merkmale handeln, bei denen z.T. nur die Extremwerte eindeutige Indikatoren darstellen: Eine Verb-Erst-Stellung ist entweder vorhanden oder nicht, ein hoher oder tiefer Offset kann mehr oder weniger ausgeprägt sein.

Wir wollen deshalb nicht nur ausschließlich intonatorische Merkmale zugrundelegen, sondern auch nicht-intonatorische. Das entspricht ja auch eher der Situation, in der sich der natürliche Sprecher/Hörer befindet: Man führt ja wohl keine komplette intonatorische Analyse einer Äußerung durch, bei der alle Modi als gleichermaßen wahrscheinlich angesetzt werden, wenn z.B. ein einleitendes W-Element die Äußerung von vornherein formal als Nicht-Aussagesatz und Nicht-Imperativsatz kennzeichnet. Natürlich soll und kann eine klassifizierende statistische Analyse kein Hörermodell darstellen, es ist aber doch sinnvoll, der Diskriminanzanalyse jeweils pro Fall nur die Modi zur Auswahl vorzulegen, denen

die Äußerung aufgrund von nicht-intonatorischen Merkmalen auch wirklich zugerechnet werden kann. (Auch ein automatisches Verfahren der Spracherkennung ist grundsätzlich in der Lage, z.B. ein einleitendes *W*-Element zu erkennen und damit Aussagesatz und Imperativsatz als Formtyp auszuschließen.) Als Möglichkeit in unserem Korpus ausgeschlossen sind durch **Verb-Erst-Stellung** Aussagesatz, durch **Verb-Zweit-Stellung** Wunschsatz und Imperativsatz, durch **Konjunktiv II** Imperativsatz, durch ein einleitendes **W-Element** Aussagesatz, Imperativsatz und Wunschsatz. Als letztes Merkmal wurden die **Modalpartikeln** berücksichtigt, die auch nur jeweils bestimmte Modi indizieren; *wohl* schließt z.B. einen Exklamativsatz aus, *etwa* einen Wunschsatz etc. Jeder Satz unseres Korpus' wurde aufgrund dieser Merkmale danach klassifiziert, welche Modi mit ihm nicht ausgedrückt werden können.

In unserem Korpus gibt es vier mögliche **Modus- (Minimalpaar- oder Tripel-) Konstellationen**; in Klammern ist jeweils die Zahl der Fälle im ganzen Korpus und bei den Prototypen angegeben:

- (1) **Fragesatz vs. Exklamativsatz**, z.B. *Hat der geflucht* (390/157)
- (2) **Fragesatz vs. Exklamativsatz vs. Imperativsatz**, z.B. *Stellt ihr euch an* (82/43)
- (3) **Fragesatz vs. Aussagesatz vs. Exklamativsatz**, z.B. *Die ist naiv* (145/45)
- (4) **Fragesatz vs. Exklamativsatz vs. Wunschsatz**, z.B. *Wäre ich glücklich* (67/12).

Hinzu kommen Fälle, bei denen jeweils nur ein einziger Modus indiziert wird; so kann der Satz *Stellt ihr euch etwa an* wegen der Modalpartikel *etwa* nur als Frage aufgefaßt werden.

Jede der Äußerungen wurde nun einer der vier Konstellationen zugeteilt oder für diesen Bearbeitungsschritt ausgeschieden. Für jede Konstellation wurden analog zur alleinigen Berücksichtigung der intonatorischen Parameter Diskriminanzanalysen durchgeführt.

5. ERGEBNISSE

Die Ergebnisse aller durchgeführten Diskriminanzanalysen sind in Tab.1 und 2 verzeichnet.

Tab.1 : Richtig klassifizierte Fälle in Prozent
(Alle Realisationen)

Werte	n-5		n-1		n	
H _z	40.4	51.1	53.4 +	67.7	55.6	73.2
H _z _{basis}	45.0	61.8	54.8 +	68.3	57.3 +	74.4
H _z _{mittel}	44.8	60.5	51.1	70.2	54.3	75.0
H _t	41.7	51.1	50.2	70.2	55.7	79.6 +
H _t _{basis}	49.7 +	66.5 +	53.1 +	72.6 +	58.6 +	80.8 +
H _t _{mittel}	51.5 +	65.9 +	53.0 +	73.6 +	56.6 +	80.1 +

Tab.2 : Richtig klassifizierte Fälle in Prozent
(Prototypen)

Werte	n-5		n-1		n	
H _z	48.6	58.0	59.2	75.1	66.3	89.5
H _z _{basis}	58.4	66.5	61.3	79.4	69.1 +	87.9
H _z _{mittel}	58.4	70.9	59.1	77.6	68.3 +	87.1
H _t	52.1	57.3	64.2 +	85.2 +	68.8 +	91.4 +
H _t _{basis}	61.2 +	72.0	66.1 +	83.9 +	68.5 +	92.4 +
H _t _{mittel}	61.8 +	79.5 +	65.7 +	83.6 +	68.8 +	92.0 +

Legende zu Tab.1 und 2:

kursiv: zusätzliche Berücksichtigung der möglichen Satzmodi.

analysierte vs. klassifizierte Stichproben ('Lern'- vs. 'Prüf'-Stichprobe):

n-5: reihum ein Sprecher analysiert, fünf klassifiziert

n-1: reihum fünf Sprecher analysiert, einer klassifiziert

n: alle Sprecher analysiert und klassifiziert.

'+' : beste Prädiktionen pro Spalte (Bereich: maximaler Prozentwert-2%).

In Tab.1 stehen die Werte für alle Realisationen, in Tab.2 nur die der Prototypen (vgl. oben Maßnahme 1); ansonsten ist der Aufbau der beiden Tabellen identisch. In der ersten Spalte ist angegeben, welche Art der Transformation den Zeilenwerten zugrundeliegt (vgl. oben Maßnahme 3). Jeweils zwei Spalten stehen für analysierte vs. klassifizierte Stichprobe (vgl. oben Maßnahme 2). Bei den nicht-kursiven Werten wurden nur die intonatorischen Variablen zugrundegelegt, bei den kursiven wurden nur die jeweils einer Moduskonstellation angehörigen Fälle gemeinsam berechnet (vgl. oben Maßnahme 4). Die Werte stehen in der linken (nicht-kursiv gekennzeichneten) Spalte von *n* für eine einzige Diskrimi-

nanzanalyse; ansonsten stellen sie Mittelwerte dar. So repräsentiert z.B. bei $n-1$, kursiv gekennzeichnete Spalte, jeder Wert einen Mittelwert aus 24 unterschiedlichen Diskriminanzanalysen (sechs Sprecher reihum bei der Analyse ausgelassen, pro Sprecher vier Moduskonstellationen). Die Mittelwerte sind **ungewichtet**, d.h. es wurde nicht berücksichtigt, daß z.B. die erste Moduskonstellation mehr Fälle repräsentiert als die anderen. Da aber nie gewichtet wurde, sind alle Werte untereinander vergleichbar. Um einen schnellen Überblick zu ermöglichen, sind die besten Werte pro Spalte (im Bereich 'maximaler Prozentwert-2%') mit '+' gekennzeichnet.

Wir fassen nun zuerst zusammen, in welchem Ausmaß die von uns vorgenommenen Maßnahmen die Prädiktion verbessern konnten. (Die dabei angegebenen Verbesserungen sind nicht genau aus den Werten von Tab.1 und 2 berechnet, sondern geben einen abgerundeten Betrag der Verbesserung wieder.)

Maßnahme 1 (Teilstichproben):

Die Prototypen können um gut 10% besser klassifiziert werden als alle anderen Realisationen zusammen. (Eine Klassifikation der Nicht-Prototypen, also der in unserem Sinne ungenügenden Realisationen, ergab eine etwa 20% schlechtere Prädiktion als die der Prototypen. Der Unterschied ist also systematisch und nicht etwa durch die geringere Anzahl der Prototypen im Verhältnis zu allen Realisationen bedingt.)

Maßnahme 2 (Analysierte vs. klassifizierte Stichprobe):

Von $n-5$ über $n-1$ nach n verbessert sich die Prädiktion um ca. 30%. (Bei $n-5$ und $n-1$ sind die Werte ein genaues Maß der Effektivität der Diskriminanzfunktion; d.h. wenn man die Funktion auf eine neue Stichprobe anwendet, so wird die Prädiktion die gleiche Güte haben. Bei n dagegen sind die Werte mit einem Bias behaftet und gelten deshalb nur für die vorliegende Stichprobe.)

Maßnahme 3 (Transformation der Hz-Werte):

Halbtonwerte resultieren in einer um ca. 5% besseren Prädiktion als Hz-Werte. Am besten ist die Prädiktion bei den transformierten Halbtonwerten, wobei es unerheblich ist, ob die Transformation sprecherspezifisch (zum Basiswert) oder äußerungsspezifisch (zum Mittelwert) erfolgte: Die Differenzen sind mit einer

Ausnahme geringer als 2%. Beide Transformationen sind also ein geeignetes Mittel, die unterschiedlichen Sprechlagen von Frauen und Männern vergleichbar zu machen. Am wichtigsten ist eine solche Transformation offensichtlich dann, wenn nur auf der Grundlage eines Sprechers oder einer Sprecherin prädiiziert wird: bei $n=5$ ist der Unterschied zwischen transformierten und nicht-transformierten Werten am höchsten.

Maßnahme 4 (Moduskonstellationen):

Bei Berücksichtigung der möglichen Moduskonstellationen verbessert sich die Prädiktion um gut 20%. Man beachte, daß sich dabei auch der Erwartungswert geändert hat; er ist nun - gewichtet aus den vier verschiedenen Moduskonstellationen und der jeweiligen Anzahl der Fälle - ca. 43%, während er sonst 20% beträgt. Ohne eine vergleichende Studie mit einem Modell, das unsere verschiedenen Konstellationen mit Zufallszahlen simuliert, kann prima facie nicht entschieden werden, in welchem Ausmaß die Verbesserung allein auf die Erhöhung des Erwartungswertes oder darauf zurückzuführen ist, daß die stärkere funktionale Belastung der Intonation in solchen Fällen eindeutiger und leichter zu klassifizierende F_0 -Werte zur Folge hat.

6. DAS FÄHNLEIN DER SIEBEN AUFRECHTEN

Vor der abschließenden Diskussion wollen wir nun die Fälle betrachten, die auch bei der besten Prädiktion von 92.4% (vgl. Tab.2, 6. Spalte, vorletzter Wert) noch fehlklassifiziert wurden. Da es sich dabei um Prototypen handelt - also um Fälle, die von den Hörern richtig klassifiziert und als natürlich bewertet wurden -, ist anzunehmen, daß dabei von uns nicht berücksichtigte Merkmale eine Rolle spielen. Tab.3 zeigt diese Fehlklassifikationen; in der ersten Spalte ist dabei die betreffende Moduskonstellation vezeichnet (vgl. Maßnahme 4).

Tab.3: Fehlklassifikationen:

Konst.	Satz	Anzahl Fälle	tatsächl. Modus	zugewiesener Modus
1	<i>Gehört das Ihnen hier *</i>	4	Frage	Exkl.
1	<i>Wie ist der reich geworden *</i>	10	Frage	Exkl.
1	<i>Wie laut ist es hier °</i>	2	Frage	Exkl.
1	<i>Stellt ihr euch vielleicht an °</i>	1	Exkl.	Frage
1	<i>Hat der geflucht °</i>	1	Exkl.	Frage
2	<i>Stellt ihr euch an °</i>	3	Imp.	Exkl.
3	<i>Er sieht was *</i>	1	Aussage	Exkl.
3	<i>Du kommst *</i>	2	Aussage	Exkl.
3	<i>Die ist naiv</i>	2	Exkl.	Aussage

Für die mit einem Stern gekennzeichneten Sätze lassen sich leicht zusätzliche Merkmale finden, mit deren Hilfe der richtige Modus klassifiziert werden kann: Die Verbsemantik schließt bei *Gehört das Ihnen hier* eine Exklamativinterpretation aus, ebenso bei *Er sieht was* und *Du kommst*. Die Realisationen von *Wie ist der reich geworden* haben alle, durch die Kontextvorgabe bedingt, den Satzakzent auf dem *W*-Wort - eine Akzentuierung, die ebenfalls eine Exklamativinterpretation ausschließt. Es ist sicher kein Zufall, daß bei allen diesen Fehlklassifikationen der Exklamativ mit beteiligt ist - ein Modus, dessen Status grundsätzlich nicht ganz eindeutig ist (vgl. Batliner 1988). Als von uns nicht berücksichtigte exklamativtypische Merkmale kommen zumindest Dehnung und die Position des Fo-Gipfels in Betracht. Nimmt man an, daß die Verwechslung von Exklamativ und Aussage bei *Die ist naiv* keine kommunikativen Schwierigkeiten nach sich ziehen würde, so bleiben sieben, in Tab.3 mit '*' gekennzeichnete 'kritische' Fehlklassifikationen übrig - kritisch insofern, als solche Verwechslungen der Intention des Sprechers beim Adressaten unadäquate Reaktionen hervorrufen. Das sind weniger als 3% der Stichprobe.

7. DISKUSSION

Linguistische oder phonetische Analysen validieren ihre Regeln meistens am aktuellen Korpus - die Überprüfung, ob diese Regeln auch auf andere Korpora, also andere Stichproben aus der Population, zutreffen, bleibt dann nachfolgenden Arbeiten überlassen. Insofern ist die zuletzt genannte Fehlerquote von weniger als 3% (die ja auch nur auf unsere Stichprobe zutrifft und bei anderen Stichproben höher ausfallen würde) vergleichbar und ohne 'doppelten Boden' erzielt.

Die Behauptung, damit sei ein Verfahren entdeckt, mit dem man in 'real life'-Situationen 97% der vom Sprecher intendierten Satzmodi richtig klassifizieren kann, wäre natürlich unsinnig. Wir möchten deshalb nun anhand einiger relevanter Punkte den Stellenwert unserer Ergebnisse diskutieren:

Sprecherunabhängigkeit: Eine Sprecherunabhängigkeit, so wie sie billigerweise bei automatischen Verfahren verlangt oder zumindest angestrebt werden sollte, ist bei unseren Berechnungen im Falle von $n-1$ simuliert. Das beste Ergebnis ist hier 73.6% für alle Realisationen. Man muß dabei allerdings bedenken, daß bei weitem nicht alle relevanten intonatorischen oder nicht-intonatorischen Merkmale in Betracht gezogen oder als Prädiktorvariablen optimiert wurden.

Produktionsbedingungen: Die Bedingungen, unter denen Korpora zustandekommen, sind ein wesentlicher Faktor für die Konsistenz dieser Korpora und damit für die Möglichkeit, Regeln aufzustellen und die Fälle erfolgreich zu klassifizieren. Auf der einen Seite stehen Korpora, bei denen erfahrene und kooperative Sprecher (z.B. die Autoren selbst) unter genau und explizit festgelegten Bedingungen die Äußerungen produzieren. Auf der anderen Seite stehen Aufnahmesituationen, die dem Ideal einer unbeobachteten und zwanglosen Kommunikation möglichst nahe kommen. Unser Korpus ist hier irgendwo dazwischen einzuordnen (vgl. zu den Aufnahmebedingungen im einzelnen Oppenrieder 1988): Einerseits werden 'naive' Sprecher nicht explizit (z.B. durch Vorgabe der Akzentposition, der Emphasestufe etc.) instruiert, wie sie produzieren sollen, sondern sie sollen nur durch eine Kontextvorgabe zur intendierten Produktion veranlaßt werden. Andererseits stehen die spezielle Minimalpaarkonstruktion sowie die Aufnahmebedingungen im schallarmen Raum 'ganz natürlich' produzierten Äußerungen entgegen. U.W. gibt es kein Maß, mit dem man den Grad der Natürlichkeit bestimmen könnte. Es gibt aber für unser Korpus doch einen gewissen Anhaltspunkt durch die Hörtests und dadurch, daß die mit deren Hilfe ausgewählten Prototypen um etwa 10% besser klassifiziert werden als alle Realisationen.

Diskursspezifität: Unser Korpus weist zwei Eigenheiten auf, die wahrscheinlich eine richtige Klassifikation anhand der von uns gewählten Prädiktorvariablen begünstigen: Zum einen sind die Äußerungen relativ kurz. Bei längeren Sätzen, die aus mehreren intonatorischen Phrasen mit jeweils einem Maximum und Minimum bestehen, dürften unsere vier Fo-Variablen allein keine so gute Prädiktion mehr gewährleisten. Zum anderen sind Fragen überrepräsentiert (gut die Hälfte der Fälle). Diese zwei Eigenheiten weisen aber auch restringierte Mensch-Maschine-Kommunikationen auf, z.B. Bahnauskünfte auf Fragen wie *Wann geht*

der Sonderzug nach Pankow? Insofern ist unser Korpus zwar nicht für die Gesamtpopulation, aber doch für bestimmte Anwendungen repräsentativ.

Prädiktorvariablen: Das stabilste intonatorische Merkmal dürfte die Offsethöhe sein, die sehr oft Fragen von Nicht-Fragen unterscheidet (vgl. Oppenrieder 1988). Da die Fragen in unserem Korpus so häufig vorkommen, ergibt der Offset allein als Prädiktorvariable eine nur um ca. 5% geringere Prädiktionsgüte (bei den besten Werten in Tab.1 und 2 z.B. 73.8 statt 80.8 und 88.8 statt 92.4). Das heißt aber nicht, daß der Offset ausreicht. Zum einen ist im Bereich von über 80% eine Verbesserung um 5% nicht zu verachten; sie zeigt, daß auch die anderen Variablen relevant sind. Zum anderen differenziert der Offset zwar gut zwischen Fragen und Nicht-Fragen, aber nicht innerhalb der Nicht-Fragen. Es ist deshalb nötig, an anderen Korpora zu untersuchen, welche weiteren Merkmale als Prädiktorvariablen in Frage kommen (vgl. dazu Batliner 1987b), um die Auswahl der Prädiktoren optimieren zu können.

8. SCHLUSSBEMERKUNGEN

Wie wir schon erwähnt haben, stand eine Optimierung der Prädiktorvariablen nicht im Mittelpunkt unserer Analysen, sondern die Frage, welche Verbesserung sich mit den von uns vorgenommenen vier Maßnahmen erzielen läßt. Wir wollen zwei Punkte zum Schluß noch einmal kommentieren:

(1) Die (plausible) Annahme, daß zu einem Vergleichswert transformierte Halbton-Werte die besten Ergebnisse bringen, hat sich voll bestätigt. Dies ist ein Beitrag zur Methodik.

(2) Bedenkt man, daß unser Korpus von seiner Konstruktion her für Verwechslungen prädestiniert ist, und daß unsere Prädiktorvariablen alles andere als optimiert waren, so läßt die erzielte Prädiktionsgüte hoffen. Dies ist ein Beitrag zur Frage, welche Aussicht eine automatische Bestimmung des Satzmodus in Zukunft haben dürfte.

LITERATUR:

- Altmann, Hans (1987): Zur Problematik der Konstitution von Satzmodi als Formtypen. In: Meibauer, Jörg (Hg.): Satzmodus zwischen Grammatik und Pragmatik. Tübingen, Niemeyer, 22-56. (=Linguistische Arbeiten 180).
- Batliner (1987a): Der Einsatz der Diskriminanzanalyse zur Prädiktion des Satzmodus. In: Tillmann, Hans G. / Willée, Gerd (Hgg.): Analyse und Synthese gesprochener Sprache. Hildesheim et al., Olms, 125-132.
- Batliner, Anton (1987b): Die intonatorische Indizierung des Fokus. Erste Ergebnisse zur Perzeption und Produktion. Ms.
- Batliner, Anton (1988): Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des Fo-Gipfels. (In diesem Band).
- Oppenrieder, Wilhelm (1988): Intonatorische Kennzeichnung von Satzmodi. (In diesem Band).
- Nöth, Elmar / Batliner, Anton / Lang, Roswitha / Oppenrieder, Wilhelm (1987): Automatische Grundfrequenzanalyse und Satzmodusdifferenzierung. In: Tillmann, Hans G. / Willée, Gerd (Hgg.): Analyse und Synthese gesprochener Sprache. Hildesheim et al., Olms, 59-66.

MODUS UND FOKUS ALS DIMENSIONEN EINER NONMETRISCHEN MULTIDIMENSIONALEN SKALIERUNG

Anton Batliner (München)

1. EINLEITUNG

Bei einer Urlaubsplanung bietet es sich an, in einem ersten Schritt die grobe Reiseroute anhand einer Karte mit einem kleinen Maßstab (1:500000 oder größer) zu bestimmen. Hat man dann die interessanten Reiseziele ausgesucht, so wird man sich in einem zweiten Schritt Karten mit einem großen Maßstab (1:100000 oder kleiner) beschaffen, um einzelne Tagesrouten, Wanderungen o.ä. genauer festlegen zu können. - Die Experimente, über die nun berichtet wird, stellen für den phonetisch-experimentellen Bereich des Münchner Intonationsprojekts einen ähnlichen ersten Schritt dar, mit dem ein Überblick über einen Teilbereich der durch intonatorische Merkmale unterscheidbaren Modus/Fokus-Konstellationen des Deutschen gewonnen und erste Hypothesen formuliert werden sollten, während ein weiterer Beitrag in diesem Band (Batliner 1988) sich - im zweiten Schritt - mit einer bestimmten einzelnen Konstellation beschäftigt. Um eine Vergleichbarkeit der Ergebnisse zu gewährleisten, wurde den Testitems ein einziger Satz zugrundegelegt, mit dem sich allerdings viele Minimalpaare bilden lassen, bei denen - neben der Verbstellung (vgl. Tab.1) - allein die Intonation Modus- und Fokusunterschiede indiziert (zur intonatorischen Minimalpaarbildung vgl. Oppenrieder 1988, zur Wahl genau dieses Testsatzes vgl. auch Batliner 1988). Bedingung war weiterhin, daß sich zu jeder Modus/Fokuskonstellation ein geeigneter Kontextsatz finden läßt.

Tab.1: Testsatz**1. Verb-Zweit-Stellung:**

unbetonter Artikel - Eigenname - gradulierbares Verb
Der Leo säuft

2. Verb-Erst-Stellung:

gradulierbares Verb - unbetonter Artikel - Eigenname
Säuft der Leo

2. MATERIAL UND EXPERIMENTELLES DESIGN

Im Tonstudio wurden vom Autor jeweils mehrere Realisationen der in Tab.2 und 3 aufgeführten Kontext- und Testsätze produziert und mit 19cm/sec auf einer Telefunken M15 aufgenommen. Nach einer Beurteilung durch mehrere kompetente Sprecher und einer Überprüfung der relevanten phonetischen Parameter anhand von Mingogrammen mit Zeitsignal, Fo- und Intensitätskurve wurde je ein typisches Exemplar ausgewählt. In Tab.2 und 3 sind die Kombinationen aus Kontext und dazugehörigem Testsatz mit den jeweiligen schematisierten Fo-Verläufen dargestellt. Der Kontextsatz hat dabei die Aufgabe, genau die angegebene Modus/Fokus-Interpretation des Testsatzes sicherzustellen. Die beiden Sätze geben immer eine zusammenhängende Äußerung und keinen Dialog wieder. Betonte Elemente sind durch Unterstreichung hervorgehoben.

Es wurden zwei randomisierte Testbänder erstellt, das erste mit dem Testsatz *Der Leo säuft* (von nun an '*DLS* '), das zweite mit dem Testsatz *Säuft der Leo* (von nun an '*SDL* '), wobei jeder Kontextsatz einmal mit jedem Testsatz kombiniert wurde. Es ergaben sich also für *DLS* $8 \times 8 = 64$ und für *SDL* $7 \times 7 = 49$ Satzpaare. (Bei *SDL* wurden noch weitere Kontextsätze mit aufgenommen, so daß im Experiment insgesamt 70 Kombinationen getestet wurden. Auf diese zusätzlichen Kombinationen werden wir hier aber nicht weiter eingehen.) Die Stimuli wurden im Tonstudio analog auf die Testbänder kopiert; das Intervall zwischen Kontextsatz und Testsatz betrug ca. 300-800 msec. Am ersten Test nahmen 25, am zweiten 22 bezahlte, freiwillige Versuchspersonen (Studenten der Linguistik oder Phonetik) teil. Die Anweisung lautete: "Sie hören Paare von Sätzen, wobei der zweite Satz immer der gleiche ist. Beide Sätze wurden von demselben Sprecher geäußert, es handelt sich also nie um einen Dialog, sondern immer um eine einzige Äußerung. Bitte entscheiden Sie, wie gut die beiden Sätze zusammenpassen." Bei *DLS* wurde den Versuchspersonen eine binäre Antwortmöglichkeit ('gut'- 'schlecht'), bei *SDL* eine fünfstufige Ratingskala (1='sehr gut', 2='gut', 3='noch gut', 4='schlecht', 5='sehr schlecht') vorgegeben. (Es sollte damit die Möglichkeit gegeben werden, unterschiedliche statistische Verfahren zu testen, die zum Teil abgestufte Daten, zum Teil aber auch kategoriale, also binäre Daten voraussetzen. Folgeexperimente mit einer fünfstufigen Ratingskala bei *DLS* zeigten, daß die unterschiedlichen Vorgaben das Ergebnis nicht beeinflussen.)

Tab.2: KONTEXTSÄTZE UND TESTSÄTZE: DER LEO SÄUFT

1. Aussagesatz; Fokus auf
- Leo*



Ihr fragt mich, wer säuft? Der Leo säuft.

2. Aussagesatz; Fokus auf
- säuft*



Was soll schon mit dem Leo sein? Der Leo säuft.

3. Aussagesatz mit Kontrastakzent; Fokus auf
- Leo*



Der Theo säuft nicht, der Leo säuft.

4. Aussagesatz mit Kontrastakzent; Fokus auf
- säuft*



Der Leo trinkt nicht, der Leo säuft.

5. Verb-Zweit-Exklamativsatz; Fokus auf
- Leo*



Gestern waren wir beim Bechern. Mann - Der Leo säuft!

6. Verb-Zweit-Exklamativsatz; Fokus auf
- säuft*



Gestern war ich mit dem Leo beim Bechern. Mann - Der Leo säuft!

7. Assertiver Fragesatz; Fokus auf
- Leo*



Wer sagst du säuft? Der Leo säuft?

8. Assertiver Fragesatz; Fokus auf
- säuft*



Stimmt das mit dem Leo? Der Leo säuft?

Tab.3: Kontextsätze und Testsätze: SÄUFT DER LEO

1. Verb-Erst-Fragesatz; Fokus auf *säuft*, steigende Kontur



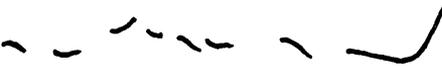
Stimmt das mit dem Leo? Säuft der Leo?

2. Verb-Erst-Fragesatz; Fokus auf *säuft*, fallende Kontur



Stimmt das mit dem Leo? Säuft der Leo?

3. Verb-Erst-Fragesatz; Fokus auf *Leo*, steigende Kontur



Wer säuft denn eigentlich? Säuft der Leo?

4. Verb-Erst-Fragesatz; Fokus auf *Leo*, fallende Kontur



Wer säuft denn eigentlich? Säuft der Leo?

5. Verb-Erst-Exklamativsatz; fokusneutraler Kontext, Akzent auf *säuft*, fallende Kontur



Also sowas hätt' ich nicht gedacht! Säuft der Leo!

6. Verb-Erst-Exklamativsatz; Fokus auf *Leo*, steigend-fallende Kontur



Gestern waren wir beim Bechern. Mann - Säuft der Leo!

7. Verb-Erst-Exklamativsatz; Fokus auf *säuft*, steigend-fallende Kontur



Gestern war ich mit dem Leo beim Bechern. Mann - Säuft der Leo!

3. ERGEBNIS

Tab.4 zeigt das Ergebnis für *DLS*, wobei die Zahlen die Häufigkeiten der 'paßt gut'-Bewertungen darstellen; das mögliche Maximum (die beste Bewertung) ist also 25, das mögliche Minimum (die schlechteste Bewertung) 0 als Zellenwert. Tab.5 zeigt das Ergebnis für *SDL*, wobei die Zahlen die Mittelwerte der Bewertungen durch die Testpersonen darstellen; das mögliche Maximum (die schlechteste Bewertung) ist also 5.00, das mögliche Minimum (die beste Bewertung) 1.00 als Zellenwert. Die Spalten stellen die Testsätze und die Zeilen die Kontextsätze dar. Die Numerierung ist dieselbe wie in Tab.2 bzw. 3 und weiter unten in Fig.1-4.

Tab.4: Rohdatenmatrix für *DLS* (*Der Leo säuft*)

	1	2	3	4	5	6	7	8
1	*25*	4	23	7	24	2	4	3
2	3	*24*	2	23	2	22	4	1
3	24	4	*24*	2	23	9	9	4
4	2	22	1	*20*	2	18	2	12
5	12	12	17	15	*23*	23	4	8
6	8	17	8	18	17	*22*	3	4
7	18	1	17	2	16	2	*23*	8
8	5	4	3	4	5	6	15	*24*

Tab.5: Rohdatenmatrix für *SDL* (*Säuft der Leo*)

	1	2	3	4	5	6	7
1	*1.36*	2.77	3.19	4.25	3.18	4.14	3.52
2	2.00	*1.55*	3.95	3.68	2.95	4.27	3.18
3	3.24	3.71	*1.41*	3.45	4.43	3.32	4.48
4	3.64	3.82	1.41	*3.48*	4.23	3.86	4.57
5	3.70	3.64	3.91	3.18	*2.68*	2.80	2.33
6	3.52	3.95	4.10	2.48	2.62	*2.64*	2.19
7	4.29	4.00	4.14	2.62	2.23	2.05	*1.77*

Das erste wichtige Ergebnis zeigt sich in den Diagonalen, also bei den mit Sternen gekennzeichneten Werten für die Originalkombinationen; diese Werte sind

fast immer sehr gut (Ausnahmen werden weiter unten besprochen), also hoch in Tab.4 und niedrig in Tab.5. Auf der anderen Seite gibt es genügend Kombinationen, die schlecht beurteilt werden. Damit sind zwei Grundvoraussetzungen für diese Art von Test erfüllt: die intendierten Originalkombinationen werden als solche 'erkannt' und heben sich ausreichend von anderen Kombinationen ab. Allerdings nicht von allen: in Tab.4 sind die Kombinationen 1/3 und 3/1 fast genauso gut wie die Originalkombinationen 1/1 und 3/3; das bestätigt die Vermutung, daß wir es bei normalem Aussagesatz (1) und Aussagesatz mit Kontrastakzent (3) nicht mit distinkten (Modus-)Kategorien zu tun haben (vgl. dazu Batliner 1988).

Es ist natürlich möglich, auf diese Art und Weise weiter zu vergleichen. Ein Vergleich aller Zellen untereinander 'via Augenschein' führt aber zu keiner übersichtlichen oder gar systematischen Lösung. Eine solche Lösung bietet erst das Verfahren der Nonmetrischen Multidimensionalen Skalierung (NMDS). Ausgangspunkt für dieses Verfahren sind Ähnlichkeitsmatrizen, wie sie in Tab.4 und Tab.5 vorliegen. (Aus den Matrizen läßt sich die Ähnlichkeit zweier Kombinationen einfach dadurch bestimmen, daß man den einen Zellenwert vom anderen subtrahiert; je größer der dabei resultierende absolute Wert ist, desto unähnlicher sind die beiden Kombinationen.) Ein iterativer Algorithmus führt zu einer Optimierung der Lösung, bei der die Objekte in einem n-dimensionalen Raum dargestellt werden, wobei unähnliche Objekte weit auseinander und ähnliche Objekte nahe beieinander zu liegen kommen.

Wir müssen an dieser Stelle von einer ausführlichen Beschreibung des Verfahrens absehen; vgl. dazu Borg (1981), der auch als illustratives Beispiel die Rekonstruktion einer Landkarte bringt, wenn nur die Entfernungen zwischen den Städten gegeben ist. Die Darstellung der Orte im Nord-Süd- und im Ost-West-Raum entspricht dabei einer zweidimensionalen NMDS-Lösung, die Entfernungen zwischen den Orten entsprechen den Ähnlichkeitsmaßen, wie sie der NMDS zugrundeliegen (Borg 1981:5ff). Die folgende Darstellung ist zuerst so wenig 'technisch' als möglich gehalten. In einem methodischen Teil werden wir anschließend Fragen und Probleme skizzieren, die mit der NMDS im allgemeinen und mit der Anwendung der NMDS auf unsere Daten im besonderen verbunden sind. Die NMDS-Lösungen wurden mit dem am Leibniz-Rechenzentrum in München implementierten Subprogramm von SPSS9 'SSA1' gerechnet (vgl. Toledano o.J.). Die Rohdatenmatrizen von Tab.4 und 5 wurden durch Mittelwertbildung aus der unteren und oberen Dreiecksmatrix symmetrisiert. Zum einen wurden dann, wie üblich, die Diagonalen auf Null gesetzt. Zum anderen wurden die Werte der Diago-

nalen, also die der Originalkombinationen, als 'Korrekturfaktor' verwendet, indem sie von den gemittelten Zellenwerten abgezogen wurden. Die Überlegung dabei war, daß dann, wenn die Originalkombination nicht gut bewertet wird, wie es etwa bei *SDL* für die Kombination 4/4 der Fall ist, diese Information mit berücksichtigt wird. Wir erhalten also vier symmetrisierte Matrizen, zwei für *DLS* und zwei für *SDL*. Für jede der Matrizen wurden ein- und zweidimensionale Lösungen errechnet. Da wir eine relativ geringe Anzahl von Items haben, sind Lösungen, die höher als zweidimensional sind, nicht mehr 'eindeutig' (vgl. zu dem Begriff Borg 1981:99ff). In Fig.1-4 sind die Alienationskoeffizienten für die gerechneten Lösungen aufgeführt. Eine Lösung wird normalerweise als genügend gut angenommen, wenn sie einen Koeffizienten aufweist, der kleiner als .15 ist. Das trifft in jedem der vier Fälle nicht auf die eindimensionale Lösung zu, die deshalb abgelehnt wird. Die zweidimensionale Lösung ist immer wesentlich besser als die eindimensionale. Sie entspricht auch unseren Strukturserwartungen, da wir die beiden Dimensionen Modus und Fokus von vornherein in unsere Items 'eingebaut' haben. Wir nehmen aus diesem Grund auch ein 'dimensionales' Modell an, d.h. wir suchen primär nicht nach Clustern (charakteristischen 'Haufenbildungen'), sondern danach, ob sich die Testitems entlang der beiden Dimensionen Modus und Fokus gruppieren.

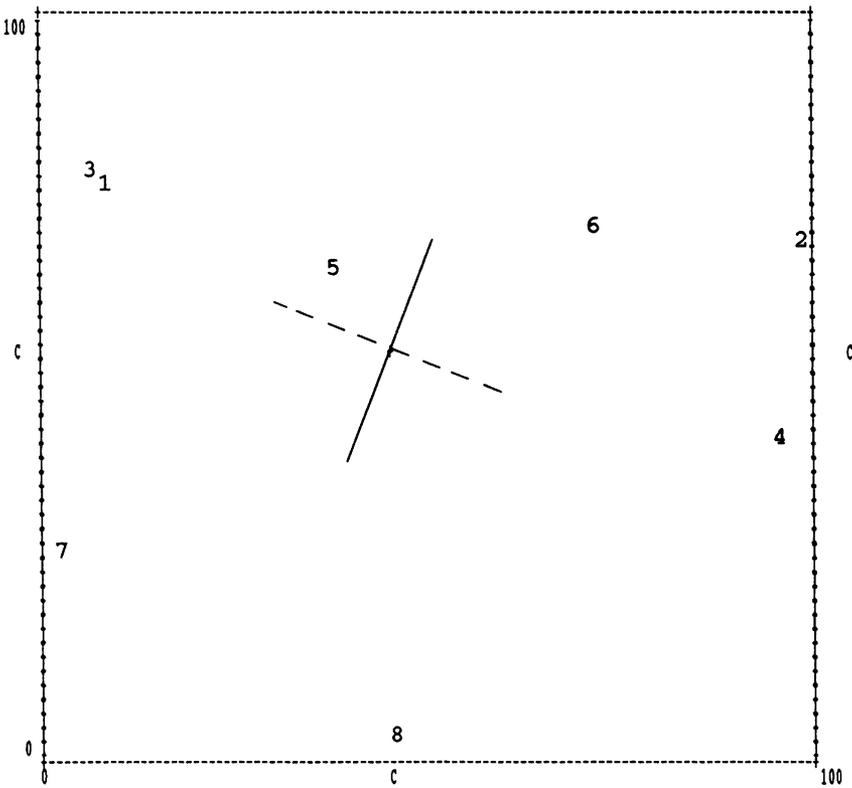
Die Gruppierung der Items im Lösungsraum ergibt sich aus einer Koordinatenmatrix, in der jedes Item n Koordinatenwerte (bei einer zweidimensionalen Lösung also zwei Werte) zugewiesen bekommt. Da diese Werte solange beliebig rotierbar und verschiebbar sind, wie sich der Abstand zwischen den Punkten nicht ändert, zeigen wir keine Koordinatenmatrizen, sondern nur die Anordnung der Items im zweidimensionalen Lösungsraum. Für unsere Interpretation nehmen wir an, daß Items, die unterschiedliche Modi bzw. Foki repräsentieren, weit auseinander liegen. Wir nehmen weiter an, daß sich die Items immer anhand unserer beiden Dimensionen gruppieren lassen, d.h. daß alle Items mit dem Fokus auf *Leo* in der einen Richtung angesiedelt sind, und alle Items mit dem Fokus auf *säuft* in der anderen, entgegengesetzten Richtung. Das gleiche gilt für den Modus. Es gilt auch der Umkehrschluß: Wenn Items auf einer Dimension nahe beieinander liegen, so unterscheiden sie sich nicht bezüglich dieser Dimension. Die Plausibilität der Interpretation und die Konvergenz der Ergebnisse mit den Ergebnissen aus anderen Verfahren entscheiden über die Qualität der Lösung.

Fig.1 und 2 zeigen die Lösungen für *DLS*, Fig.3 und 4 die für *SDL*. Es lassen sich immer zwei Dimensionen feststellen, die wir mit einer durchgezogenen Linie für Modus und einer gestrichelten Linie für Fokus verdeutlichen. (Man beachte,

daß die Linien keine Grenzen darstellen, sondern die Richtung, in der sich die Dimensionen ändern. Um einen gemeinsamen Bezugspunkt zu haben, wurden die Linien jeweils durch das Zentrum gelegt; diese Entscheidung ist beliebig.) Fig.1 und 2 unterscheiden sich nur unwesentlich voneinander. Die Fragesätze sind (links) unten, die Nicht-Fragesätze oben angesiedelt. Es sei nochmals darauf hingewiesen, daß diese Ortsangaben nicht absolut zu verstehen sind, da sie z.B. durch eine Rotation um 180 Grad ins Gegenteil verkehrt werden können. Entscheidend ist nur, daß die Fragesätze entgegengesetzt von den Nicht-Fragesätzen angesiedelt sind. Items mit dem Fokus auf *Leo* sind links (oben), solche mit dem Fokus auf *säuft* rechts angesiedelt. Normale Aussagesätze ohne und Aussagesätze mit Kontrastakzent liegen bei gleichem Fokus dicht beieinander, d.h. wir haben es hier entsprechend unseren Erwartungen nicht mit unterschiedlichen Modi zu tun. Bei den Exklamativsätzen ist die Fokusunterscheidung deutlich schwächer ausgeprägt als bei den Aussagesätzen, ebensowenig lassen sich Exklamativsätze und Aussagesätze auf der Dimension 'Frage vs. Nicht-Frage' differenzieren.

Um herauszufinden, ob sich Aussagesätze und Exklamativsätze nicht doch besser voneinander trennen lassen, wurde auch eine weitere zweidimensionale NMDS-Lösung ohne die assertiven Fragesätze, also nur mit Item 1-6, gerechnet. Dabei ließ sich wirklich eine Dimension feststellen, entlang derer Aussagesätze vs. Exklamativsätze angesiedelt sind. Diese Unterscheidung ist aber viel weniger dominant als die zwischen Frage und Nicht-Frage, da sie erst zum Tragen kommt, wenn man die Fragen aus der Analyse herausnimmt.

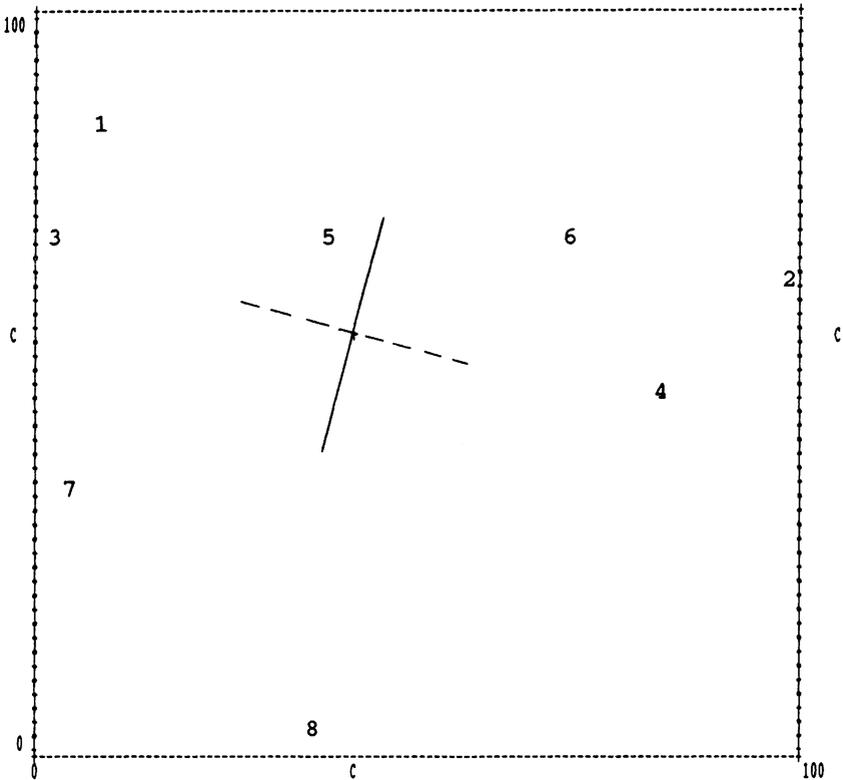
**Fig.1: Zweidimensionale NMDS-Lösung für *Der Leo säuft*,
Diagonale auf Null gesetzt**



- 1: Aussagesatz, Fokus auf *Leo*
- 2: Aussagesatz, Fokus auf *säuft*
- 3: Aussagesatz mit Kontrastakzent, Fokus auf *Leo*
- 4: Aussagesatz mit Kontrastakzent, Fokus auf *säuft*
- 5: Verb-Zweit-Exklamativsatz, Fokus auf *Leo*
- 6: Verb-Zweit-Exklamativsatz, Fokus auf *säuft*
- 7: assertiver Fragesatz, Fokus auf *Leo*
- 8: assertiver Fragesatz, Fokus auf *säuft*

Alienationskoeffizienten: eine Dimension: .320
zwei Dimensionen: .069

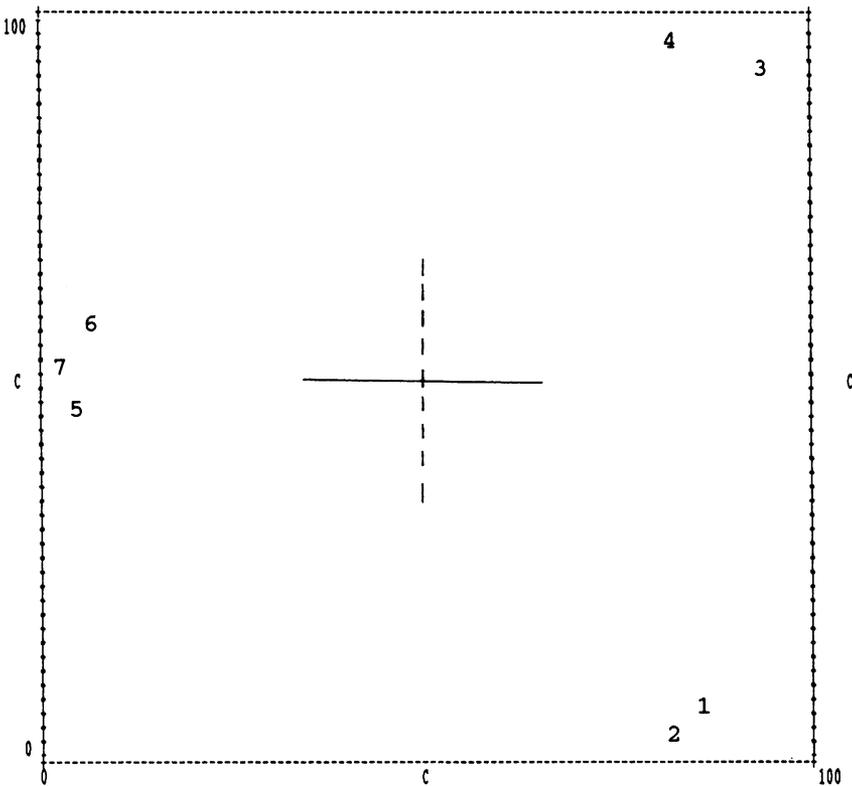
**Fig.2: Zweidimensionale NMDS-Lösung für *Der Leo säuft*,
'korrigiert' mit den Werten in der Diagonalen**



- 1: Aussagesatz, Fokus auf *Leo*
- 2: Aussagesatz, Fokus auf *säuft*
- 3: Aussagesatz mit Kontrastakzent, Fokus auf *Leo*
- 4: Aussagesatz mit Kontrastakzent, Fokus auf *säuft*
- 5: Verb-Zweit-Exklamativsatz, Fokus auf *Leo*
- 6: Verb-Zweit-Exklamativsatz, Fokus auf *säuft*
- 7: assertiver Fragesatz, Fokus auf *Leo*
- 8: assertiver Fragesatz, Fokus auf *säuft*

Alienationskoeffizienten: eine Dimension: .312
zwei Dimensionen: .079

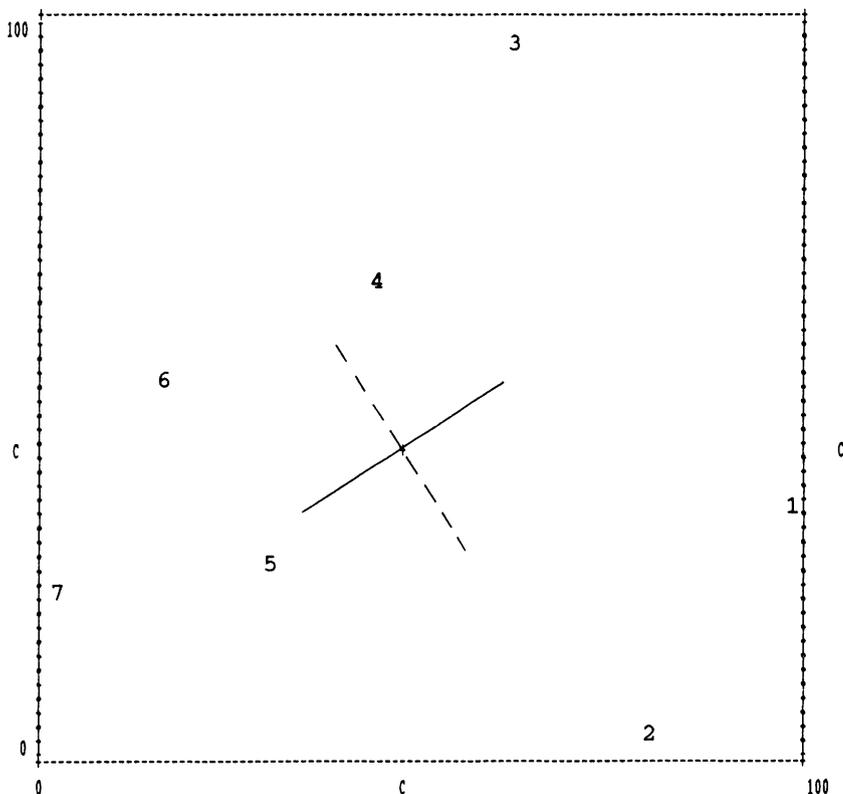
Fig.3: Zweidimensionale NMDS-Lösung für *Säuft der Leo*,
Diagonale auf Null gesetzt



- 1: Verb-Erst-Fragesatz, Fokus auf *säuft*, steigende Kontur
- 2: Verb-Erst-Fragesatz, Fokus auf *säuft*, fallende Kontur
- 3: Verb-Erst-Fragesatz, Fokus auf *Leo*, steigende Kontur
- 4: Verb-Erst-Fragesatz, Fokus auf *Leo*, fallende Kontur
- 5: Verb-Erst-Exklamativsatz, neutraler Fokus (Akzent auf *säuft*), fallende Kontur
- 6: Verb-Erst-Exklamativsatz, Fokus auf *Leo*, steigend-fallende Kontur
- 7: Verb-Erst-Exklamativsatz, Fokus auf *säuft*, steigend-fallende Kontur

Alienationskoeffizient: eine Dimension: .436
zwei Dimensionen: .015

Fig.4: Zweidimensionale Lösung für *Säuft der Leo*,
'korrigiert' mit den Werten in der Diagonalen



- 1: Verb-Erst-Fragesatz, Fokus auf *säuft*, steigende Kontur
- 2: Verb-Erst-Fragesatz, Fokus auf *säuft*, fallende Kontur
- 3: Verb-Erst-Fragesatz, Fokus auf *Leo*, steigende Kontur
- 4: Verb-Erst-Fragesatz, Fokus auf *Leo*, fallende Kontur
- 5: Verb-Erst-Exklamativsatz, neutraler Fokus (Akzent auf *säuft*), fallende Kontur
- 6: Verb-Erst-Exklamativsatz, Fokus auf *Leo*, steigend-fallende Kontur
- 7: Verb-Erst-Exklamativsatz, Fokus auf *säuft*, steigend-fallende Kontur

Alienationskoeffizient: eine Dimension: .44
zwei Dimensionen: .07

Im Gegensatz zu Fig.1 und 2 unterscheiden sich Fig.3 und 4 doch deutlich voneinander. In Fig.3 gibt es wieder zwei eindeutige Dimensionen: Die Exklamativsätze sind links, die Fragesätze rechts angesiedelt. Bei den Fragesätzen liegen die Items mit dem Fokus auf *Leo* oben, die mit dem Fokus auf *säuft* unten. Bei den Exklamativsätzen gibt es überhaupt keine Fokusdifferenzierung. In Fig.4, wo wir die Zellenwerte in Relation zu den Originalkombinationen gesetzt haben, sind die Dimensionen zwar dieselben und ebenfalls orthogonal aufeinander, die Verteilung ist aber weniger extrem: Es gibt bei den Exklamativsätzen eine rudimentäre Fokusdifferenzierung. Am interessantesten ist aber die Verteilung bei den Fragesätzen: Die Items mit fallender Intonationskontur (2 und 4) liegen jetzt **zwischen** den Fragesätzen mit steigendem Konturverlauf und den Exklamativsätzen. Die Bezeichnungen 'steigend' und 'fallend' beschreiben ja fürs erste Intonationsverläufe und keine (gar kognitiven) Dimensionen. Man kann allerdings die kategoriale Unterscheidung 'Frage vs. Nicht-Frage' weiter differenzieren und Frage als graduierbare Kategorie auffassen, also unterschiedliche Ausprägungen der 'Fragehaltigkeit' annehmen, die u.a. indiziert werden durch steigender Intonationsverlauf = stark fragehaltig und fallender Intonationsverlauf = weniger stark fragehaltig. (Weitere Überlegungen zur Fragehaltigkeit und erste experimentelle Bestätigungen dieses Konzepts finden sich in Batliner 1987.)

4. METHODISCHE ÜBERLEGUNGEN ZUR NMDS UND ZU IHRER ANWENDUNG AUF UNSERE DATEN

"Multidimensional scaling aims to determine a metric space in which the dimensions correspond to meaningful and relevant attributes of the stimulus." (Anderson 1981:364). Wir könnten uns damit zufrieden geben, daß wir genau das mit unserer Analyse gezeigt haben. Nun gibt es aber einige kritische Punkte bei der NMDS im allgemeinen und auch bei ihrer Anwendung auf unsere Daten im besonderen, auf die wir in diesem Abschnitt eingehen wollen.

Auf der methodischen Seite wird kritisiert, daß es keine definitiven Kriterien (etwa vergleichbar der Signifikanz des Ergebnisses in der Prüfstatistik) für die Güte des Modells und die Zahl der angenommenen Dimensionen gibt (Anderson 1981:364ff; zu den vorhandenen Kriterien vgl. wieder Borg 1981 passim). U.E. kann man in diesem Manko auch eine Tugend sehen, da damit verhindert wird, daß man glaubt, mit einem einzigen signifikanten Ergebnis etwas 'bewiesen' zu haben (vgl. dazu Rozeboom 1960:passim und Guttman 1977:86). Die Replikation von Experimenten und die Konvergenz mit anderen Ergebnissen erhalten also eher den ihnen zukommenden Stellenwert.

Auf der inhaltlichen Seite gibt es zwei wichtige Einwände, die auch für unsere spezielle Anwendung zu kommentieren sind:

(1) "[...] most of our subjective experience is based on continuous underlying dimensions [...] our perception and understanding of things and events is seldom

clearcut and exact. [Unsere Konzepte korrespondieren also nicht mit] a distinct point along some dimension or in some multidimensional space." (Oden 1979:409).

(2) "[...]valuation is asymmetrical. [Vergleicht man z.B. Adjektive, die Charaktereigenschaften beschreiben, also Adjektiv A mit Adjektiv B:] a *prudent* man would seem unlikely to be *cruel*, but a *cruel* man may well be *prudent*. The similarity models of multidimensional scaling require symmetry [but] the pervasive asymmetries of the valuation operation are not amenable to analysis by the standard spatial representations of multidimensional scaling." (Anderson 1981: 366; vgl. auch die Überlegungen zum sog. Ordnungseffekt in Schiefer/Batliner 1988).

ad (1): Die Bemerkung von Oden trifft u.E. auch auf Modus und Fokus zu: Es handelt sich hierbei nicht um klar distinkte binäre Kategorien, sondern um Kategorien, die oft prototypische Ausprägungen annehmen, die aber auch einen breiten Übergangsbereich und in sich Abstufungen aufweisen (zum Begriff der Kategorialität vgl. auch Schiefer/Batliner 1988). Hinzu kommt, daß unsere Realisationen der Kontext- und Testsätze zwar sicher prototypisch, aber nicht die einzig möglichen sind. Andere Realisationen dieser Sätze oder gar andere Sätze dürften auch eine andere NMDS-Lösung ergeben. Wir nehmen aber an, daß die **grundsätzliche** Lage der Items entlang der Dimensionen Modus und Fokus zueinander sich nicht ändern wird und daß größere Abweichungen immer sinnvoll interpretiert werden können, so wie das auch bei den Abweichungen von der Symmetriebedingung der Fall ist, s.u. Um es mit einem Bild zu sagen: Solche unterschiedlichen NMDS-Lösungen dürften sich zueinander wie Momentaufnahmen des gleichen Gesichts verhalten, das unterschiedliche Ausdrücke annehmen kann, aber doch immer das gleiche Gesicht bleibt. Hier zeigt sich auch ein wichtiger Aspekt, in dem die Landkartenanalogie aus der Einleitung **nicht** auf die NMDS-Lösung übertragbar ist: nicht nur aus methodischen Gründen (da Ordinaldaten vorliegen), sondern auch aus inhaltlichen Gründen ist es unsinnig, exakte Entfernungen zwischen etwa Frage und Exklamativ aus Fig.1 bis 4 herauslesen zu wollen, so wie man auf einer Landkarte die Entfernung zwischen Orten ablesen kann.

ad (2): Dem inhaltlichen Problem der Asymmetrie ist das methodische vorgeordnet, daß Ordinaldaten streng genommen nicht gemittelt werden dürften - obwohl das allgemeiner Usus ist. Man dürfte also weder über die Versuchspersonen mitteln noch zwischen der oberen und der unteren Dreiecksmatrix, "[...] da die Werte ja in jeder beliebigen Weise transformierbar sind, solange nur ihre Ordnung erhalten bleibt." (Borg 1981:283). Bei der Mittelung werden die Werte aber nicht in irgendeiner beliebigen, sondern eben nur in einer ganz bestimmten Weise transformiert. Das Argument würde, zu Ende gedacht, z.B. sämtliche Anwendungen parametrischer Verfahren auf Daten, die nicht physikalische Meßgrößen darstellen, verbieten. Wenn man die obere und die untere Dreiecksmatrix getrennt einer Analyse unterzieht, so wird dieser Forderung in einem Aspekt entsprochen. Dieser Weg ist aber in unserem Fall nicht gangbar, da die Positionen der Kontext- und der Testsätze zwar grundsätzlich, aber nicht immer bei dieser Form der Minimalpaarbildung austauschbar sind: So ist zwar die Kombination 3 in Tab.2 umkehrbar: *Der Leo säuft, der Theo säuft nicht*, nicht aber (ohne Bedeutungsänderung) etwa die Kombination 1 in Tab.2: *Der Leo säuft. Ihr fragt mich, wer säuft?* Und damit kommen wir zum eigentlich kritischen, aber auch interessanten Punkt: Das Verhältnis von Kontext- und Testsatz läßt sich im Prinzip zwar in der zeitlichen Anordnung, nicht aber phänomenal umkehren. Der Kontext **korrespondiert** im von ihm indizierten Modus und Fokus nicht nur mit dem Testsatz, er **evoziert** (wenn er an erster Stelle steht) oder präzisiert zumindest die Modus/Fokus-Lesart des Testsatzes. Dadurch sind die nicht zufälligen, sondern systematischen Asymmetrien bedingt, die wir im nächsten Abschnitt

behandeln wollen. Wir sind allerdings nicht der Ansicht, daß sich deshalb eine Anwendung der NMDS auf unsere Daten verbietet - genausowenig wie wir meinen, daß die in der Psychophysik und in Experimenten zur Kategorialen Wahrnehmung übliche Mittelung von AB und BA bei z.B. Diskriminationstests alle dabei gewonnenen Ergebnisse wertlos macht (vgl. auch Schiefer/Batliner 1988). Man muß sich nur immer darüber im klaren sein, daß bei dieser Mittelung bestimmte Phänomene unter den Tisch fallen.

Ein weiteres mögliches Argument sei kurz kommentiert: Warum testen wir Kombinationen von Kontext- und Testsatz und nicht einfach die Testsätze in ihrer unterschiedlichen intonatorischen Ausprägung? Auch bei dieser üblichen Art von Test lassen sich interessante Ergebnisse erzielen, vgl. Batliner (1987). Allerdings ist es bei der vorliegenden Art von Test viel sicherer, daß die Versuchspersonen nicht rein intonatorische oder psychophysische Parameter beurteilen, sondern wirklich das **Zusammenspiel** von intonatorischen und anderen grammatischen Merkmalen, also der Merkmale, die Modus und Fokus kennzeichnen.

5. ASYMMETRIE UND KONVERGENZ

Wir wollen nun anhand einiger asymmetrischer Zellenwerte zeigen, welche systematischen Faktoren für diese Asymmetrien verantwortlich sind. In Tab.6 sind für drei Modus/Fokus-Kombinationen die Rohdaten aus Tab.4 und 5 noch einmal aufgetragen. Die Unterschiede, die wir kommentieren wollen, sind mit '>' (größer als), '<' (kleiner als), '≈' (ungefähr gleich) und '=' (gleich) gekennzeichnet. Die Spalten zeigen die Modus/Fokus-Kombination B, die Zeilen zuerst AB und dann BA, wobei A die zu kommentierende Kombination darstellt. Die erste Stelle im Paar steht immer für den Kontextsatz, die zweite für den Testsatz.

Tab.6: Asymmetrische Zellenwerte

		1	2	3	4	5	6	7	8
(1)	5 _{DLS} B	12 = 12		17 ≈ 15		23	23	4	8
		< >		< >					
	B5 _{DLS}	24 > 2		23 > 2		23	17	16	5
(2)	7 _{DLS} B	18	1	17	2	16	2	23	8
		>	>	>	>				
	B7 _{DLS}	4	4	9	2	4	3	23	15
(3)	4 _{SDL} B	3.64	3.82	1.41	3.48	4.23	3.86	4.57	
		<	<	>	>	>	>	>	
	B4 _{SDL}	4.25	3.68	3.45	3.48	3.18	2.48	2.62	

(1) 5_{DLS}: Verb-Zweit-Exklamativsatz, Fokus auf *Leo*

(2) 7_{DLS}: assertiver Fragesatz, Fokus auf *Leo*

(3) 4_{SDL}: Verb-Erst-Fragesatz, Fokus auf *Leo*, fallende Kontur

(1) Verb-Zweit-Exklamativsatz, Fokus auf Leo (5DLs):

Legt man als Kontextsatz den zum Exklamativsatz 5DLs passenden zugrunde, so zeigt sich bei den Kombinationen mit Aussagesatz (1 und 2) bzw. Aussagesatz mit Kontrastakzent (3 und 4), daß alle als mittelgut eingestuft werden; es gibt also hier keinen Fokuskonflikt. Wenn 5DLs an zweiter Stelle steht, so zeigt sich dagegen ein eindeutiger Fokuskonflikt: die Kombinationen 1/5 und 3/5 werden so gut bewertet wie die Originalkombinationen 1/1 bzw. 3/3, wogegen die Kombinationen 2/5 und 4/5 so schlecht wie die entsprechenden Aussagekombinationen mit Fokuskonflikt 1/2 und 2/1 bzw. 4/3 und 3/4 bewertet werden, vgl. Tab.4. Diese Unterschiede lassen sich so erklären, daß Aussagekontexte immer einen Fokus indizieren, und daß eine exklamativsatztypische Intonationskontur beim Testsatz auch mit einem (emphatischen) Aussagekontext verträglich ist, solange die Foki übereinstimmen. Einem exklamativindizierenden Kontextsatz muß nicht unbedingt ein Testsatz mit einer exklamativtypischen Intonationskontur folgen, da der Ausdruck der Exklamativität auch vorverlagert sein kann, also z.B. in das *Mann-* des Kontextsatzes. Ebensowenig muß eine Übereinstimmung in der Fokusstruktur (im Gegensatz zum Aussagesatz) bestehen. (Bei Item 6 in Tab.4, also dem Exklamativsatz mit dem Fokus auf *säuft*, ist im Gegensatz dazu eine (rudimentäre) Fokusstruktur festzustellen; wir können noch nicht angeben, wodurch diese Unterschiede bedingt sind.) Eine fehlende bzw. nur rudimentär vorhandene Fokusstruktur und das Fehlen einer exklamativtypischen Intonationskontur beim Exklamativsatz sowie die Möglichkeit, auch beim Aussagesatz akzeptable exklamativtypische Intonationskonturen beim Testsatz zu finden, zeigten sich auch in den weiterführenden Perzeptionsexperimenten, die in Batliner (1988) beschrieben sind.

(2) Assertiver Fragesatz, Fokus auf Leo (7DLs):

Die Kombinationen 7/1, 7/3 und 7/5 von assertivem Fragesatz als Kontext und Nicht-Fragesatz als Testsatz mit übereinstimmendem Fokus auf *Leo* werden unerwartet gut bewertet. Postexperimentelle Befragungen ergaben, daß die Versuchspersonen trotz des Hinweises, daß es sich immer um die Äußerung eines einzigen Sprechers handelt, in diesem Fall oft einen Dialog beurteilten - und als Dialog ist die Abfolge *Wer sagst du, säuft?* - *Der Leo säuft* natürlich voll akzeptabel. Damit erklärt sich auch die Lage der Items zueinander in Fig.1 und 2, wo Item 7 um einiges näher an den Aussagen liegt als Item 8, bei dem eine Interpretation als Dialog unwahrscheinlich ist.

(3) Verb-Erst-Fragesatz, Fokus auf *Leo*, fallende Kontur (4_{SDL}):

Bei der Auswahl der Teststimuli waren sich die Beurteiler darüber einig, daß dieser Testsatz seltsam klingt; offenbar ist ein fallender Intonationsverlauf und eine Fokussierung des finalen Elements nicht gut mit einer Frageinterpretation verträglich. Wir vermuten, daß es sich dabei um ein systematisches 'gap' in den möglichen Merkmalkombinationen handelt. Um alle Kombinationen durchspielen zu können, wurde dieses Item aber trotzdem in den Test mit aufgenommen. Es zeigte sich denn auch in Fig.4, daß dieses Item den Exklamativsätzen am nächsten liegt; so wurden auch die Kombinationen mit exklamativindizierenden Kontexten durchgehend besser bewertet, vgl. Tab.6, besser sogar als die Originalkombination 4/4. Da der entsprechende Kontextsatz diese intonatorische Charakteristik nicht aufweist, wird auch die Kombination 4/3 viel besser bewertet als die Kombination 3/4 (vgl. Tab.5: 1.41 vs. 3.45).

Wir haben nun die augenfälligsten Asymmetrien behandelt, was nicht heißt, daß andere, weniger auffällige nicht auch systematische Ursachen haben können. Wie wir oben in Teil 4 betont haben, sind wir aber nicht der Ansicht, daß die NMDS wegen dieser Asymmetrien ein fehlerhaftes Instrument darstellt. Es kommt auf den Verwendungszweck an: Ein Brotmesser ist kein Chirurgenkalpell, leistet aber, richtig angewandt, auch gute Dienste. (Wie ein Brotmesser ist auch die NMDS ein relativ robustes Instrument. Wir rechneten z.B. Analysen ohne die problematischen Items 7_{DLS} und 4_{SDL}; dabei veränderte sich die Lage der restlichen Punkte zueinander nur minimal.) Wie der Vergleich mit den Übersichtskarten in der Einleitung deutlich machen sollte, kann man mit der NMDS einen anschaulichen und zugleich systematischen Überblick über viele unterschiedliche Modus/Fokus-Konstellationen erhalten, so wie es auf anderem Weg nicht möglich wäre. Die inhaltlichen Folgerungen, die wir in den Schlußbemerkungen zusammenfassen, müssen in einem zweiten Schritt durch Perzeptionsexperimente, die ganz bestimmte Konstellationen einzeln untersuchen, abgesichert werden. Zwei solcher Konstellationen wurden schon genauer untersucht (vgl. Batliner 1987 und Batliner 1988). Die Ergebnisse konvergieren mit denen der NMDS.

6. SCHLUSSBEMERKUNGEN

Abschließend seien die Ergebnisse für die beiden Dimensionen Modus und Fokus kurz zusammengefaßt:

Die Dimension Fokus ist orthogonal (also im rechten Winkel) zur Dimension Modus, d.h. die beiden Dimensionen sind voneinander unabhängig. Der Fokus ist bei Frage und Aussage relevant, beim Exklamativ wenig relevant bis irrelevant. Es bietet sich deshalb an, beim Exklamativ nicht von 'Fokusakzent', sondern nur von 'Hauptakzent' zu sprechen.

Beim Modus ist Frage vs. Nicht-Frage eine deutliche Dimension, sowohl bei Satzfrage vs. Exklamativ als auch bei assertiver Frage vs. Nicht-Frage (Aussage und Exklamativ). Dieser Unterschied ist dominant im Gegensatz zum Modusunterschied Aussage vs. Exklamativ, der nur sichtbar wird, wenn die Fragen aus der Analyse genommen werden. Die Verteilung entlang der Modusdimension legt nahe, besonders bei der Frage keine diskrete, sondern eine abgestufte Kategorie anzusetzen. Folgeexperimente bestätigen diese Interpretation.

LITERATUR

Anderson, Norman H. (1981): Foundations of Information Integration Theory. New York u.a., Academic Press.

Batliner, Anton (1987): Kategorialität und Kontexteffekte bei Frage- und Exklamativmodus im Deutschen. - Perzeptionsexperimente zur Rolle des Fo-Verlaufs. Ms.

Batliner, Anton (1988): Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des Fo-Gipfels. (In diesem Band)

Borg, Ingwer (1981): Anwendungsorientierte Multidimensionale Skalierung. Berlin u.a., Springer.

Guttman, Louis (1977): What is not What in Statistics. The Statistician 26, 81-107.

Oden, Gregg C. (1979): Fuzzy Propositional Approach to Psycholinguistic Problems: An Application of Fuzzy Set Theory in Cognitive Science. In: Gupta, Madan M. / Ragade, Rammohan K. / Yager, Ronald R. (Hgg.): Advances in Fuzzy Set Theory and Applications. Amsterdam u.a., North-Holland Publishing Company, 409-420.

Oppenrieder, Wilhelm (1988): Intonatorische Kennzeichnung von Satzmodi. (In diesem Band)

Rozeboom, William W. (1960): The Fallacy of the Null-Hypothesis Significance Test. Psychological Bulletin 57, 416-428.

Schiefer, Lieselotte / Batliner, Anton (1988): Intonation, Ordnungseffekt und das Paradigma der Kategorialen Wahrnehmung. (In diesem Band)

Toledano, Shlomo (o.J.): Smallest Space Analysis. A New Subprogram 'SSA1' in the SPSS Package. Xerox, Hebrew University of Jerusalem, Computation Center.

DER EXKLAMATIV: MEHR ALS AUSSAGE ODER DOCH NUR MEHR ODER WENIGER AUSSAGE? EXPERIMENTE ZUR ROLLE VON HÖHE UND POSITION DES F₀-GIPFELS.

Anton Batliner (München)

1. EINLEITUNG

LINGUISTISCHE KLASSIFIKATION: Ausgangspunkt der im folgenden beschriebenen Experimente ist das von H. Altmann entwickelte Satzmodussystem, vgl. Altmann (1987); **Thema** ist die intonatorisch unterschiedliche Markierung der Modi **Aussage** und **Exklamativ**, und zwar bei derjenigen Minimalpaarkonstellation, bei der die Intonation allein die beiden Modi unterscheidet, also bei Verb-Zweit-Stellung ohne zusätzliche Merkmale (wie Modalpartikel o.ä.). Im Gegensatz zu den Hauptkategorien wie Aussage, Frage und Imperativ ist dabei der Status des Exklamativs weniger eindeutig bestimmbar; z.T. wird er von anderen Autoren als ein Spezialfall der (emphatischen) Aussage, also nicht als eigenständiger Modus betrachtet; vgl. dazu Näf (1987). Daß der Exklamativ der Aussage näher steht als der Frage und auch von der Aussage nicht so eindeutig zu trennen ist, zeigt sich auch in der in Batliner (1988) beschriebenen Analyse im Rahmen einer Nonmetrischen Multidimensionalen Skalierung. Nimmt man realistischerweise an, daß die Kategorialität der Modi grundsätzlich nur in ihren Kernbereichen deutlich ausgeprägt ist, daß es aber an ihren Rändern Mischtypen gibt (vgl. das Konzept der 'Fuzziness', s. u.a. Oden 1979), so kann man weiter vom Exklamativ annehmen, daß wir es hier mit einer Kategorie zu tun haben werden, die besonders 'fuzzy' ist.

LINGUISTISCHER VERSUS PHONETISCHER GESICHTSPUNKT: Von Haus aus interessieren sich Linguisten mehr für die formalen, strukturellen und semantischen Eigenschaften des Exklamativs und behandeln seine phonetischen Eigenschaften eher global; Näf (1987:146) etwa begnügt sich mit der Charakterisierung "portato-ritando-Intonation". Auf der anderen Seite spricht die Phonetik normalerweise nicht von Exklamativ, sondern unspezifiziert von Emphase, und begnügt sich dabei mit einem allgemeinen vortheoretischen Verständnis, ohne den

Begriff abzuleiten. Man kann auch in den allermeisten Fällen Einverständnis darüber erzielen, welche Äußerung in welcher Situation emphatisch gemeint ist. (Ein anderer Grund dürfte sein, daß Phonetiker solche Begriffsbestimmungen lieber den Linguisten überlassen.) Versucht man nun, wie in diesem Beitrag, einer phonetischen Untersuchung ein streng linguistisch definiertes Material zugrundezulegen, so erkaufte man sich mit einem grundsätzlichen Vorteil eine praktische Schwierigkeit: vorteilhaft ist, daß die phonetischen Ergebnisse ohne weiteres in das linguistische Modell inkorporierbar sind; die Schnittstelle zwischen Linguistik und Phonetik ist also gegeben. Nachteilig kann sein, daß die linguistischen Vorgaben die Wahl des Stimulusmaterials stark einschränken.

INTONATORISCHE MARKIERUNG: Grob gesprochen, zeichnet sich der Exklamativ durch ein 'Mehr' aus: auf der funktionalen, inhaltlichen Seite durch größeres emotionales Engagement, Überraschung, Erstaunen; auf der formalen Seite bei der Intonation durch stärkere Ausprägungen der relevanten Parameter. Kennzeichen sind also im Fo-Bereich größere **Maxima**, im Zeitbereich größere **Dauer**, und eventuell größere **Intensität**. Um es phonologisch auszudrücken, wobei man sich dabei weniger binäre Oppositionen als vielmehr n-dimensionale Merkmale vorstellen sollte: die Aussage ist der Normalfall, also **unmarkiert**; der Exklamativ ist eine **markierte** Kategorie und wird deshalb auch durch markierte Parameterausprägungen indiziert. Markiert können dabei nicht nur die drei erwähnten Standardparameter sein, sondern auch andere, z.B. die Fo-**Gipfelposition**.

EXPERIMENTELLE OPERATIONALISIERUNG: Es wurde ein Testsatz gewählt, der bestmöglich den linguistischen und phonetischen Anforderungen entspricht. D.h. es sollte ein Satz sein, mit dem - durch unterschiedliche intonatorische Markierung - möglichst viele verschiedene Modus- und Fokuskonstellationen indiziert werden können. Wie sich zeigen wird, lassen sich die Modi auch durch eine unterschiedliche Behandlung der Fokusstruktur kennzeichnen; Modus und Fokus können also nur bedingt isoliert voneinander betrachtet werden. Phonetisch-experimentelle Kriterien sprechen für einen kurzen Satz mit möglichst vielen stimmhaften Segmenten. Die linguistischen Kriterien für einen Exklamativ erfordern hier einen definiten Ausdruck sowie ein graduierbares Prädikat. Wir entschieden uns für den Satz *Der Leo säuft*, mit dem u.a. die in Tab.1 aufgeführten Konstellationen erzeugbar sind. Wenn der Kontextsatz einen Satzakzent auf *Leo* indiziert, so sprechen wir vom 'frühen', wenn ein Hauptakzent auf *säuft* indiziert wird, vom 'späten' Fokus. In der Tabelle sind auch die Kontextsätze aufgeführt, mit denen im Experiment genau die jeweiligen Konstellationen indiziert werden sollten. Eine Bemerkung zur Terminologie: Im folgenden verwen-

den wir statt einiger expliziter Begriffe oft Kurzformen: KONTRAST etwa steht für 'Aussagesatz mit Kontrastakzent', AUSSAGE für 'normaler Aussagesatz ohne Kontrastakzent' usw. Diese Kurzformen werden auch später beim Versuchsplan für die Faktoren und ihre Klassen verwendet. 'RISE', 'LEVEL' und 'FALL' stehen für 'Anstieg', 'ebener Verlauf' bzw. 'Abfall' der Fo-Kontur, 'Gipfel' für 'Maximum und Wendepunkt der Fo-Kontur'.

Tab.1: Modus/Fokus-Konstellationen und Kontextsätze

AUSSAGE, FRÜHER FOKUS:	<i>Ihr fragt mich, wer säuft?</i>
AUSSAGE, SPÄTER FOKUS:	<i>Was soll schon mit dem Leo sein?</i>
KONTRAST, FRÜHER FOKUS:	<i>Der Theo säuft nicht,</i>
KONTRAST, SPÄTER FOKUS:	<i>Der Leo trinkt nicht,</i>
EXLAMATIV, FRÜHER FOKUS:	<i>Gestern waren wir im Wirtshaus. Mann -</i>
EXKLAMATIV, SPÄTER FOKUS:	<i>Gestern war ich mit dem Leo im Wirtshaus. Mann -</i>

GRUNDLEGENDE FRAGESTELLUNG: Wir interessieren uns in diesem Beitrag für die schon im Titel angedeutete Frage, ob sich der Exklamativ immer durch eine 'stärkere' intonatorische Markierung gegenüber der Aussage auszeichnet. Es wird sich zeigen, daß das nicht notwendig der Fall sein muß, daß es aber trotzdem auch in bezug auf die intonatorische Markierung sinnvoll ist, vom Exklamativ als einer eigenen Kategorie zu sprechen.

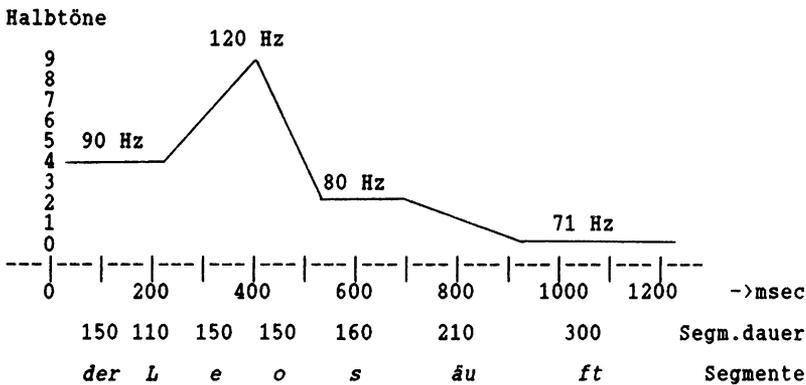
2. TESTAUFBAU

Um Redundanzen in der Darstellung zu vermeiden, werden die folgenden Punkte, soweit sie auf alle Experimente zutreffen, gesammelt abgehandelt. Bei den einzelnen Experimenten wird dann nur noch kurz darauf eingegangen.

MATERIAL: Der Autor produzierte die Kontextsätze mit natürlicher Intonation so wie den Testsatz *Der Leo säuft* mit monotoner Intonation mehrmals im schallarmen Raum des Instituts für Phonetik in München (Telefunken M15, Geschwindigkeit 19/sec). Die monotone Intonation sollte dabei einen möglichst neutralen Ausgangspunkt für die Manipulation darstellen. Die besten Realisationen wurden auf einer PDP11/50 mit einer Sample-Rate von 20 KHz digitalisiert und mit einem Tiefpaß von 8 KHz gefiltert. Die Kontextsätze blieben unmanipuliert, der Testsatz wurde mit einem Segmentierungsprogramm periodenweise geschnitten und diente damit als Grundlage zur pitchsynchrone Manipulation auf Halbtonbasis. Einer der Stimuli taucht in jedem der Experimente auf. Dieser 'Bezugsstimulus'

wird nun anhand von Fig.1 beschrieben; die anderen Stimuli können leicht von ihm abgeleitet werden. Auf der y-Achse sind in Fig.1 die Halbtonwerte aufgetragen, auf der x-Achse der Zeitverlauf. Der LEVEL-Verlauf des Anfangsteils wurde nicht manipuliert, sondern von der natürlichen Äußerung übernommen. Ebenfalls nicht manipuliert waren die stimmlosen Konsonanten in *säuft*. Im vokalischen Teil von *Leo* erfolgt ein RISE um fünf Halbtöne und, nach einem Gipfel bei der Transition zwischen [e] und [o], ein FALL von sieben Halbtönen. Der Diphthong in *säuft* fällt um zwei Halbtöne. Der Intensitätsverlauf blieb unmanipuliert, d.h. monoton, aber mit den geringfügigen intrinsischen Schwankungen der natürlichen Produktion.

Fig.1: Bezugsstimulus



In Tab.2 sind die Stimulusserien I-V aufgeführt, die in den Experimenten getestet wurden; sie seien nun kurz erläutert. Wir interessierten uns hauptsächlich für den Gipfel als Indikator des Exklamativs und dabei besonders für die **Gipfelerhöhung** und **Gipfelverschiebung**. Der Intensitätsverlauf wurde deshalb nicht verändert. Die Dauer wurde für die Untersuchung des verschobenen Gipfels (Kontinua II-V) kovariierend manipuliert, da sonst zu steile Fo-Verläufe entstanden wären.

Tab.2: Dauer- und Halbtonwerte der Kontinua

		Dauerrelation (RISE/FALL) pro Kontinuum				
		I	II	III	IV	V
Halbtöne						
1 / 3	150/150	-	-	-	-	-
3 / 5	150/150	0/270	90/180	135/135	180/90	
5 / 7	150/150	0/300	100/200	150/150	200/100	
7 / 9	150/150	0/330	110/220	165/165	220/110	
9 / 11	150/150	0/360	120/240	180/180	240/120	
11 / 13	-	0/390	130/260	195/195	260/130	

Im ersten Kontinuum (I) bleibt die Dauer des manipulierten Bereichs in *Leo* konstant auf 300 msec; der Gipfel befindet sich in der Mitte bei der Transition zwischen [e] und [o]. Vor dem Schrägstrich ist immer die Dauer des RISE, danach die des FALL angegeben. Der Gipfel wird in Zwei-Halbton-Schritten von einem Halbton bis neun Halbtöne über dem Ausgangswert angehoben. Man beachte, daß wir die Höhe des Gipfels immer mit dem Wert des RISE angeben; der FALL ist jeweils um 2 Halbtöne größer. Da dieses Kontinuum noch zu wenig im Exklamativbereich lag, wurde in den Kontinua II-V der Gipfel in Zwei-Halbton-Schritten von 3 Halbtönen auf 11 Halbtöne angehoben. Die Dauer kovariert dabei in 30-msec-Schritten zwischen 270 msec und 390 msec. Die durch die pitchsynchrone Manipulation bedingten Schwankungen um +/-5 msec um die in Tab.2 angegebenen Zielwerte sind perceptiv unauffällig. Im Kontinuum IV war der Gipfel analog zum Kontinuum I in der **Mitte**, bei der Transition zwischen [e] und [o]. Im Kontinuum III kam der Gipfel **früh**, nach dem ersten Drittel im [e]; die Transition blieb in der Mitte. Im Kontinuum V kam der Gipfel **spät**, nach dem zweiten Drittel zusammen mit der Transition zwischen [e] und [o]. Die Ergebnisse der ersten Testphase führten zur Konstruktion eines **sehr frühen** Kontinuums II, bei dem der Gipfel noch im [l] lag, d.h. der RISE im [e] hat eine Dauer von Null. Bei diesem letzten Kontinuum blieb der Ausgangswert nicht konstant auf 90 Hz, sondern erhielt jeweils einen LEVEL-Verlauf, der zwei Halbtöne unter dem Wert des Gipfels lag.

HYPOTHESEN: Es wurden keine exakten Alternativhypothesen aufgestellt, sondern eine globale Hypothese: ein 'Mehr' an Ausprägung, an Markiertheit, entspricht einem 'Mehr' an Exklamativität. Das 'Mehr' wird dabei kodiert als Erhöhung des Gipfels bzw. als Verschiebung dieser Gipfelposition aus einer unmarkierten Lage heraus. Eigene, frühere Ergebnisse (vgl. Batliner 1988) lassen weiter vermuten, daß der Fokus beim Exklamativ weniger indiziert wird als bei der Aussage. Weitere Hypothesen, die sich aus eigenen Vorarbeiten ergaben, werden wir jeweils bei den einzelnen Experimenten gesondert erläutern.

VERSUCHSPERSONEN waren Studenten der Germanistik oder Phonetik, die für ihre freiwillige Teilnahme bezahlt wurden.

DESIGN: Es wurden Identifikationstests, Diskriminationstests und Kontexttests durchgeführt. Die beiden ersten Arten werden üblicherweise bei Untersuchungen zur Kategorialität phonetischer Parameter verwendet. Der Kontexttest wurde wegen seiner größeren Nähe zur natürlichen Kommunikationssituation hinzugenommen. Die Versuchspersonen saßen dabei immer im Sprachlabor des Instituts für Phonetik vor einem Abstimmkästchen und hörten die Stimuli über Kopfhörer. Genau nach jedem Stimulus(-paar) wurde die Abstimmung freigegeben; dies wurde durch eine Lampe an jedem Kästchen angezeigt. Die Versuchspersonen hatten dann vier Sekunden Zeit, ihre Antwort zu überlegen und die entsprechende Taste zu drücken. Die Antworten wurden auf einer PDP11/3 gesammelt und zur weiteren Verarbeitung aufbereitet. Beim **Identifikationstest** lautete die Instruktion: "Bitte entscheiden Sie, ob es sich bei diesem Stimulus um eine Aussage oder einen Ausruf handelt, und drücken Sie die linke Taste für 'Aussage' oder die rechte für 'Ausruf'." Das geläufigere 'Ausruf' wurde für die Instruktion dem Fachterminus 'Exklamativ' vorgezogen. Es wurde darauf hingewiesen, daß man sich auch eine nachdrückliche (kontrastierende) Aussage vorstellen kann und daß beim Ausruf ein Merkmal des Erstaunens, der Überraschung hinzukommen muß. Beim **Diskriminationstest** lautete die Instruktion: "Bitte entscheiden Sie, ob die beiden Stimuli innerhalb eines Paares in allen Belangen gleich klingen oder nicht, und drücken Sie dann die linke Taste für 'gleich' oder die rechte für 'verschieden'." Bei den **Kontexttests** mußten die Versuchspersonen ihre Antworten auf einer fünf-stufigen Ratingskala abgeben. Sie entschieden damit, ob die beiden Sätze als Äußerung ein und desselben Sprechers 'sehr gut'=1, 'gut'=2, 'noch gut'=3, 'schlecht'=4 oder 'sehr schlecht'=5 zusammenpaßten. (Diese Skala wurde für die Auswertung und die Darstellung in eine Skala von 0 bis 4 umgerechnet.) Da die Items öfter als einmal vorkamen (vgl. die jeweiligen Angaben zu den einzelnen Experimenten), erhielten wir für die Identifikations- und Diskriminationstests pro Versuchsperson und Item einen Häufigkeitswert für eine der Antwortmöglichkeiten und für die Kontexttests einen Mittelwert aus den Einzelbeurteilungen.

STATISTISCHE AUSWERTUNG: Die statistische Auswertung wurde mit dem Programmpaket SPSSPC+ vorgenommen, d.h. bei erfüllten Voraussetzungen, gegebenenfalls nach Transformation der Daten, mit der dort verfügbaren Prozedur MANOVA. Dabei wurde das multivariate Modell zugrundegelegt. Wenn die Voraussetzungen nicht gegeben waren, so wurde zuerst für die Teilgruppen der Stimuli, die die gleiche Gipfelhöhe aufwiesen, ein nicht-parametrischer Test für k abhängige Stichproben (Kendalls Konkordanzkoeffizient) durchgeführt. War dieser Test signifikant, so wurde mit dem Wilcoxon-Test paarweise verglichen. Zur Vermeidung eines Fehlers der ersten Art wurde dabei immer das Signifikanzniveau nach der Formel

$$1 - (1 - \alpha)^{1/c}$$

abgesenkt, wobei α für ein Signifikanzniveau von 5% steht und c für die Anzahl der durchgeführten Tests.

TERMINOLOGIE: Wie schon erwähnt, verwenden wir für die untersuchten Konstellationen statt linguistisch exakter Bezeichnungen abkürzende, die zugleich die Faktoren und ihre Klassen im varianzanalytischen Design bezeichnen; sie sind im weiteren Text in diesem Zusammenhang mit Großbuchstaben geschrieben:

MODUS (AUSSAGE, KONTRAST, EXKLAMATIV);
 FOKUS (FRÜH, SPÄT);
 POSITION (SEHR FRÜH, FRÜH, MITTE, SPÄT);
 HALBTON (1, 3, 5, 7, 9, 11).

Die Tests werden ebenfalls von nun an mit einer abkürzenden Schreibweise bezeichnet: 'IT' steht für 'Identifikationstest', 'DT' für 'Diskriminationstest' und 'KT' für 'Kontexttest'.

3. DIE EXPERIMENTE

An Exp.1 und Exp.3 sowie an Exp.2 und Exp.4 nahmen je 20 Versuchspersonen teil, die auf zwei Gruppen aufgeteilt wurden; der einen Gruppe wurde zuerst der IT, dann der KT (aufgeteilt auf zwei Sitzungen) dargeboten, die andere durchlief die Tests in umgekehrter Reihenfolge. Die Tests fanden in drei aufeinanderfolgenden Wochen zum gleichen Termin statt.

EXPERIMENT 1

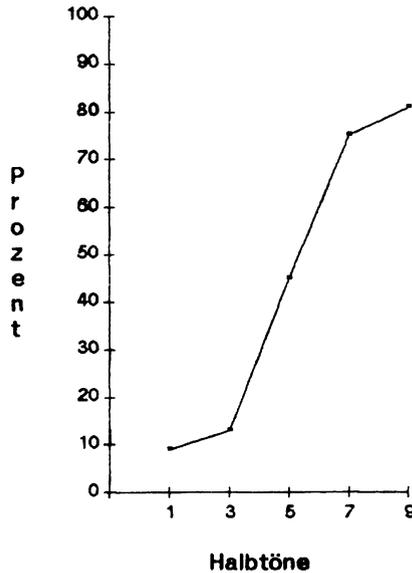
FRAGESTELLUNG: Es sollte untersucht werden, ob eine Gipfelerhöhung ohne gleichzeitige Dehnung im IT eine (kategoriale) Grenze zwischen AUSSAGE und EXKLAMATIV ergibt.

MATERIAL UND DESIGN: Die fünf Stimuli aus dem Kontinuum I mit einem MITTLEREN Gipfel von einem bis neun Halbtönen ohne kovariierende Dauervariation wurden in einem IT je zehnmal randomisiert dargeboten.

ERGEBNIS: Fig.2 zeigt die Zuweisungen zum EXKLAMATIV, wobei auf der y-Achse die Antworten in Prozent, gemittelt über die Versuchspersonen, und auf der x-Achse die Stimuli aufgetragen sind. Die 50%-Grenze liegt im Bereich des Stimulus mit einem Gipfel von 5 Halbtönen; dieser Stimulus bildet deshalb einen 'Schnittstimulus' (der Identifikationskurve mit der 50%-Grenze). Die Eckstimuli liegen bei 9% bzw. 80.5%. Die Identifikation kann also als schwach kategorial bezeichnet werden. Eine ausgeprägte Kategorialität wäre gegeben, wenn die Kurve steiler verlaufen würde und zumindest die Eckstimuli nahe bei 0% bzw. 100 % liegen würden. (Zum Begriff 'Kategorialität' vgl. Repp 1984; wir verwenden hier diesen Begriff allerdings in einer nicht strikt definierten Weise, vgl. die Diskussion weiter unten.) Zwei systematische Erklärungen für den relativ niedrigen Wert des 9-HALBTON-Stimulus bieten sich an: zum einen ist auch eine

emphatische Aussage mit einem hohen Gipfel innerhalb unseres Satzmodussystems möglich, zum anderen mag der 9-HALBTON-Stimulus, da er nicht zusätzlich gedehnt ist, noch nicht im typischen Exklamativ-Bereich liegen.

**Fig.2: Identifikationstest:
Exklamativzuweisung**

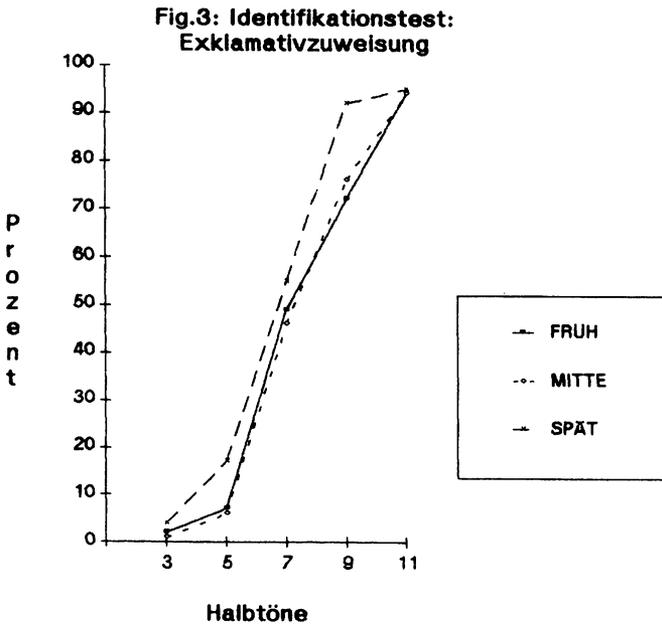


EXPERIMENT 2

FRAGESTELLUNG: Die Fragestellung von Exp.1 wurde erweitert und modifiziert: anhand von drei unterschiedlichen Gipfelpositionen soll untersucht werden, welche dieser Positionen eher exklamativtypisch ist. Die Kontinua sollten durch eine Verschiebung im 'Range'-Bereich (3-11 HALBTÖNE) und eine gleichzeitige kovariierende Dehnung den EXKLAMATIV besser indizieren können: wie aus Tab.2 ersichtlich ist, geht eine Erhöhung um zwei HALBTÖNE immer einher mit einer Dehnung um 30 msec.

MATERIAL UND DESIGN: Analog zu Exp.1 wurden nun die Stimuli aus den Kontinua III, IV und V mit einem Gipfel von 3 bis 11 HALBTÖNEN und FRÜHER, MITTLERER und SPÄTER Gipfelposition in einem IT je fünfmal dargeboten. Wir erwarteten dabei zumindest beim Eckstimulus mit dem SPÄTEN Gipfel von 11 HALBTÖNEN eine eindeutige und insgesamt beim SPÄTEN Kontinuum eine frühere Exklamativzuweisung.

ERGEBNIS: Fig.3 zeigt für die drei Positionen grundsätzlich den gleichen Verlauf, wobei die Eckstimuli diesmal eindeutiger, nämlich mit 1%, 2% und 4% bzw. 94%, 94% und 95 % als EXKLAMATIV kategorisiert werden. Kendalls Konkordanzkoeffizient zeigt für die Positionen einen signifikanten Unterschied bei den Stimuli mit einem Gipfel von 5 und 9 HALBTÖNEN. Ein paarweiser Vergleich mit einem Wilcoxon-Test ergab für das abgesenkte Signifikanzniveau von .014 bei einseitiger Fragestellung und dem 9-HALBTON-Gipfel einen signifikanten Unterschied zwischen dem SPÄTEN und dem MITTLEREN Gipfel ($p < .003$) und zwischen dem SPÄTEN und dem FRÜHEN Gipfel ($p < .006$). Damit ist unsere Hypothese, daß ein SPÄTER Gipfel eher Exklamativität indiziert, für den Bereich von 9 HALBTÖNEN bestätigt. Zwei Unterschiede zu Exp.1 können beobachtet werden:



(1) Die Eckstimuli werden eindeutiger beurteilt - sie entsprechen auch eher den beiden unterschiedlichen Modi: die 3-HALBTON-Stimuli in Exp.2 sind um 30 msec kürzer als ihr Pendant in Exp.1 und die 11-HALBTON-Stimuli sind 90 msec länger und zwei HALBTÖNE höher als der entsprechende Eckstimulus in Exp.1. Bei den Eckstimuli ist kein Unterschied in der Gipfelposition zu beobachten. Offensichtlich kennzeichnet hier die Gipfelhöhe allein schon hinreichend die Modi.

(2) Die 50%-Grenze liegt für alle drei Positionen im Bereich des 7-HALBTON-Stimulus - obwohl die kovariierende Dauermanipulation einen früheren Schnittpunkt erwarten ließe. Diese Frage wird weiter unten in Exp.5 aufgegriffen.

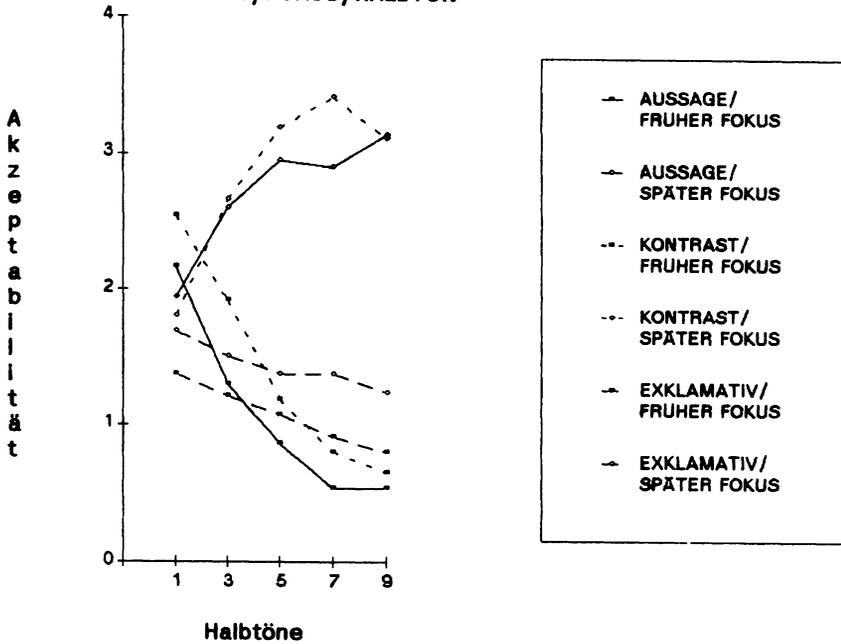
EXPERIMENT 3

FRAGESTELLUNG: Es soll untersucht werden, ob sich im KT eine zum IT analoge Kategoriengrenze ergibt und ob sich die Modi bei der Fokusindizierung unterscheiden. Der Aussagemodus wurde zusätzlich unterteilt in normale Aussage (AUSSAGE) und Aussage mit Kontrastakzent (KONTRAST), um herauszufinden, ob der Kontrastakzent anders bewertet wird als der normale Hauptakzent.

MATERIAL UND DESIGN: Paare aus allen sechs Kontextsätzen (vgl. Tab.1) und den Stimuli aus Kontinuum I wurden in einem KT je viermal randomisiert dargeboten. Damit ergab sich ein Versuchsplan mit den Faktoren MODUS (AUSSAGE, KONTRAST und EXKLAMATIV), FOKUS (FRÜH und SPÄT) sowie HALBTON (1, 3, 5, 7, 9). Wir erwarteten, ausgehend von der in Batliner (1988) beschriebenen Analyse, bei AUSSAGE und KONTRAST klare, bei EXKLAMATIV nur rudimentäre Fokusedifferenzierung.

ERGEBNIS: In Fig.4 sind auf der x-Achse die Stimuli und auf der y-Achse die Mittelwerte der Akzeptabilitätsurteile aufgetragen; '0' steht für 'passen sehr gut zusammen', '1' für 'gut', '2' für 'noch gut', '3' für 'schlecht' und '4' für 'sehr schlecht'. Die Varianzanalyse ergab signifikante Haupteffekte und Interaktionen mit Ausnahme der Interaktion MODUS mit HALBTON (vgl. Tab.3). Für die Interpretation kann also auf die einfachen Haupteffekte zurückgegriffen werden. D.h. daß z.B. der Faktor MODUS nicht gemittelt über die beiden FOKI und alle fünf HALBTON-Höhen betrachtet wird, sondern für jeden FOKUS und jede HALBTON-Höhe getrennt. Das gleiche gilt für die beiden anderen Faktoren. In Tab.3 sind die signifikanten Ergebnisse für die einfachen Haupteffekte bei MODUS und FOKUS mit einem Stern gekennzeichnet. Die HALBTÖNE unterscheiden sich signifikant bei AUSSAGE und KONTRAST, nicht aber beim EXKLAMATIV.

Fig.4: Kontexttest:
MODUS/FOKUS/HALBTON



Tab.3: Haupteffekte, Interaktionen und einfache Haupteffekte von Exp.3

Faktoren/Interaktionen:	df	F	Sig.
MODUS	2,18	12.33	.000
FOKUS	1,19	65.09	.000
HALBTON	4,16	6.98	.002
MODUS mit FOKUS	2,18	30.15	.000
MODUS mit HALBTON	8,12	2.54	.071
FOKUS mit HALBTON	4,16	18.50	.000
MODUS mit FOKUS mit HALBTON	8,12	11.07	.000

einfache Haupteffekte:	bei HALBTON				
	1	3	5	7	9
FOKUS bei AUSSAGE	*	*	*	*	*
FOKUS bei KONTRAST	*	*	*	*	*
FOKUS bei EXKLAMATIV	*	*		*	*
MODUS bei FRÜHEM FOKUS	*	*			
MODUS bei SPÄTEM FOKUS		*	*	*	*

Fokusedifferenzierung: Bei AUSSAGE und besonders bei KONTRAST indiziert ein Gipfel von 3 oder mehr HALBTÖNEN einen Fokus auf *Leo*. Offenbar genügt ein Gipfel von 5 HALBTÖNEN für die Fokusindizierung, da sich bei den höheren Gipfeln die Kurvenverläufe nicht mehr wesentlich ändern. Die Verläufe von AUSSAGE und KONTRAST sind sehr ähnlich; das entspricht unseren Erwartungen, da es sich in beiden Fällen um denselben Modus handelt. Für den Kontrastakzent wird üblicherweise eine lokal stärker ausgeprägte Markierung angenommen (vgl. Bannert 1985:300). Anhand unseres Ergebnisses läßt sich der Kontrastakzent vom normalen Hauptakzent nur beim Stimulus mit einem Gipfel von einem Halbton unterscheiden, der bei der AUSSAGE indifferent ist, beim KONTRAST aber signifikant eher einen SPÄTEN FOKUS auf *säuft* indiziert. Beim EXKLAMATIV verlaufen die Kurven für FRÜHEN und SPÄTEN FOKUS ganz anders als bei den Aussagen, nämlich parallel und leicht abfallend. Der SPÄTE FOKUS wird durchgehend schlechter bewertet. Dieses Ergebnis stimmt mit Batliner (1988) überein, wo sich auch nur eine rudimentäre Fokusedifferenzierung beim EXKLAMATIV zeigt. Die Differenzierung sieht so aus, daß grundsätzlich bei einem Kontext, der einen Fokus auf *säuft* erwarten läßt, alle Stimuli mit ihren weniger oder mehr ausgeprägten Akzenten auf *Leo* schlechter bewertet werden. Die Differenzierung führt aber nicht dazu, daß in so einem Fall ein deutlicherer Gipfel auf *Leo* wie bei den Aussagen eine entsprechend schlechtere Bewertung hervorruft.

Modusedifferenzierung: Auf die Differenzierung **innerhalb** der Aussagemodi (AUSSAGE vs. KONTRAST) wurde schon im letzten Abschnitt eingegangen; es gibt keinen Unterschied im grundsätzlichen Verlauf der Kurven. Anders bei der Differenzierung **zwischen** den Modi: beim FRÜHEN FOKUS setzt die EXKLAMATIV-Kurve deutlich tiefer an, d.h. diese Kombination wird besser bewertet als die entsprechenden Aussagekombinationen; die Kurve fällt dann aber nicht so steil ab und kreuzt sogar die der Aussagen. Die Verbesserung mit zunehmender Gipfelhöhe ist also beim EXKLAMATIV bei weitem nicht so ausgeprägt wie bei den Aussagen – ein Ergebnis, das auf den ersten Blick kontraintuitiv aussieht, da wir ja davon ausgehen, daß sich der EXKLAMATIV durch ein 'Mehr', also z.B. durch ein 'Mehr' an Gipfelhöhe auszeichnet. Wir nehmen zur Erklärung zwei Faktoren an, deren jeweiligen Stellenwert wir im Augenblick allerdings nicht gegeneinander abwägen können:

(1) Der EXKLAMATIV ist weniger lokal als vielmehr **global** markiert. Die Exklamativität, der Ausdruck des Erstaunens, muß nicht auf dem Testsatz liegen, sondern wird auch schon im Kontext durch das *Mann*- indiziert; der Testsatz kann also auch nur noch einen intonatorischen 'Nachlauf' ohne besondere

Hervorhebung bilden. Dieser Fall kann bei den Stimuli mit einem niedrigen Gipfel vorliegen. (Es ließe sich nun einwenden, daß in einem solchen Fall der Testsatz eben kein Exklamativsatz sei, sondern ein Aussagesatz. Der Testsatz wäre aber in diesem Kontext und ohne exklamativindizierende intonatorische Hervorhebung statt mit Verb-Zweit-Stellung auch mit Verb-Erst-Stellung denkbar: (*Mann -) Säuft der Leo.* Konsequenterweise wäre dieser Satz dann als Fragesatz zu betrachten, da er aufgrund der Verb-Stellung nur Exklamativsatz oder Fragesatz sein kann - eine etwas seltsam anmutende Lösung. Möglicherweise ergibt sich für dieses Phänomen eine allseits zufriedenstellende Erklärung erst dann, wenn man als Beschreibungseinheit nicht den Satz, sondern die Äußerung ansetzt.)

(2) Wir befinden uns nicht in einem für den EXKLAMATIV typischen Parameterbereich, da die Dauer ja gleich blieb.

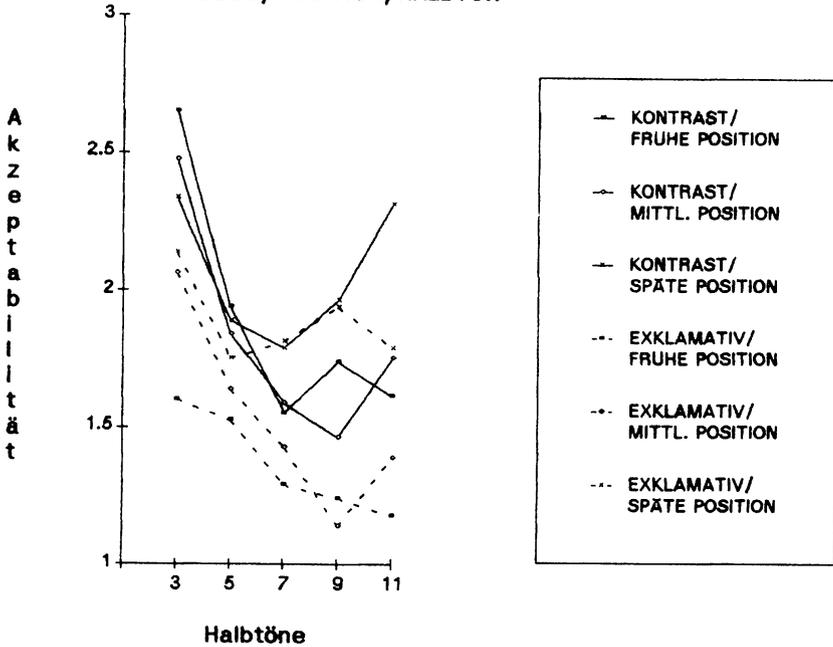
Die Globalität des EXKLAMATIVS ist auch dafür verantwortlich, daß man keinen direkten Vergleich zwischen dem IT und dem KT ziehen kann - man sucht ja vergeblich nach einem Pendant zum Schnittstimulus von ca. 5 HALBTÖNEN im Exp.1. Im IT gab es nur ein Indiz, nur einen Parameter, der verschiedene Werte annimmt: die Gipfelhöhe. In einem solchen Fall wird natürlich dieses 'Mehr' an Ausprägung in einem 'forced choice test' auch mit einem 'Mehr' an Exklamativität gleichgesetzt. Beim KT kann aber auch der Kontext schon eine exklamativindizierende Funktion übernehmen.

EXPERIMENT 4:

FRAGESTELLUNG: Es soll untersucht werden, ob sich die ausgeprägtere Kategorialität der Differenzierung zwischen AUSSAGE und EXKLAMATIV sowie die bessere Indizierung des EXKLAMATIVS durch einen SPÄTEN Gipfel nicht nur im IT, sondern auch im KT nachweisen läßt.

MATERIAL UND DESIGN: Die beiden Kontexte KONTRAST/FRÜHER FOKUS und EXKLAMATIV/FRÜHER FOKUS wurden mit den Kontinua III, IV und V gepaart und in einem KT je viermal randomisiert dargeboten. Daraus ergab sich ein Versuchsplan mit den Faktoren MODUS (KONTRAST, EXKLAMATIV), POSITION (FRÜH, MITTE und SPÄT) sowie HALBTON (3,5,7,9,11). Die Kontinua decken einen anderen Bereich als das Kontinuum I ab, damit auf alle Fälle auch exklamativtypische Konturen enthalten sind; dem diente auch die kovariierende Dauermanipulation (Dehnung in 30 msec-Schritten von 270 bis 390 msec auf [eo]).

**Fig.5: Kontexttest:
MODUS/POSITION/HALBTON**



ERGEBNIS: Der erste Eindruck von Fig.5 mag täuschen, da hier, um die Unterschiede deutlicher zu machen, nur eine Teilskala von 1-3 für die y-Achse gewählt wurde. Die Urteile sind also indifferenter, es fehlen 'sehr gute' und 'sehr schlechte' Bewertungen. Auf der einen Seite hat es sich zwar schon in Exp.3 angedeutet, daß beim MODUS keine so ausgeprägten Differenzierungen zu erwarten sind wie beim FOKUS. Auf der anderen Seite ist es aber doch unerwartet, daß die Stimuli mit einem Gipfel von 9 und 11 HALBTÖNEN, die durch die kovariierende Dauermanipulation noch exklamativtypischer werden sollten, mit dem Exklamativkontext relativ schlecht beurteilt wurden, schlechter als der Eckstimulus in Exp.3. U.E. handelt es sich hierbei um einen kombinierten Effekt der spezifischen Art des Tests und der Art der Materialmanipulation und -zusammenstellung: Es wurde ein zwar resynthetisierter, aber sehr gut klingender Kontextsatz gekoppelt mit einem schematisch manipulierten Testsatz; damit ergab sich vor allen Dingen bei den hohen Gipfeln mit starker Dehnung eine gewisse Diskrepanz im globalen Zeitbereich. Da den Versuchspersonen bei dieser Art des Tests natürlich freisteht, bestimmte Kombinationen grundsätzlich weniger akzeptabel zu finden - aber nicht aus den vom Versuchsleiter intendierten

Gründen -, so können, wie in diesem Fall, weniger trennscharfe Kurven entstehen. Entscheidend ist aber, ob sich trotzdem signifikante und sinnvoll interpretierbare Unterschiede ergeben. Tab.4 zeigt das Ergebnis der Varianzanalyse: wieder sind alle Haupteffekte und zwei der Interaktionen signifikant; wir greifen also auf die einfachen Haupteffekte zurück. Wie bei Exp.3 sind die signifikanten Ergebnisse für die einfachen Haupteffekte bei MODUS und POSITION mit einem Stern gekennzeichnet. Die HALBTÖNE unterscheiden sich beim KONTRAST signifikant in allen drei POSITIONEN, beim EXKLAMATIV nur in der MITTLEREN. Wir beschränken uns im folgenden auf eine Interpretation der Fig.6-8, anhand derer die wichtigsten Punkte gut herausgearbeitet werden können.

Tab.4: Haupteffekte, Interaktionen und einfache Haupteffekte von Exp.4

Faktoren/Interaktionen:	df	F	Sig.
MODUS	1,19	10.79	.004
POSITION	2,18	7.78	.004
HALBTON	4,16	6.60	.002
MODUS mit POSITION	2,18	6.67	.007
MODUS mit HALBTON	4,16	2.68	.069
POSITION mit HALBTON	8,12	3.85	.018
MODUS mit POSITION mit HALBTON	8,15	1.95	.142
einfache Haupteffekte:			
	bei HALBTON	3	5 7 9 11
MODUS bei FRÜHEM GIPFEL		*	* *
MODUS bei MITTLEREM GIPFEL			* *
MODUS bei SPÄTEM GIPFEL			*
POSITION bei KONTRAST			*
POSITION bei EXKLAMATIV		*	* * *

Fig.6 zeigt die Mittelwerte für die Interaktion POSITION mit MODUS bei kombiniertem HALBTON. Die FRÜHE POSITION ist beim EXKLAMATIV deutlich besser als die anderen; wie Fig.5 zeigt, sind dafür besonders die niedrigeren Gipfelhöhen verantwortlich. Die SPÄTE POSITION wird am schlechtesten bewertet. Das gilt auch für den KONTRAST, wobei sich hier FRÜHER und MITTLERER Gipfel nicht so stark unterscheiden.

Fig.7 zeigt die Interaktion HALBTON mit MODUS bei kombinierter POSITION. Beim EXKLAMATIV ist der Verlauf ähnlich wie in Exp.3 für beide Fokuspositionen, nämlich leicht fallend. Beim KONTRAST ist er, wieder analog zu Exp.3, anfangs stärker fallend; dann folgt aber ein Wendepunkt beim Stimulus mit dem Gipfel

von 7 HALBTÖNEN. Bei sieben HALBTÖNEN liegt aber auch der Schnittstimulus von Exp.2, d.h. hier zeigt sich auch im KT die Grenze zwischen KONTRAST und EXKLAMATIV, wenn Gipfelerhöhung und -dehnung zusammen eingesetzt werden. Bei solchen Stimuli werden die höheren Gipfel als für den KONTRAST nicht mehr adäquat bewertet.

Fig.8 zeigt die Interaktion HALBTON mit POSITION bei kombiniertem MODUS. Ab 7 HALBTÖNEN ist der Verlauf für den SPÄTEN Gipfel schlechter. Die Verschlechterung ist am deutlichsten beim KONTRAST, vgl. Fig.5. Das wurde erwartet, da der SPÄTE Gipfel Exklamativität indizieren soll. Aber auch beim EXKLAMATIV wird der SPÄTE Gipfel deutlich schlechter bewertet als die beiden anderen Positionen. Dieses Ergebnis steht in Widerspruch zu Exp.2, bei dem der SPÄTE Gipfel signifikant früher Exklamativität indiziert als die beiden anderen. U.E. liegt die Erklärung dafür in einer Diskrepanz im Zeitbereich, wie sie oben für die weniger ausgeprägte Trennschärfe in diesem Experiment überhaupt gegeben wurde, allerdings in einer spezielleren Diskrepanz, nämlich in einem schlechten 'timing' bei der Manipulation des Gipfels im Verhältnis zu den beiden vokalischen Elementen [e] und [o]. Damit wird bewirkt, daß der SPÄTE Gipfel per se als unnatürlich eingestuft wird und das natürlich besonders dann, wenn dieser Gipfel prominent ist, wie bei den höheren Gipfeln im Kontinuum. (Wir können noch nicht angeben, welches 'timing' nötig wäre, um einen besser klingenden Stimulus zu erzeugen, wollen dieser Frage aber in einer eigenen Teiluntersuchung nachgehen.) Der vermeintliche Widerspruch zu Exp.2 erklärt sich also wieder aus der unterschiedlichen Aufgabenstellung von IT und KT: bei einem 'forced choice' muß auch ein schlechter Stimulus akzeptiert, d.h. einer der beiden Kategorien zugeordnet werden. Entsprechend unserer Hypothese wird dann der SPÄTE Gipfel eher dem EXKLAMATIV zugeordnet. Bei einem KT steht es den Versuchspersonen frei, die Kombination zu bewerten, und wenn eine Stimulusserie weniger gut generiert ist, so wird sie einfach durchgehend schlechter bewertet, egal, ob sie mit einem KONTRAST-Kontext oder einem EXKLAMATIV-Kontext gepaart ist.

Fig.6: Haupteffekte bei
komb. HALBTON

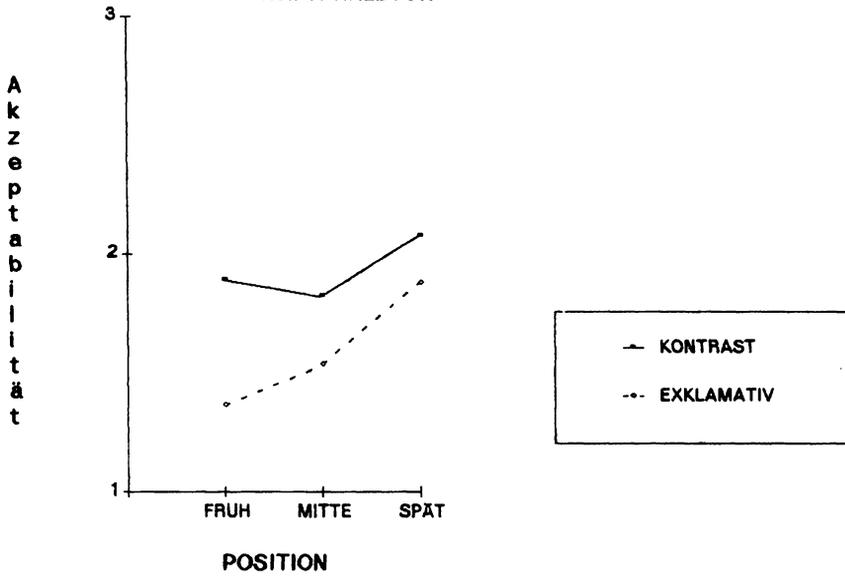


Fig.7: Haupteffekte bei
komb. POSITION

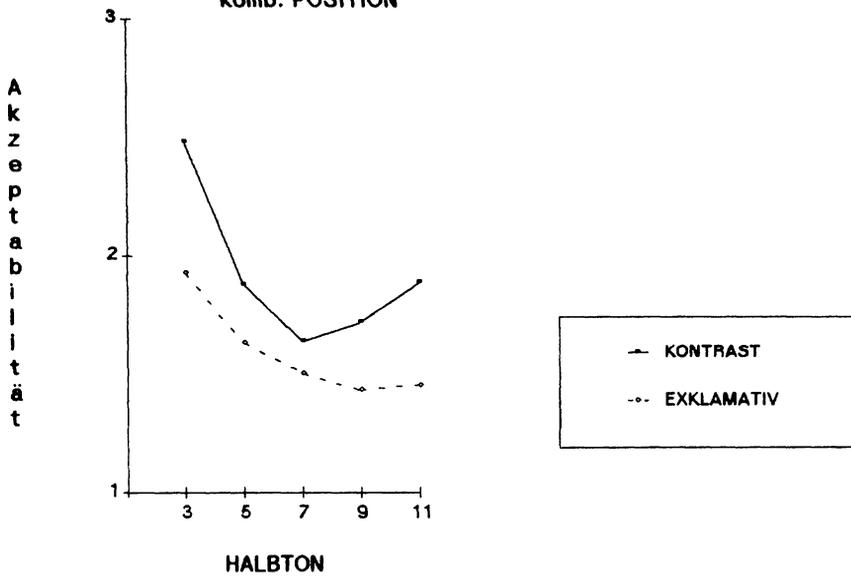
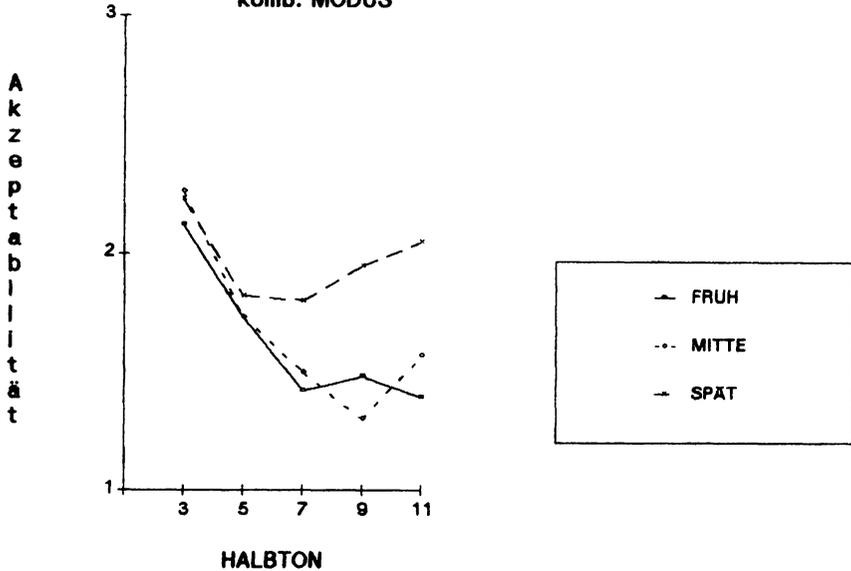


Fig.8: Haupteffekte bei komb. MODUS



4. FOLGEEXPIMENTE

Mit anderen Gruppen von Versuchspersonen wurden weitere Experimente durchgeführt, um bestimmte Fragen, die sich aus Exp.1-4 ergaben, klären zu können.

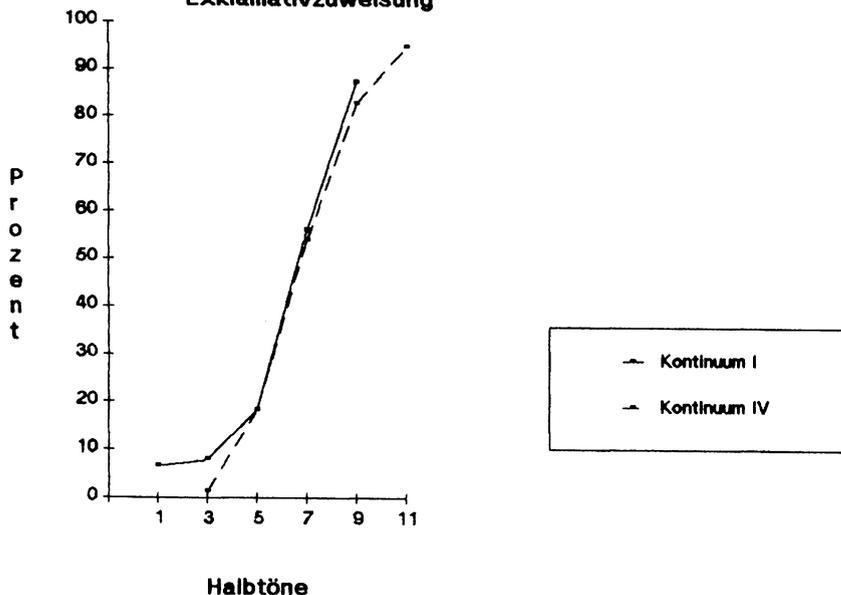
EXPERIMENT 5

FRAGESTELLUNG: Dieses Experiment sollte überprüfen, ob sich unterschiedliche Schnittstimuli für die Kontinua I und IV im IT ergeben, wenn beide Kontinua zusammen getestet werden.

MATERIAL UND DESIGN: 15 Versuchspersonen nahmen an einem IT teil, bei dem das Kontinuum I und das Kontinuum IV mit 10 randomisierten Wiederholungen pro Stimulus dargeboten wurden. (Da der Bezugsstimulus beiden Kontinua angehört, kam er zwanzigmal vor.) Beim Kontinuum I wird nur der Gipfel angehoben, die Dauer bleibt unverändert. Beim Kontinuum IV werden Gipfelhöhe und Dauer variiert. Die 3-, 7- und 9-HALBTON-Stimuli unterscheiden sich also nur in der Dauer, der 5-HALBTON-Stimulus gehört zu beiden Kontinua.

ERGEBNIS: In Fig.9 bestätigt der geringe Unterschied zwischen den Kontinua bei einer Gipfelhöhe von 5, 7 und 9 HALBTÖNEN, daß zumindest für diese Schrittgrößen die Gipfelerhöhung relevanter ist als die Dehnung. (Die Dauerdifferenz beträgt zwischen 0 und 60 msec, vgl. Tab.2). Bei beiden Kontinua liegt der Schnittstimulus kurz unterhalb des 7-HALBTON-Stimulus. Ein Designeffekt der Art, daß die Versuchspersonen die Kategoriengrenze in die Mitte zwischen den Eckstimuli, also zwischen die 5- und 7-HALBTON-Stimuli legen, ist nicht auszuschließen. Wenn wir weiter annehmen, daß der Fo-Range und damit auch in gewissem Rahmen die Grenze zwischen AUSSAGE und EXKLAMATIV bei der Produktion sprecherspezifisch variieren können und daß der Hörer - wieder in einem gewissen Umfang - sich bei der Kalibrierung der intonatorischen Parameter darauf einstellt, so läßt sich zwischen dem Designeffekt und den Bedingungen der natürlichen Kommunikation eine Verbindung herstellen: in beiden Fällen gibt es keine exakte Grenze in absoluten Werten, sondern nur eine relativ zum dargebotenen Fo-Range.

**Fig.9: Identifikationstest:
Exklamativzuweisung**



EXPERIMENT 6 UND 7

FRAGESTELLUNG: Der wesentliche Punkt bei Überlegungen zur Kategorialen Wahrnehmung von Intonationsphänomenen läßt sich wie folgt skizzieren: Im festen Paradigma der Kategorialen Wahrnehmung wird davon ausgegangen, daß der Mensch nur dann einen Unterschied zwischen zwei Stimuli wahrnimmt, wenn sie verschiedenen Kategorien angehören. Damit wird implizit angenommen, daß die Antworten der Versuchspersonen im Experiment ihre **Wahrnehmung** wiedergeben und nicht ihre **Klassifizierung** der Stimuli. Je mehr Zeit zwischen Präsentation der Stimuli und Reaktion der Versuchspersonen verstreicht, desto unwahrscheinlicher ist diese Annahme - bewiesen werden im eigentlichen Sinne konnte sie auch für ganz kurze Stimuli und Reaktionszeiten nicht. Sie ist im Gegenteil auch bei diesen Fällen fragwürdig (vgl. im einzelnen Massaro/Cohen 1983). Bei unserem Material vergehen zwischen Präsentation der Stimuli und Antwort der Versuchsperson mindestens zwei Sekunden - ergo wird die Annahme für solche Intonationsphänomene noch unwahrscheinlicher. Eine eingehende Auseinandersetzung mit dem Paradigma der Kategorialen Wahrnehmung findet sich in Schiefer/Batlner (1988).

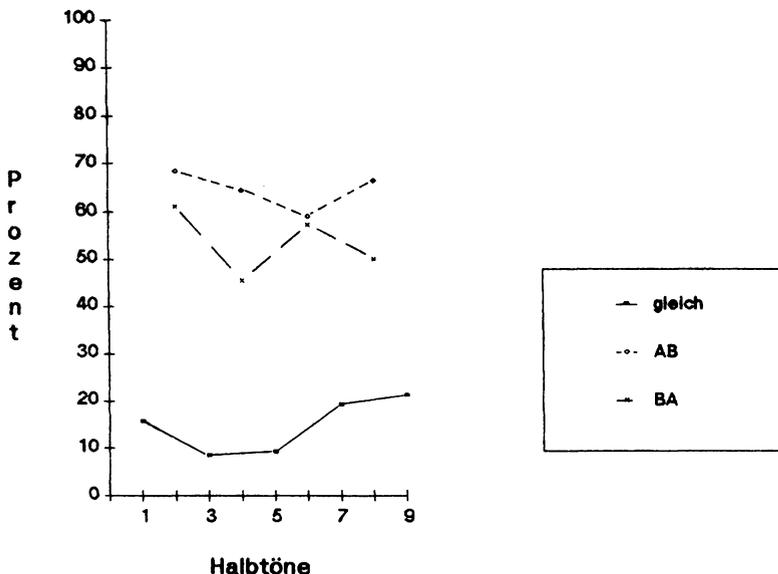
Auf der anderen Seite bedeuten diese Argumente gegen das strikte Paradigma natürlich nicht, daß man den Begriff 'Kategorialität' abschafft. Auch wenn man annimmt, daß in solchen Tests immer klassifiziert und nie nur wahrgenommen wird, daß aber dann, wenn die beiden Stimuli unterschiedlichen Kategorien angehören, die Diskrimination erleichtert wird, kann es sinnvoll sein, einen DT zur Überprüfung dieser Annahme durchzuführen.

MATERIAL UND DESIGN: Um nun zu überprüfen, ob sich eine Kategoriengrenze zwischen AUSSAGE und EXKLAMATIV auch bei der Diskrimination zeigt, wurde getrennt ein DT mit den Kontinua I und IV und jeweils fünfmaliger randomisierter Wiederholung pro Anordnung (AB, BA, AA, BB) durchgeführt, an dem 11 (Kontinuum I) bzw. 13 (Kontinuum IV) Versuchspersonen teilnahmen. Da die Paare im 'Einerschritt' zusammengestellt waren, gab es innerhalb der Anordnung AB bzw. BA immer eine Differenz von 2 HALBTÖNEN; hinzu kam beim Kontinuum IV die Dauerdifferenz.

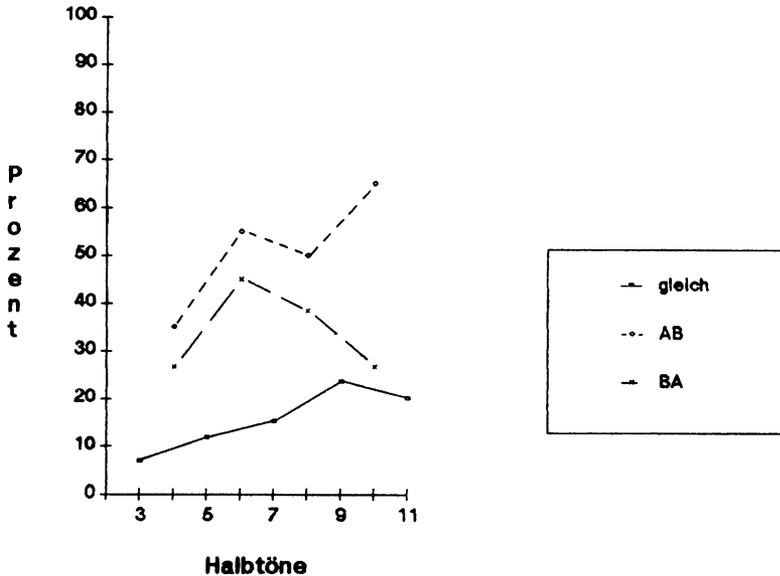
ERGEBNIS: Während sich in Fig.10 beim Kontinuum I für die Stimuli ohne kovariierende Dauervariation kein Diskriminationsgipfel zeigt, liegt in Fig.11 für das Kontinuum IV mit Dauervariation bei der Anordnung BA ein Diskriminationsgipfel zwischen dem 5- und dem 7-HALBTON-Stimulus vor. Bei dieser Anordnung folgt der Stimulus mit dem niedrigeren F_0 -Gipfel im Paar dem Stimulus

mit dem höheren Fo-Gipfel. Dieser Gipfel liegt beim Schnittstimulus aus den Exp.1, 2 und 5. Bei der anderen Anordnung AB zeigt sich zwar auch ein solcher Gipfel, aber zugleich ein weiterer, noch nicht sinnvoll erklärbarer Gipfel am rechten Ende des Kontinuums, der sich bei der Anordnung BA nicht wiederfindet. Eine aus den beiden Anordnungen AB und BA gemittelte Kurve würde denn auch keinen ausgeprägten Gipfel zeigen. Wir können dieses Ergebnis hier nicht eingehender diskutieren; in Schiefer/Batliner (1988) vermuten wir, daß die Anordnungen AB und BA den gleichen Status wie die Schrittgröße zwischen den Stimuli haben. D.h. wir können in diesem Fall nur die weniger prominente Anordnung BA diskutieren, die einen deutlicheren Gipfel aufweist. Man kann also sagen, daß wir im DT beim Kontinuum IV eine, wenn auch ebenso wie im IT schwach ausgeprägte, Kategorialität von AUSSAGE und EXKLAMATIV wiederfinden - vorausgesetzt, wir begnügen uns introspektiv mit der Übereinstimmung des Diskriminationsgipfels mit dem Schnittstimulus der Identifikation und versuchen nicht, die Kategorialität der Wahrnehmung im Sinne des Paradigmas nachzuweisen. Der fehlende Gipfel bei den Diskriminationskurven in Fig.10 ist ein weiterer Hinweis darauf, daß für einen kategorialen Übergang von AUSSAGE zu EXKLAMATIV eine Dehnung der Hauptakzentsilbe unerlässlich ist.

**Fig.10: Diskriminationstest:
Kontinuum I**



**Fig.11: Diskriminationstest:
Kontinuum IV**



5. EIN WEITERES KONTINUUM: DER 'SEHR FRÜHE' GIPFEL

Kohler (1987) hat auf der Grundlage des natürlich produzierten Satzes *Sie hat ja gelogen* ein Kontinuum mit 11 Stimuli erzeugt, bei dem der Gipfel von einer prä vokalischen Position im [l] zu einer späten Position im [o:] von *gelogen* wandert. Er schließt aus den mit diesem Kontinuum durchgeführten Experimenten Folgendes (Kohler 1987:136):

Das Gehör spricht ganz offensichtlich kategorial auf die Positionierung des F₀-Gipfels relativ zur Grenze zwischen Silbenperipherie und Silbenkern [...] an. Und diese kategoriale Differenzierung wird im Deutschen semantisch genutzt für eine "begründende Äußerung am Ende einer Argumentationskette" gegenüber dem "Beginn einer Argumentation" diesseits und jenseits der Grenze in der 11-er Serie von links nach rechts. Eine weitere Verschiebung nach rechts erhöht den Grad der Emphase/Empörung.

Auch wenn bei Kohler die Gipfelhöhe immer konstant blieb und sowohl seine linguistischen Vorgaben als auch seine Art der Manipulation nur bedingt mit unserem Vorgehen vergleichbar sind, so bildet doch die Verschiebung des Gipfels aus einer frühen in eine späte Position ein *tertium comparationis* - allerdings

ein unvollständiges, da Kohlers früher Gipfel zwischen Konsonant und Vokal in unseren bisherigen Kontinua nicht auftritt: unser FRÜHER Gipfel liegt schon 100 msec im Vokal. Damit wäre auch unser weniger eindeutiges Ergebnis bei Exp.4 zu erklären, wenn man annimmt, daß im Material keine Kategoriengrenze enthalten ist. Um diese Frage zu überprüfen, wurde ein zusätzliches Kontinuum (II in Tab.2) generiert, bei dem der Gipfel der Fo-Kontur noch im [l] liegt und [eo] einen Fall aufweist.

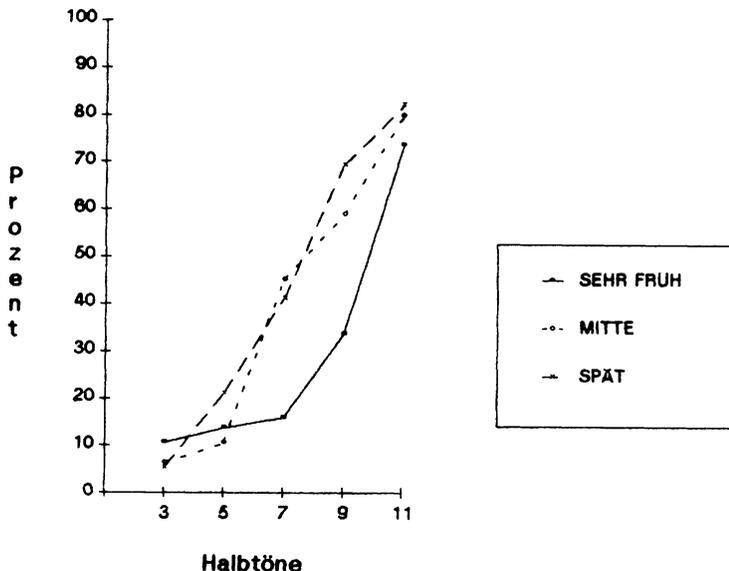
EXPERIMENT 8

FRAGESTELLUNG: Es sollte untersucht werden, ob sich die Annahme, daß ein SEHR FRÜHER Gipfel eher eine AUSSAGE als einen EXKLAMATIV indiziert, im IT bestätigt.

MATERIAL UND DESIGN: Die Stimuli mit einem SEHR FRÜHEN, MITTLEREN und SPÄTEN Gipfel (Kontinua II, IV und V in Tab.1) wurden in einem IT je fünfmal randomisiert 19 Versuchspersonen dargeboten. Sie sollten entscheiden, ob es sich bei den Stimuli um eine Aussage oder einen Ausruf handelt.

ERGEBNIS: Kendalls Konkordanzkoeffizient zeigt für das in Fig.12 dargestellte Ergebnis signifikante Unterschiede bei den Stimuli mit einem Gipfel von 7 und 9 HALBTÖNEN. Für den paarweisen Vergleich nach Wilcoxon bei einer einseitigen Fragestellung wurde das Signifikanzniveau auf .017 abgesenkt. Der SEHR FRÜHE Gipfel unterschied sich bei beiden Gipfelhöhen signifikant vom MITTLEREN und SPÄTEN Gipfel. Für 7 Halbtöne galt: SEHR FRÜH vs. MITTE: $p < .010$, SEHR FRÜH vs. SPÄT: $p < .017$. Für 9 Halbtöne galt: SEHR FRÜH vs. MITTE: $p < .016$, SEHR FRÜH vs. SPÄT: $p < .003$. Wie bei Exp.2 zeigte sich bei dem höchsten und den niedrigen Gipfeln kein signifikanter Unterschied. Wir vermuten, daß bei den niedrigen Gipfeln der Unterschied in der POSITION perzeptiv weniger auffällig ist und daß beim höchsten Gipfel mit 11 HALBTÖNEN die Gipfelhöhe gegenüber der POSITION dominiert. Der MITTLERE und der SPÄTE Gipfel liegen dabei mit 80% bzw. 82.2% Exklamativzuweisung um gut 10% unter den Werten aus Exp.2. 'Störfaktor' sind hier einzelne Versuchspersonen, die - im Sinne unserer Annahmen - 'irregulär' urteilten. Wir haben aber grundsätzlich davon abgesehen, solche Versuchspersonen unberücksichtigt zu lassen.

**Fig.12: Identifikationstest:
Exklamativzuweisung**



EXPERIMENT 9

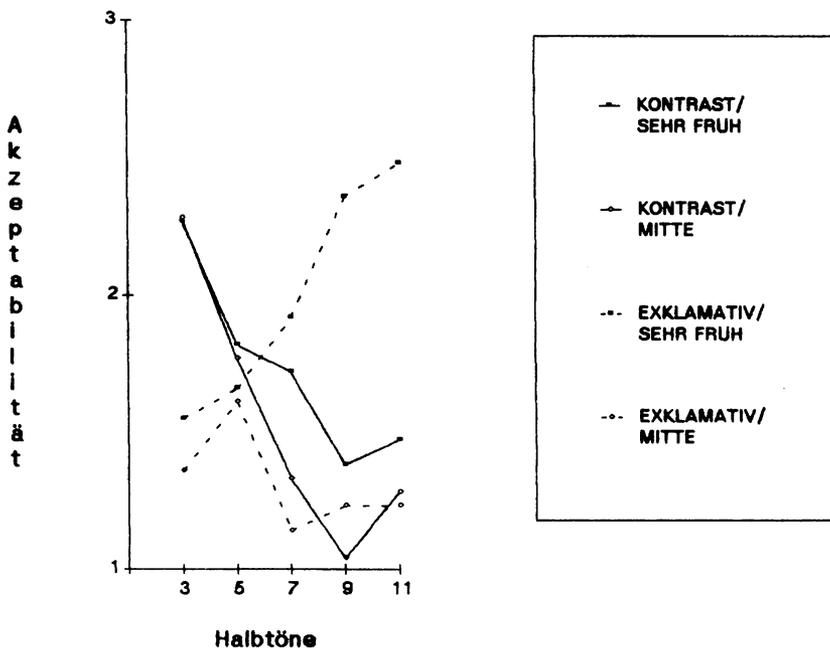
FRAGESTELLUNG: Es sollte untersucht werden, ob sich die schlechtere Exklamativzuweisung an die Stimuli mit einem SEHR FRÜHEN Gipfel auch im KT replizieren läßt.

MATERIAL UND DESIGN: 13 Versuchspersonen wurden in einem KT die Stimuli aus den Kontinua II und IV, also die SEHR FRÜHEN und MITTLEREN Gipfel, gepaart mit den Kontextsätzen KONTRAST, FRÜHER FOKUS und EXKLAMATIV, FRÜHER FOKUS (vgl. Tab.1) je dreimal randomisiert dargeboten.

ERGEBNIS: Kendalls Konkordanzkoeffizient ergibt für die Verläufe in Fig.13 einen signifikanten Unterschied bei den Stimuli mit einem Gipfel von 3 und 9 HALBTÖNEN. Da im Anschluß daran 12 paarweise Vergleiche nach Wilcoxon durchgeführt wurden, mußte das Signifikanzniveau für die einseitige Fragestellung auf .008 abgesenkt werden. Dabei unterschied sich der SEHR FRÜHE Gipfel mit 9 Halbtönen in der Kombination mit dem Exklamativkontext signifikant vom MITTLEREN Gipfel mit Kontrastkontext ($p < .008$) und Exklamativkontext ($p < .004$). Dieses

Ergebnis entspricht der Annahme, daß der SEHR FRÜHE Gipfel exklamativuntypisch ist. Die geringe Anzahl der Versuchspersonen, die Streuung der Antworten und die notwendige Absenkung des Signifikanzniveaus dürften weitere signifikante Unterschiede verhindert haben. Es mag aber doch erlaubt sein, die Kurvenverläufe insgesamt zu interpretieren: mit Ausnahme des SEHR FRÜHEN Gipfels im KONTRAST-Kontext gleichen sie den Verläufen der SPÄTEN und MITTLEREN Gipfel von Exp.4 in Fig.5. Im Gegensatz zum SPÄTEN Gipfel haben wir es aber hier nicht mit einer Verschlechterung in beiden Kontextkonstellationen zu tun. D.h. es liegt nicht an einer mangelhaften Qualität der Stimuli, sondern die Verschlechterung beim Exklamativkontext ist wirklich durch die Unverträglichkeit des SEHR FRÜHEN Gipfels mit dem EXKLAMATIV bedingt.

**Fig.13: Kontexttest:
MODUS/POSITION/HALBTÖN**



Die Ergebnisse von Exp.8 und 9 stützen die Annahme von Kohler, daß ein prävokalischer Gipfel eher eine abgeschlossene Aussage und ein Gipfel im Vokal eher Emphase (und damit cum grano salis Exklamativität in unserem Sinne) indiziert. Allerdings haben wir kein Pendant zu seiner Unterscheidung zwischen einem kategorialen und einem allmählichen Übergang beim Wechsel von der prävokalischen Position in eine Position im Vokal einerseits und bei einer weite-

ren Verschiebung nach rechts andererseits gefunden. Möglicherweise ist das durch die unterschiedliche Materialauswahl und Fragestellung bedingt. Unabhängig davon zeigen aber unsere Ergebnisse, daß die Gipfelhöhe ein Hauptmerkmal, die Gipfelposition dagegen ein Nebenmerkmal sein dürfte, da die Position nur in einem begrenzten Fo-Range-Bereich wirksam wird.

6. DISKUSSION

CAVEAT: Die üblichen Schlüsse aus solchen Experimenten mit einem beschränkten Material auf 'die deutsche Sprache' können natürlich nur vorsichtig gezogen werden und sollten gestützt werden durch konvergierende Ergebnisse mit anderem Material und anderen Methoden. Ebenso nötig ist es, die 'Ontologie' der Begriffe zu überprüfen: Ist z.B. die Gipfelposition auf der Hauptakzentsilbe das entscheidende Merkmal, oder ist es nicht doch der Position dieser Hauptakzentsilbe im Satz untergeordnet? Dieses zweite Merkmal wurde in unseren Tests nicht berücksichtigt. Ein Modell, das lokale Akzentveränderungen über einen globalen Intonationsverlauf der Äußerung legt, wird eher die Relevanz dieses Merkmals annehmen; vgl. auch die Überlegungen weiter unten.

METHODE: Es hat sich als sinnvoll erwiesen, alle drei verschiedenen Arten von Tests einzusetzen, wobei wir auf alle Fälle für die kombinierte Verwendung von IT und KT plädieren. Der IT ist trennschärfer bei der Entdeckung von relevanten Merkmalen, der KT setzt einen natürlicheren Rahmen und kann dabei helfen, unnatürliche Kombinationen bzw. Stimuli zu entdecken.

DIE INTONATORISCHE KENNZEICHNUNG VON AUSSAGE UND EXKLAMATIV: Ein nach rechts verschobener Gipfel begünstigt eine Interpretation als EXKLAMATIV, ein nach links (aus dem vokalischen Bereich heraus) verschobener Gipfel spricht eher für einen Nicht-Exklamativ. Bei der Gipfelhöhe deutet alles darauf hin, daß wir es mit einem breiten Übergangsbereich zu tun haben, wobei die Kategorien-grenze bei einem Gipfel von etwa 7 HALBTÖNEN liegt. Der Nicht-Exklamativ, insbesondere der KONTRAST, kann natürlich auch einen höheren Gipfel als 7 HALBTÖNE aufweisen, er überschreitet aber nicht einen gewissen Dauerbereich. Exklamativtypisch ist also auch - und damit bestätigen wir Bekanntes - die Kombination von hohem Gipfel und Dehnung. Da wir die Dehnung nur kovariierend zum Gipfel untersucht haben, können wir für sie keine Grenzzone - etwa einen Verhältniswert von Dauer der Hauptakzentsilbe zur durchschnittlichen Dauer der übrigen Silben - angeben. Die stabilste Unterscheidung zwischen

AUSSAGE und EXKLAMATIV findet sich, um unseren Begriff aus der Eingangsdiskussion wieder aufzunehmen, wohl nicht bei einem 'Mehr' oder 'Weniger' an intonatorischer Ausprägung, sondern dort, wo gar nichts ist, nämlich bei der rudimentären oder fehlenden Fokussdifferenzierung des EXKLAMATIVS. Beim EXKLAMATIV ist also die Akzentuierung freier als bei der AUSSAGE; er besitzt deshalb keinen Fokusakzent, sondern nur einen Hauptakzent. Damit erhält man eine 'negative' Definition des EXKLAMATIVS, die auch operationalisiert werden kann: Wenn die Akzentuierung einer Äußerung in scheinbarem Widerspruch zur Fokusindizierung durch den Kontext steht, so muß es sich um einen EXKLAMATIV handeln - sieht man von schlichten Fehlakzentuierungen ab. Kennzeichnend für den EXKLAMATIV ist auch, daß andere sprachliche Mittel als die Intonation der betreffenden Teiläußerung den Ausdruck der Exklamativität übernehmen können - z.B. ein vorhergehender Kontext. Diese Ansicht setzt natürlich voraus, daß man ein intonatorisches 'Mehr' zwar als prototypisch, aber nicht als konstituierend für den EXKLAMATIV ansieht.

DIE FRAGE NACH DEM TIEFEREN GRUND: Es mag am Ende erlaubt sein, etwas zu spekulieren. Wir haben ganz zu Anfang die These aufgestellt, daß sich der Exklamativ gegenüber der Aussage durch ein 'Mehr' an intonatorischer Ausprägung auszeichnet. (Die Modi werden im folgenden wieder klein geschrieben, da diese Überlegungen über die speziellen Konstellationen unserer Experimente hinausreichen.) Die übliche Annahme, daß ein höherer Gipfel und eine größere Dauer exklamativtypisch sind, hat sich denn auch in unseren Experimenten bestätigt. Was haben aber ein fehlender Fokusakzent, das gänzliche Fehlen von intonatorischen Ausprägungen und der SPÄTE Gipfel (bzw. ein Gipfel, der nicht SEHR FRÜH in unserem Sinn positioniert ist) mit einem 'Mehr' zu tun? Ohne eine Erklärung dieser Ergebnisse müßte die Antwort auf die im Titel dieses Beitrags formulierte Frage ein entschiedenes, aber unbefriedigendes 'sowohl - als auch' enthalten: Der Exklamativ zeichnete sich dann sowohl durch ein 'Mehr' aus als auch gelegentlich durch etwas anderes oder durch gar nichts.

Eine Lösung dieses Dilemmas liegt u.E. darin, daß man nicht automatisch ein 'Mehr' an intonatorischer Ausprägung mit einem 'Mehr' an Information gleichsetzt. Nehmen wir an, daß der markierte Modus Exklamativ sich durch ein solches 'Mehr' auf der funktionalen Seite (Ausdruck des Erstaunens, größeres Engagement, usw.) auszeichnet. Wir können nun diesem Exklamativ den unmarkierten Normalfall einer 'unauffälligen' Aussage gegenüberstellen, die einen leicht fallenden Deklinationsverlauf und einen normalen, mit dem Kontext verträglichen Fokusakzent ohne allzu große Hervorhebungen aufweist. Wenn wir annehmen, daß

ein Abweichen von dieser Normalstruktur markiert und deshalb per se geeignet zur Kennzeichnung des Exklamativs ist, so kann auch ein 'Weniger' an intonatorischer Ausprägung ein 'Mehr' an Information bedeuten. Das Fehlen des normalen Fokusakzents bzw. eine auf den ersten Blick nicht mit der durch den Kontext indizierten verträgliche Akzentuierung bedeutet ein 'Mehr' an Information, ebenso eine Abweichung von der fallenden Deklination. Ein SPÄTER Gipfel, der absolut gesehen die gleiche Höhe wie ein früherer Gipfel aufweist, ist natürlich relativ zur Deklinationslinie prominenter.

Diese letzte Feststellung gilt dann, wenn man den Gipfel in bezug setzt zur **globalen Fo-Kontur**. Betrachtet man die Verhältnisse **innerhalb** der Silbe, so kann ein weiterer Effekt vorliegen. Der FALL ist beim SEHR FRÜHEN und beim SPÄTEN Gipfel gleich groß, im signifikanten Bereich von Exp.8 und 9 hat aber der RISE auf dem betonten Element beim SPÄTEN Gipfel einen Umfang von 9 HALBTÖNEN, wogegen der RISE beim SEHR FRÜHEN Gipfel nur einen Umfang von 2 HALBTÖNEN aufweist. (Ein ähnlicher Unterschied ist auch bei den Stimuli von Kohler 1987 zu beobachten.) Bei insgesamt gleicher Dauer der Silbe wird also beim SPÄTEN Gipfel ein bedeutend längerer und stärkerer RISE produziert. Es ist nachgewiesen (vgl. Sundberg 1979), daß die Produktion eines RISE länger dauert als die Produktion eines FALL gleichen Umfangs. Das könnte auch dazu führen, daß in der Perzeption bei physikalisch gleicher Dauer beim SPÄTEN Gipfel der Eindruck einer größeren Dauer entsteht (vgl. Ohala 1978:30) – und die Dehnung ist ja ein stabiles Kennzeichen des Exklamativs. Damit ließe sich der arbiträr aussehende Unterschied der Gipfelpositionen doch einem 'Mehr' oder 'Weniger' an intonatorischer Ausprägung unterordnen.

Es muß dahingestellt bleiben, welche dieser Erklärungsmöglichkeiten zutrifft. Sie schließen sich natürlich nicht gegenseitig aus, und sie sind alle gleichermaßen attraktiv, weil sie einen gemeinsamen Nenner bilden für die Antwort auf die Frage, warum denn gerade so unterschiedlich aussehende Merkmale wie Gipfelposition, fehlende Fokuszuweisung und größere intonatorische Ausprägung den Exklamativ kennzeichnen. In allen diesen Fällen ist mit einem 'Mehr', aber auch mit einem 'Weniger' an intonatorischen Ausprägung immer ein 'Mehr' an Information verbunden.

LITERATUR

- Altmann, Hans (1987): Zur Problematik der Konstitution von Satzmodi als Formtypen. In: Meibauer, Jörg (Hg.): Satzmodus zwischen Grammatik und Pragmatik. Tübingen, Niemeyer, 22-56. (= Linguistische Arbeiten 180).
- Bannert, Robert (1985): Fokus, Kontrast und Phrasenintonation im Deutschen. Zeitschrift für Dialektologie und Linguistik, 52, 289-305.
- Batliner, Anton (1988): Modus und Fokus als Dimensionen einer Nonmetrischen Multidimensionalen Skalierung. (In diesem Band).
- Kohler, Klaus J. (1987): Funktionen von F0-Gipfeln im Deutschen. In: Tillmann, Hans G. / Willée, Gerd (Hgg.): Analyse und Synthese gesprochener Sprache. Hildesheim u.a., Olms, 133-140.
- Massaro, Dominic W. / Cohen, Michael M. (1983): Categorical or Continuous Speech Perception: A New Test. Speech Communication, 2, 15-35.
- Näf, Anton (1987): Gibt es Exklamativsätze? In: Meibauer, Jörg (Hg.): Satzmodus zwischen Grammatik und Pragmatik. Tübingen, Niemeyer, 140-160. (= Linguistische Arbeiten 180).
- Oden, Gregg C. (1979): Fuzzy Propositional Approach to Psycholinguistic Problems: An Application of Fuzzy Set Theory in Cognitive Science. In: Gupta, Madan M. / Ragade, Rammohan K. / Yager, Ronald R. (Hgg.): Advances in Fuzzy Set Theory and Applications. Amsterdam u.a., North-Holland Publishing Company, 409-420.
- Ohala, John J. (1978): Production of Tone. In: Fromkin, Victoria A. (Hg.): Tone. A Linguistic Survey. New York u.a., Academic Press, 5-39.
- Repp, Bruno H. (1984): Categorical Perception: Issues, Methods, Findings. In: Lass, Norman J. (Hg.): Speech and Language. Orlando u.a., Academic Press, 243-335.
- Schiefer, Lisa / Batliner, Anton (1988): Intonation, Ordnungseffekt und das Paradigma der Kategorialen Wahrnehmung. (In diesem Band).
- Sundberg, Johan (1979): Maximum Speed of Pitch Changes in Singers and Untrained Subjects. Journal of Phonetics, 7, 71-79.

INTONATION, ORDNUNGSEFFEKT UND DAS PARADIGMA DER KATEGORIALEN WAHRNEHMUNG

Lieselotte Schiefer und Anton Batliner (München)

1. EINLEITUNG

Etablierte, genau beschriebene Modelle haben idealerweise zwei positive Eigenschaften: (1) Wenn man innerhalb des Modells definierte Begriffe verwendet, weiß jeder, wovon man spricht. (2) Eine strikte Definition ist auch Voraussetzung für die Falsifizierbarkeit des Modells bzw. einzelner Teilannahmen. Oft gibt es allerdings mehrere, sich ergänzende, aber auch miteinander konkurrierende Definitionen und Kriterien - so auch beim Paradigma der Kategorialen Wahrnehmung. Dieses Paradigma ist als Instrument gedacht, mit dem man kategoriale Wahrnehmung von nicht-kategorialer, d.h. kontinuierlicher Wahrnehmung abgrenzen kann. Es wurde entwickelt auf segmenteller Materialbasis bei der Wahrnehmung von Plosiven (vgl. Libermann et al. 1957) und mit der Zeit erweitert auf z.B. Vokale und nicht-sprachliche Stimuli. Wir legen im folgenden die wohl bekannteste Spielart des Modells, die der Haskins-Laboratories, zugrunde. Eine ausführliche Diskussion würde den Rahmen dieser Arbeit übersteigen; sie findet sich auch schon im Sammelreferat von Repp (1984), auf das wir uns in diesem Beitrag öfter beziehen werden.

2. DAS PARADIGMA DER KATEGORIALEN WAHRNEHMUNG

Ein Experiment im Paradigma der Kategorialen Wahrnehmung besteht typischerweise aus einem Identifikations- und einem Diskriminationstest (von nun an 'IT' und 'DT'). Im IT werden äquidistante Stimuli eines physikalischen Kontinuums, das zwei (gelegentlich auch mehrere) Kategorien enthält, in randomisierter Folge Hörern dargeboten, die diese einer der vorgegebenen Kategorien zuordnen müssen (*forced choice*). Beim DT werden benachbarte oder auch weiter entfernte Stimuli, die zu Paaren (AX-Test) zusammengestellt sind, auf Gleichheit beurteilt. Es werden dabei drei Stimulusabfolgen (AB, BA, und AA bzw. BB) getestet. Die erhaltene Diskriminationskurve setzt sich aus den richtigen Antworten (*hits*) zu allen Stimulusfolgen zusammen. In den meisten Untersu-

chungen wurde statt des AX-Tests der ABX-Test verwendet, in dem Stimulus-Tripel dargeboten werden, bei denen die beiden ersten Stimuli stets verschieden sind, Stimulus X dagegen Stimulus A oder B sein kann. Da in unseren Experimenten ausschließlich der AX-Test Verwendung fand, gehen wir auf den ABX-Test nicht ausführlicher ein. Wir entschieden uns für den AX-Test, da die Beanspruchung des Gedächtnisses (*memory load*) und damit die Wahrscheinlichkeit, daß die dargebotenen Stimuli zuerst klassifiziert und dann miteinander verglichen werden, beim Vergleich von nur zwei Stimuli (AX) geringer ist als beim Vergleich von drei Stimuli (ABX). Die Hypothese, daß der Mensch kategorial wahrnimmt und nicht nur klassifiziert (wie der Name sagt: 'Kategoriale Wahrnehmung', nicht 'Kategoriale Klassifikation'), wird also mit dem AX-Test einer strengeren Prüfung unterzogen (vgl. Repp 1984:266).

Kategoriale Wahrnehmung in dem strikten Design des Paradigmas ist dann anzunehmen, wenn die folgenden vier Kriterien (Repp 1984: 253) erfüllt sind:

1. Labeling probabilities change abruptly somewhere along the continuum; in other words, the identification functions have a rather steep slope. The point of maximum slope is the **category boundary** (equivalently defined as the point at which responses in two adjacent categories are equiprobable).
2. Discrimination functions show a peak at the category boundary; that is, stimuli are more easily discriminated when they fall on opposite sides of the boundary than when they fall on the same side.
3. Discrimination performance within each category is at or near chance level.
4. Discrimination functions are perfectly predictable from the labeling probabilities (using one of the simple formulae provided by the Haskins model [...]). This implies that (a) the discrimination peak is in exactly the right place and of the right height, and (b) the labeling probabilities are appropriate; that is, they apply independently of the context in which they were observed.

Das 1. Kriterium (*steepness of labeling functions*) ist schwach (vgl. Repp 1984:253), da die Steilheit der Identifikationsfunktion nicht unwesentlich von den Abständen der Stimuli zueinander (*spacing* bzw. Schrittgröße) abhängt. Das 2. Kriterium fordert einen Gipfel in der Diskriminationskurve im Bereich der Kategoriengrenze, an der Stelle der maximalen perzeptiven Ungleichheit benachbarter Stimuli (*phoneme boundary effect*). Dieser Aspekt wird von Repp (1984, 253) als wesentlicher als die Kurvensteilheit beurteilt. Daß dieses Kriterium jedoch trotz gegebener Steilheit der IT-Kurve nicht immer erfüllt sein muß, wird im folgenden noch gezeigt werden. Das 3. Kriterium impliziert, daß Stimuli, die der gleichen Kategorie angehören, nicht oder nur zufällig (*chance level*) diskriminiert werden können. Und das 4. Kriterium schließlich setzt voraus, daß

die Diskriminationsleistung eindeutig aus der Identifikationsleistung voraussagbar ist z.B. nach der Haskins-Formel

$$pcor = .5 + [(p_1 - p_2)^2 / 2].$$

$pcor$ steht für die Wahrscheinlichkeit einer korrekten Diskrimination, p_1 für die Wahrscheinlichkeit, daß A einer der Kategorien und p_2 für die Wahrscheinlichkeit, daß B der gleichen Kategorie zugeordnet wird (vgl. auch Cutting 1982). Die berechnete (*predicted*) kann mit der erhaltenen (*obtained*) Kurve unter Verwendung eines statistischen Verfahrens (z.B. Varianzanalyse, s. Repp et al. 1979) auf die Güte der Anpassung (*goodness of fit*) verglichen werden. Somit scheint einzig dieses Kriterium einer mathematisch-statistischen Falsifizierung zugänglich.

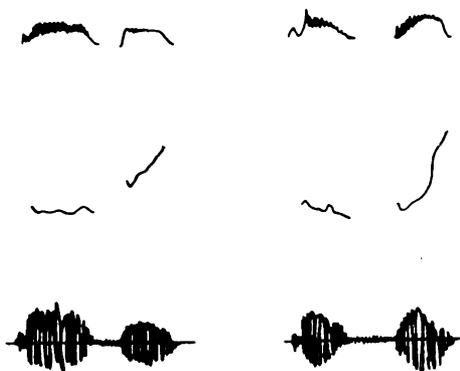
3. ZWEI EXPERIMENTE ZUR AKZENTWAHRNEHMUNG

Testaufbau

Material: Einer der Autoren (A.B.) produzierte im Tonstudio des Instituts für Phonetik in München mehrmals den Fragesatz *Kommen Sie?* in zwei Varianten, wobei der Akzent zum einen auf *Kommen*, zum anderen auf *Sie* lag (Telefunken M15, 19 m/sec.) Als Eckstimuli für die Generierung eines Kontinuums wurden diejenigen Realisationen ausgewählt, die von mehreren kompetenten Beurteilern als die natürlichsten eingestuft wurden und deren intonatorische Parameter die gewünschten Ausprägungen hatten (auditiv überprüft und kontrolliert mit Hilfe von Fo-Meter und Intensity-Meter). Fig.1 zeigt schematisierte Mingogramme dieser beiden Eckstimuli mit dem Zeitsignal, dem Grundfrequenz- (Fo-) Verlauf sowie der Intensität.

Fig.1: Eckstimuli (*Kommen Sie*):

links: Akzent auf *kommen*
rechts: Akzent auf *Sie*
oben: Intensität
Mitte: Fo (Grundfrequenz)
unten: Zeitsignal



Mit den auf einer PDP11/50 digitalisierten und segmentierten Eckstimuli wurde ein 10-stufiges Kontinuum unter Verwendung eines Programmpakets zur Interpolation zwischen natürlich produzierten Sprachsignalen erzeugt. (Dieses Paket ist genauer beschrieben in Simon 1983; zu den in der Zwischenzeit vorgenommenen Verbesserungen vgl. Hadersbeck 1987). Mit dem Verfahren können die Parameter Fo-Verlauf, Intensitätsverlauf, spektraler Energieverlauf, segmentale und damit auch suprasegmentale Zeitstruktur jeweils isoliert oder, wie in diesem Fall, gemeinsam behandelt werden. Für die Experimente ergaben sich also neben den resynthetisierten Eckstimuli acht synthetisierte, bezüglich der perzeptiv relevanten Parameter in ihrer physikalischen Ausprägung äquidistante Zwischenstufen.

Versuchspersonen waren acht Studenten der Linguistik bzw. der Phonetik, die freiwillig an den Experimenten teilnahmen.

Design: Es wurde ein IT und darauffolgend, mit einer Woche Abstand, ein DT durchgeführt. Beim IT wurden die zehn Testitems je zehnmal mit einem zeitlichen Abstand von 3.5 sec zwischen den einzelnen Items dargeboten. Beim DT wurden die Items im Einerschritt (1/2, 2/3, ... 2/1, 3/2 ...) gepaart und die drei unterschiedlichen Anordnungen AA, BB, AB bzw. BA mit einem Abstand von 3.5 sec zwischen den Paaren und 500 msec innerhalb der Paare je fünfmal randomisiert dargeboten. Die Versuchspersonen saßen dabei im Sprachlabor des Instituts für Phonetik vor einem Abstimmkästchen und hörten die Stimuli über die Raumlautsprecher. Genau nach jedem Stimulus(-paar) wurde die Abstimmung freigegeben; dies wurde durch eine Lampe an jedem Kästchen angezeigt. Die Versuchspersonen hatten dann drei Sekunden Zeit, ihre Antwort zu überlegen und die entsprechende Taste zu drücken. Die Antworten wurden auf einer PDP11/03 gesammelt und zur weiteren Verarbeitung aufbereitet. Beim IT lautete die Instruktion: "Bitte drücken Sie die linke Taste, wenn der Akzent Ihrer Ansicht nach auf *kommen*, oder die rechte, wenn er auf *Sie* liegt." Beim DT lautete die Instruktion: "Bitte entscheiden Sie, ob die beiden Stimuli innerhalb eines Paares in allen Belangen gleich klingen oder nicht, und drücken Sie dann die linke Taste für 'gleich' oder die rechte für 'verschieden'."

Statistische Auswertung: Es wurden zweifaktorielle Varianzanalysen gerechnet. Im Rahmen unserer Argumentation ist allerdings eine adäquate Prüfstatistik problematisch, da wir verschiedene Konstellationen mit dem gleichen Datenmaterial berechnen müssen. Es ist aber auch nicht unser Ziel, mit prüfstatistisch signifikanten Ergebnissen zu argumentieren. Wir wollen nur illustrieren, welche Ergebnisse man erhält, wenn man den eigenen Entscheidungskriterien unterschiedliche Berechnungen zugrunde legt. D.h. wir tun so, als ob wir jeweils nur eine Berechnung durchführen würden. Wir geben deshalb normalerweise nur die Wahrscheinlichkeitswerte p an und sprechen nicht davon, daß das Ergebnis signifikant ist, sondern daß es signifikant sein würde - bei Zugrundelegung jeweils nur einer bestimmten Datenkonstellation.

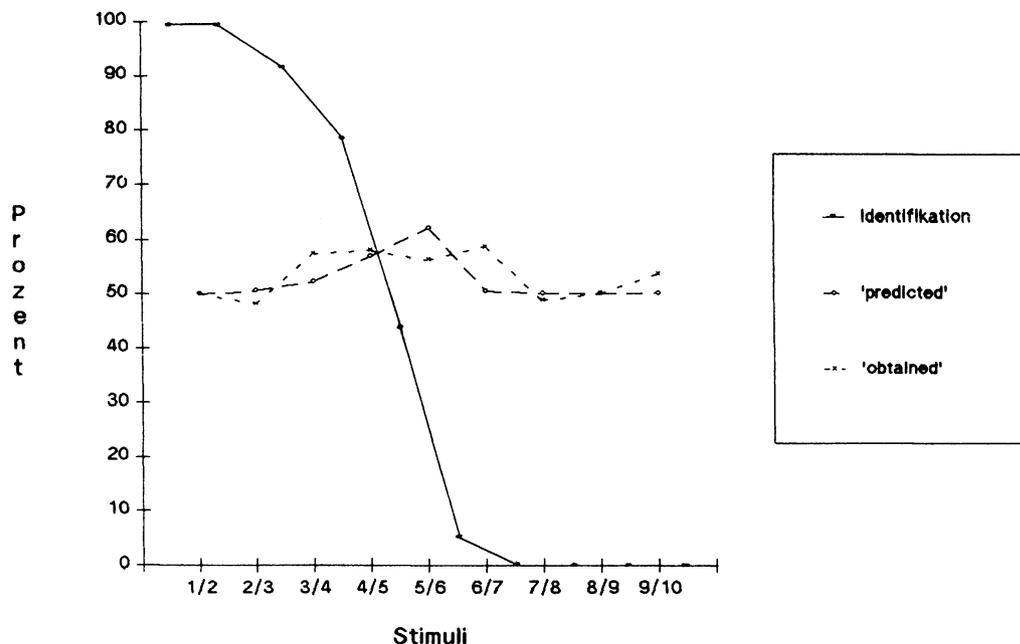
Ergebnisse

Die Versuchspersonen unterteilten im IT das Kontinuum in zwei Kategorien (Fig.2). Dabei wurde in den ersten vier Stimuli der Akzent auf *kommen*, in den letzten fünf Stimuli auf *Sie* wahrgenommen. Stimulus fünf konnte keiner Kategorie eindeutig zugeordnet werden.

Die aus den Ergebnissen des IT errechnete *predicted*-Kurve (ebenfalls Fig.2) weist einen niedrigen Diskriminationsgipfel bei Stimulus-Paar 5/6 auf. Die *obtained*-Kurve zeichnet sich durch ein niedriges Plateau von Paar 3/4 bis Paar 6/7 aus, das seinen niedrigsten Wert bei Paar 5/6 (sic!) hat, also genau an derjenigen Stelle, an der die *predicted*-Kurve ihren Gipfel aufweist. Trotz unterschiedlicher Form wären beide Kurven jedoch nicht signifikant verschieden ($p < .5$).

Haben also die bei der Generierung der Test-Stimuli vorgenommenen Manipulationen der akustischen Parameter zu einer kategorialen Wahrnehmung des Akzents geführt oder muß man einen kontinuierlichen Übergang von einer Kategorie in die andere annehmen? Zur Beantwortung dieser Frage seien die in Teil 2 diskutierten Kriterien auf die Test-Ergebnisse angewandt. Kriterium 1 (*labeling*

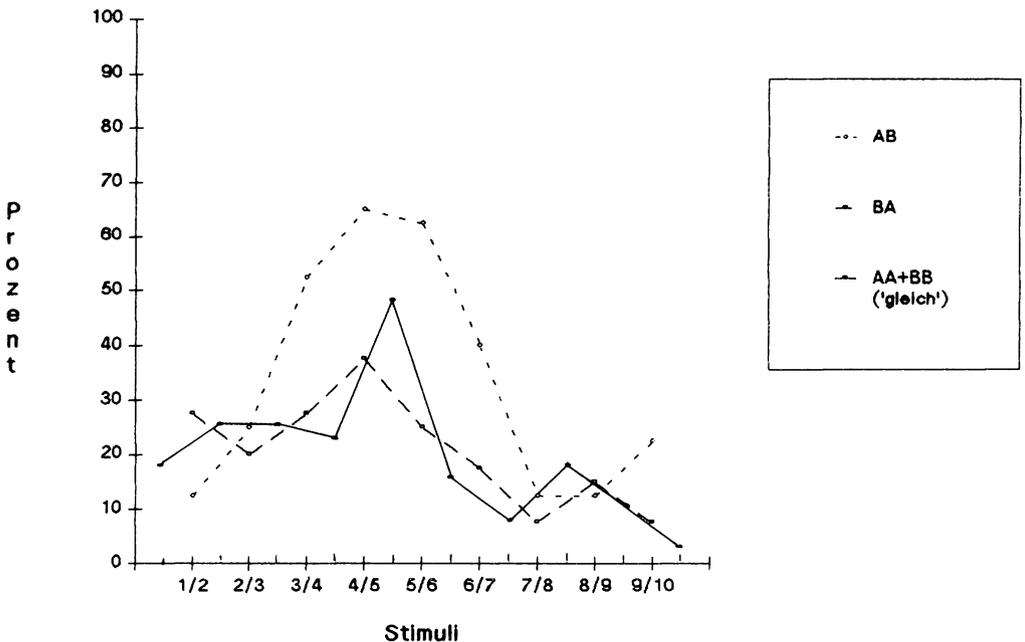
Fig.2: Identifikation und Diskrimination ('predicted' vs. 'obtained')



probabilities change abruptly) setzt einen abrupten Übergang von der einen zur anderen Kategorie im IT voraus. Dieses Kriterium scheint in unserem Fall nicht voll erfüllt, da Stimulus 5 keiner Kategorie eindeutig zugeordnet wurde. (Das mag mit der Schrittgröße (*spacing*) im Kontinuum zu tun haben; vgl. dazu unten Teil 4.) Kriterium 2 (*discrimination functions show a peak at the category boundary*) ist in keiner Weise erfüllt, da kein eindeutiger Diskriminationsgipfel existiert und darüberhinaus sogar eine Erniedrigung der Kurve bei Stimulus-Paar 5/6 vorliegt. Dagegen scheint das 3. Kriterium (*discrimination performance within each category is at or near chance level*) voll erfüllt. Die Ergebnisse entsprechen jedoch ebenfalls nicht Kriterium 4, da die Diskriminationsfunktion nicht aus dem IT voraussagbar ist (daran ändert auch die Tatsache nichts, daß die zu vergleichenden Kurven statistisch nicht signifikant verschieden wären).

Da drei der vier Kriterien nicht (voll) erfüllt sind, sprechen die Ergebnisse gegen eine kategoriale Wahrnehmung. Bedeutet dies jedoch, daß kontinuierliche Wahrnehmung anzunehmen ist? Bevor ein solches Urteil gefällt werden kann, müssen u.E. die Ergebnisse für die unterschiedlichen Darbietungsformen (AB, BA, AA bzw. BB) einzeln untersucht werden. Diese sind in Fig.3 getrennt dargestellt.

Fig.3: Diskrimination (AB, BA, 'gleich')



Drei wesentliche Ergebnisse sind zu diskutieren:

(1) Die 'verschieden'-Paare in Anordnung AB werden häufiger als verschieden perzipiert als die entsprechenden Paare in der Anordnung BA ($p < .001$). Beide Kurven zeigen einen Diskriminationsgipfel bei Paar 4/5; allerdings ist bei der AB-Kurve der Wert für das benachbarte Paar 5/6 fast genauso hoch. Das Phänomen der von der Anordnung der Stimuli innerhalb der Test-Paare abhängigen unterschiedlichen Beurteilung ansonsten gleicher Paare bezeichnen wir als **Ordnungseffekt**.

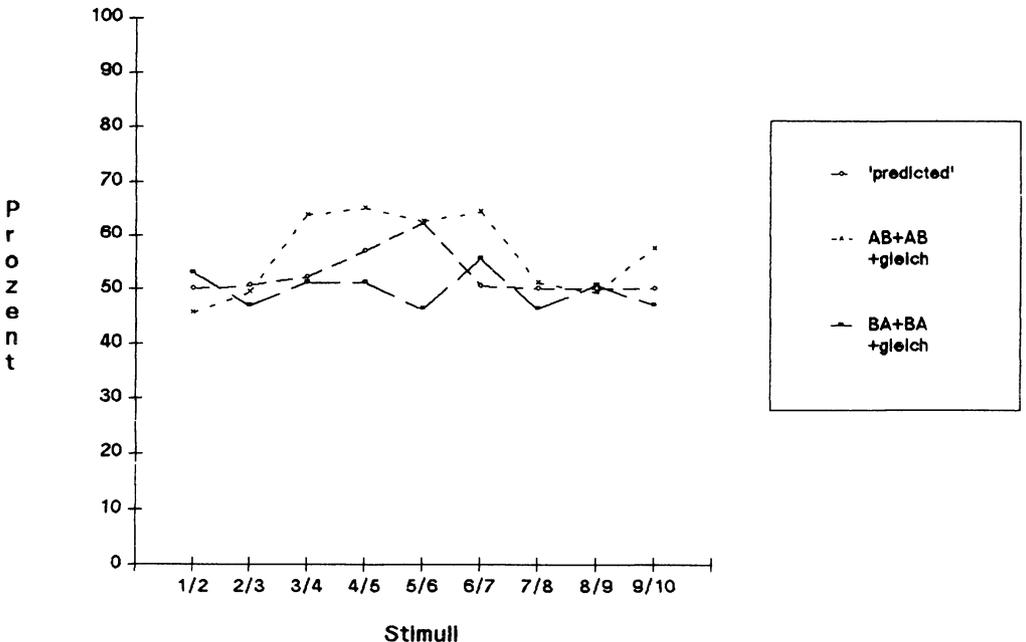
(2) Paar 1/2 wird in der Anordnung BA besser diskriminiert als in der Anordnung AB, während die Ergebnisse für Paar 9/10 in der AB-Darbietung besser sind. Beide Paare enthalten einen unmanipulierten (Stimulus 1 resp. 10) und einen manipulierten Stimulus (Stimulus 2 resp. 9). (Das erklärt auch die besseren Ergebnisse für die Eckpaare in allen berechneten Diskriminationskurven. Wir vermuten, daß bedingt durch ein Artefakt der Stimulusgenerierung die Intensität dieser resynthetisierten natürlichen Ausgangsstimuli im Verhältnis zu den synthetisierten Zwischenstufen zu hoch ist. Dies führt dann zu einer unverhältnismäßig guten Diskriminierbarkeit der Paare, bei denen der (prominente) Eckstimulus an zweiter Stelle steht. Der Fehler wurde inzwischen behoben, vgl. Hadersbeck 1987.)

(3) Bei den Ergebnissen für die 'gleich'-Paare (AA bzw. BB) weist Stimulus-Paar 5/5 einen eindeutigen Gipfel auf (nahe 50%). Dieses Ergebnis ist besonders interessant, da der Gipfel im Bereich der Kategorien-Grenze liegt, nämlich bei demjenigen Stimulus, der im IT keiner Kategorie eindeutig zugeordnet werden konnte. Vergleichbare Ergebnisse wurden von uns auch bei DTs zur Plosivwahrnehmung (Schiefer, bisher unveröffentlicht) gefunden und deuten darauf hin, daß ein systematischer Faktor dafür verantwortlich sein muß. Eine mögliche Erklärung wäre, daß der Hörer auch die im DT zum Vergleich dargebotenen Stimuli kategorisiert (klassifiziert), wobei zunächst der 1. Stimulus einer der vorgegebenen Kategorien zugeordnet wird. Der 2. Stimulus wird nicht auf seine akustische Struktur, sondern lediglich daraufhin überprüft, ob er der gleichen Kategorie angehört, die man dem 1. Stimulus zugewiesen hat. Ist dies der Fall, wird mit 'gleich' geantwortet, andernfalls mit 'verschieden'. Handelt es sich um einen kategoriell nicht eindeutig identifizierbaren Stimulus (in unserem Fall Stimulus 5), so wird zwar beim 1. Stimulus der gleiche Prozeß der Klassifizierung wie bei einem eindeutigen Stimulus durchlaufen, d.h. er wird einer der Kategorien zugeordnet, etwa 'Akzent auf *kommen*'. Der 2. Stimulus wird nun daraufhin

überprüft, ob der Akzent **eindeutig** auf *kommen* liegt. Da der Stimulus 'unentscheidbar' ist, ist dies nicht der Fall und es wird mit 'verschieden' geantwortet.

Wieweit beeinflussen nun diese Ergebnisse die Gesamt-Diskriminationsleistung (*obtained*-Kurve)? Es ist eindeutig, daß die Abflachung der *predicted*-Kurve im Bereich der Kategoriengrenze (die Paare 4/5 und 5/6) durch die *false alarms*, d.h. durch falsche 'verschieden'-Antworten auf 'gleich'-Paare bewirkt wird, da an beiden Paaren Stimulus 5 beteiligt ist. Basiert die Berechnung der Diskriminationskurve dagegen nicht auf AB und BA, sondern **nur** auf AB (Formel: $(AB+AB+AA+BB)/4$; vgl. Fig.4), so ist, ebenso wie bei der *obtained*-Kurve, kein Diskriminationsgipfel vorhanden, sondern die Ergebnisse für die Paare 3/4, 4/5, 5/6 und 6/7 sind fast identisch. Die Kurve liegt grundsätzlich **über** der *predicted*-Kurve ($p < .004$). Wird dagegen die Diskriminationskurve aufgrund der Ergebnisse für die BA-Darbietungsrichtung berechnet, so erhält man eine Kurve, die grundsätzlich **unter** der *predicted*-Kurve liegt ($p < .09$). Ein eindeutiger Gipfel fehlt auch hier; die beste Diskriminationsleistung erzielt Paar 6/7. Bei keiner der durchgeführten Berechnungen tritt ein Diskriminationsgipfel an der erwarteten Stelle auf.

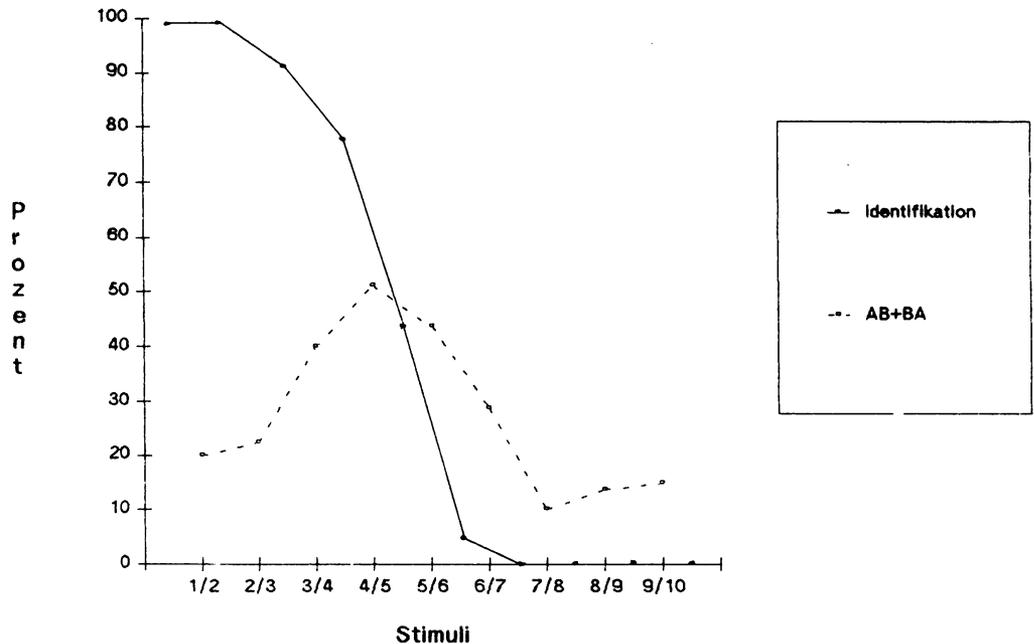
Fig.4: Diskrimination ('predicted', AB+AB+'gleich', BA+BA+'gleich')



Zusammenfassend kann für jede Art der Berechnung festgehalten werden: (1) Die *false alarms* für die 'gleich'-Paare führen zu einer Abflachung der Diskriminationskurve. (2) Die Einbeziehung der BA-Paare führt zu einer Erniedrigung, die der AB-Paare zu einer Erhöhung der Diskriminationskurve.

Es ist schließlich noch zu fragen, ob der Diskriminationsgipfel, der sich bei Addition der Antworten für die 'verschieden'-Paare (AB und BA) ergibt (gemittelte Kurve), wenn nicht in der Höhe so doch in der Position mit den Ergebnissen aus dem IT-Test übereinstimmt. Dies ist in Fig.5 dargestellt. Die Kurve zeigt einen Gipfel bei Paar 4/5 und entspricht damit den Ergebnissen aus dem IT, die einen Schnittpunkt der IT-Kurve mit der 50%-Linie zwischen Stimulus 4 und 5 aufweisen. (Man beachte, daß bei der Berechnung der *predicted*-Kurve der Schnittpunkt zwischen der IT-Kurve und der 50%-Linie nicht berücksichtigt wird, sondern ausschließlich die rechnerische Differenz zwischen den Antworthäufigkeiten für die einzelnen Stimuli maßgebend ist.)

Fig.5: Identifikation und Diskrimination (AB+BA)



Kehten wir nun zur Frage nach der Kategorialität der Ergebnisse zurück, die bei Heranziehung des Paradigmas in seiner strikten Form sicher zu verneinen ist. Die Alternative dazu - kontinuierliche Wahrnehmung - ist allerdings bei genauer Betrachtung der Kurven ebenfalls unplausibel. Ein Weg aus dem Dilemma ist eine weniger strikte Anwendung des Paradigmas, wobei etwa lediglich die Form der IT-Kurve und die gemittelte DT-Kurve einer heuristischen Beurteilung unterzogen werden. Bei einem solchen Vorgehen wird man sicher eine kategoriale Wahrnehmung annehmen. Der Preis dafür ist das Fehlen einer exakten Definition und der objektiven Meßbarkeit von "Kategorialität". Wir werden auf diesen Punkt in der Diskussion zurückkommen.

4. DER ORDNUNGSEFFEKT

Für die Ergebnisse aus IT und DT sind zumindest zwei Faktoren mit 'verantwortlich', die primär nichts mit Kategorialität, sondern mit der Konstruktion der Teststimuli und ihrer Darbietung im Experiment zu tun haben. Der eine Faktor ist bekannt: Die Schrittgröße, d.h. die Dichte der Abfolge der Stimuli im physikalischen Kontinuum (*spacing*). Ein zu geringer Abstand der Stimuli zueinander führt im IT zu einem weniger abrupten Übergang von einer zur anderen Kategorie. Im DT resultiert daraus einerseits eine Abflachung der Diskriminationskurve bei den 'verschieden' Paaren und eine Erhöhung der *false alarms* an der Kategoriengrenze bei den 'gleich'-Paaren. Ein zu groß gewählter Schritt dagegen würde im IT zwar zu dem gewünschten abrupten Kategorienwechsel, im DT jedoch zu einer zu starken Erhöhung der Diskriminationskurve (*ceiling effect*) für die 'verschieden'-Paare führen, während mit einer Reduzierung der *false alarms* bei den 'gleich'-Paaren zu rechnen ist. Man muß also gegebenenfalls in Pilotexperimenten die adäquate Schrittgröße ermitteln, die dann im eigentlichen Experiment getestet werden kann.

Ein zweiter, die Ergebnisse maßgeblich beeinflussender Faktor, ist der Ordnungseffekt, der sich in einer (häufig signifikant) besseren Diskriminierbarkeit von Stimulus-Paaren in einer der beiden Darbietungsrichtungen AB oder BA äußert. Der Ordnungseffekt wurde bereits im 19. Jahrhundert durch Fechner entdeckt, der ihn den 'constanten Fehler' bei der Ermittlung des 'Masses der Unterschiedsempfindlichkeit' nannte (Fechner 1964: 90ff). In der Folgezeit, besonders durch die Arbeiten von Woodrow (1935, 1951), Woodworth (1950), Stott (1933, 1935) sowie Woodrow/Stott (1936) wurde der Ordnungseffekt (in der englischsprachigen Literatur als *time-order error* bezeichnet) in der experimen-

tellen Psychologie einer intensiven Erforschung unterzogen. Ein *time-order error* konnte z.B. nachgewiesen werden bei der Wahrnehmung von Dauer, Länge, Lautstärke, Helligkeit und Geschmack. Als Erklärung für das Phänomen wurde zunächst eine Verknüpfung mit Gedächtnis- und Perzeptionsprozessen angenommen, während spätere Untersuchungen einen *subject bias* dafür verantwortlich machen, sei es in Form eines *simple response bias* (Luce 1959) oder eines *criterion bias* (Wickelgren 1968). Eine endgültige Erklärung für das Phänomen steht jedoch nach wie vor noch aus.

In der Psychophysik wird der Ordnungseffekt normalerweise durch eine Mittelung der Ergebnisse neutralisiert. Die Phonetik verfährt ähnlich; dort wurde mit wenigen Ausnahmen (s. etwa Repp et al. 1979; Smith 1976; Rosen 1977; Chuang-Wang 1978a, 1978b) der Ordnungseffekt nicht zur Kenntnis genommen und unseres Wissens nicht hinsichtlich seiner Auswirkung auf Experimente im Paradigma der Kategorialen Wahrnehmung thematisiert. Die Ergebnisse aus dem oben beschriebenen Experiment lassen die Annahme zu, daß der Ordnungseffekt eine systematische 'Störvariable' im Paradigma ist, die nicht als Test-Artefakt erklärbar ist.

5. EXPERIMENTE ZUM 'DESIGN-EFFEKT'

Mit den im folgenden geschilderten Experimenten wurde die Hypothese getestet, daß eine gemischte (durchrandomisierte) und eine getrennte Darbietung, bei der entweder nur die Anordnung AB oder die Anordnung BA auftritt, gleichermaßen zu einem signifikanten Ordnungseffekt führen. Dieser Annahme nach kann der Ordnungseffekt zwar durchaus vom experimentellen Design beeinflußt werden, er ist aber nicht **ausschließlich** auf dieses Design zurückzuführen.

Testaufbau

Material: Ausgangsmaterial für die Stimuli war ein natürlich produziertes, monotonen *ja* (Aufnahmebedingungen wie oben in Teil 3 beschrieben). Dieser kurze und durchgehend stimmhafte Ausgangsstimulus wurde gewählt, um eine exakte Manipulation der interessierenden Parameter zu gewährleisten. Der Stimulus wurde unter Verwendung eines Segmentationsprogramms (für Einzelheiten s. Batliner/Schiefer 1987) in einzelne Perioden segmentiert. Die Fo-Manipulation der sieben generierten Stimuli begann stets im quasi-stationären Teil des Vokals, dessen Fo bei 85 Hz lag. Der Fo-Offset war bei allen Stimuli konstant bei 114 Hz. Bei Stimulus 1 erstreckte sich der Fo-Anstieg über die letzten 4 Perioden (das entspricht einer Dauer von 39.3 msec); für die übrigen Stimuli wurde der Fo-Anstieg um jeweils zwei Perioden verlängert, so daß Stimulus 7 einen Anstieg

über 16 Perioden (entsprechend 159.5 msec) aufwies. Das auf diese Weise generierte Kontinuum basierte somit auf unterschiedlicher Dauer und Steilheit (*slope*) des Fo-Anstiegs, während der Fo-Offset, der Fo-Range und die mittlere Fo in allen Stimuli unverändert war. Dauer und Steilheit des Fo-Anstiegs waren äquidistant; die Dauerunterschiede zwischen den einzelnen Stimuli betragen ca. 20 msec, die durch die pitchsynchrone Manipulation (periodenweise Segmentierung, kein fixes Zeitfenster) bedingten Schwankungen lagen im Bereich von +/- 1.5 msec. Außer mit dem beschriebenen Kontinuum wurde mit zwei weiteren Kontinua gearbeitet, die auf der Manipulation anderer Fo-Parameter basierten. Damit sollte zum einen die perzeptive Relevanz der unterschiedlichen Parameter wie Fo-Offset, Dauer und Steilheit des Fo-Anstiegs und mittlere Fo überprüft werden; zum anderen sollten diese Experimente ermitteln, welche Anordnung innerhalb der Stimulus-Paare bei den verschiedenen Manipulationen besser diskriminierbar ist. Da die Ergebnisse aus den drei Kontinua sich in den hier interessierenden Punkten nicht unterscheiden, beschränken wir uns auf die Darstellung der Ergebnisse eines Kontinuums. (Es wird im folgenden aus Platzgründen auch nicht auf alle damit durchgeführten Tests eingegangen, da eine Diskussion der ITs und damit der Frage, ob überhaupt und in welchem Ausmaß unterschiedliche Kategorien im Kontinuum enthalten sind, zu weit führen würde und in unserem Zusammenhang auch nicht interessiert. Über diese Ergebnisse werden wir an anderer Stelle berichten.)

Design: Für die AX-DTs wurden die 7 Stimuli im 2er Schritt gepaart (1/3, 2/4, 3/5 ... 3/1, 4/2, 5/3 ...) und in randomisierter Folge je 5mal wiederholt. Das Interstimulus-Intervall betrug 500 msec, die Pause zwischen den Paaren 3.5 sec. Im 1. DT wurden alle Darbietungsfolgen (AB, BA, AA bzw. BB) randomisiert dargeboten. In den beiden weiteren DT wurden die Darbietungsrichtungen AB und BA jeweils mit den 'gleich'-Paaren getrennt randomisiert und getestet. Im 2. DT wurden zunächst die BA-Paare, dann die AB-Paare dargeboten, während im 3. DT die Reihenfolge umgekehrt war. Dies ergibt die folgenden Testkonfigurationen:

1. Diskriminationstest: AB, BA, AA/BB gemischt randomisiert
2. Diskriminationstest: 1.Teil: BA, AA/BB; 2.Teil: AB, AA/BB
3. Diskriminationstest: 1.Teil: AB, AA/BB; 2.Teil: BA, AA/BB

Die Darbietung der Tests erfolgte mit einer Revox 77 über Raumlautsprecher. Testpersonen waren die Teilnehmer (n = 10, 11, 15) eines Seminars. (Zur Instruktion und dem allgemeinen Prozedere vgl. oben Teil 3.) Da die Tests mit mindestens einer Woche Abstand durchgeführt wurden, betrachteten wir sie als 'unabhängig', also nicht als wiederholte Messungen an den gleichen Versuchspersonen. Für jeden Test rechneten wir zwei-faktorielle Varianzanalysen mit den Faktoren 'Paar' (1/3, 2/4, 3/5, 4/6, 5/7) und 'Darbietungsrichtung' (AB, BA).

Ergebnisse

Die Ergebnisse sind in Fig. 6-8 dargestellt. In allen DTs ist die Diskrimination der AB-Paare besser als die der BA-Paare (1.DT: $F(1,9)=79.83$; $p<.001$; 2.DT: $F(1,10)=24.35$; $p<.001$; 3.DT: $F(1,14)=75.77$; $p<.001$). Wir vergleichen die Tests zwar nicht prüfstatistisch untereinander, da dafür die Voraussetzungen fehlen, es läßt sich jedoch an Hand der Abbildungen feststellen, daß das unterschiedliche Design Einfluß auf die Ergebnisse hatte: Beginnt der Test mit der weniger prominenten Stimulus-Abfolge BA (Fig.7), so werden diese Paare besser diskriminiert als die gleiche Abfolge im durchrandomisierten Test (Fig.6); stehen dagegen die prominenten Paare AB an erster Stelle (Fig.8), so werden diese etwas schlechter diskriminiert als im durchrandomisierten Test (Fig.6). Wir haben es hier also mit einer gegenseitigen Beeinflussung der Darbietungsfolgen zu tun: die Anwesenheit der prominenten Paare AB im gleichen Test führt zu einer schlechteren Diskrimination der BA-Paare, während diese ihrerseits zu einer besseren Diskrimination der AB-Paare führen (*anchoring*; vgl. den folgenden Abschnitt). Die Ergebnisse aus den zweiten Hälften der nicht-durchrandomisierten

Fig.6: randomisiert

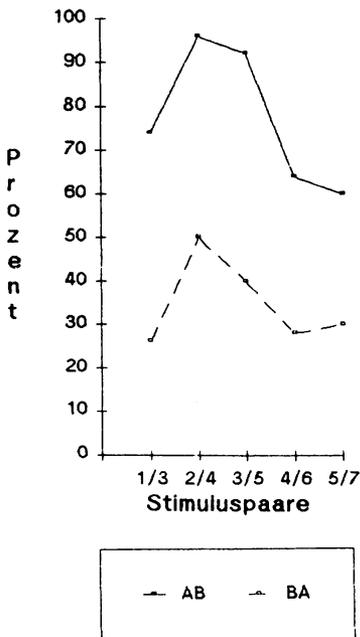


Fig.7: Teil1:BA, Teil2:AB

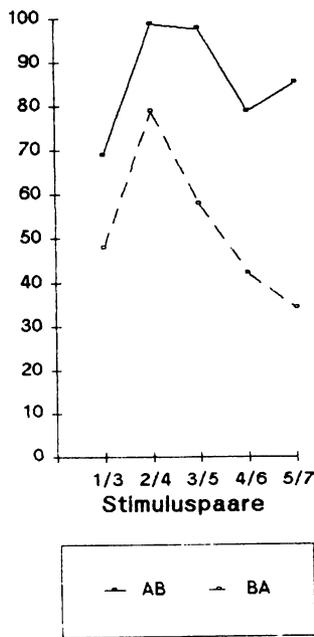
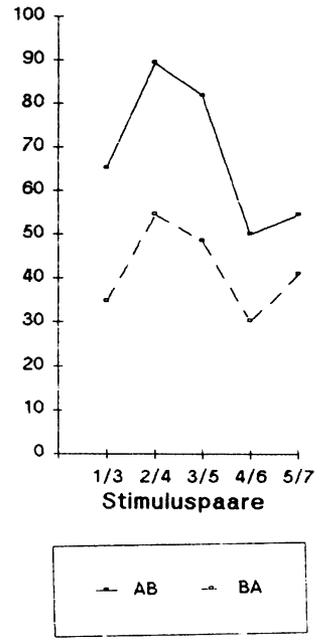


Fig.8: Teil1:AB, Teil2:BA



Tests unterscheiden sich nur geringfügig von den entsprechenden Verläufen des durchrandomisierten Tests; wir vermuten, daß dafür verschiedene, sich gegenseitig in ihrer Wirkung neutralisierende Faktoren verantwortlich sind.

Für eine Überprüfung dieser Ergebnisse wurde bei einem Fo-Kontinuum, das im Prinzip gleich aufgebaut war, aber eine andere Schrittgröße (30 msec statt 20 msec) hatte, das Testdesign geändert. Dabei prüften wir die Hypothese, daß trotz gemischter Darbietung von Paaren unterschiedlicher Schrittgröße (1/2, 2/3, 3/4 vs. 1/3, 2/4, 3/5 vs. 1/4, 2/5 usw.) auch bei den Paaren mit geringerem Unterschied (1er Schritt) ein signifikanter Ordnungseffekt auftritt. Diese Hypothese nimmt also an, daß sich ein Ankereffekt (*anchoring*), der bewirkt, daß die Versuchspersonen ihre Bewertung anhand der deutlichsten Unterschiede innerhalb eines Paares kalibrieren, nicht entscheidend auf den Ordnungseffekt auswirkt. Es wurden zwei DTs erstellt. Im 1. DT hatten die 'verschieden'-Paare die Schrittgröße 1, während im 2. DT alle möglichen Schrittgrößen (1er bis 6er Schritt) vorlagen. Die AB-Paare wurden gemäß unserer Annahme beim 1er und auch beim 2er Schritt signifikant besser als die BA-Paare diskriminiert ($p < .001$). Ab dem 3er Schritt stellte sich ein *ceiling effekt* ein: beide Anordnungen wurden sehr gut diskriminiert. Die Ergebnisse entsprechen ansonsten den oben diskutierten Experimenten.

Die Ergebnisse dieser beiden Testserien sprechen dafür, den Ordnungseffekt nicht als simplen Designeffekt aufzufassen. (Damit ist natürlich nicht ausgeschlossen, daß er grundsätzlich ein experimenteller Effekt ist. Es ist noch nicht klar, welche Rolle er in natürlichsprachlicher Kommunikation spielt.) Weitere Experimente zu den folgenden zwei Fragen wurden mit dem gleichen Material durchgeführt und sollen an anderer Stelle dargestellt werden: (1) Handelt es sich beim Ordnungseffekt um ein systematisches Phänomen, d.h. lassen sich bei der Wahrnehmung von Tonhöhen Bedingungen für die prominenten Paare angeben? (2) Ist ein unterschiedlicher Ordnungseffekt festzustellen, wenn Tonhöhenunterschiede bei sprachlichem bzw. nicht-sprachlichem Material wahrgenommen werden? Ohne der Diskussion dieser Ergebnisse vorzugreifen, kann jedoch festgestellt werden, daß immer eine prominente (besser diskriminierbare) Anordnung AB einer nicht-prominenten Anordnung BA gegenübersteht. Als prominent erwiesen sich bisher Anordnungen, bei denen (a) der höhere Offset (s. Teil 3), (b) die längere Dauer des Fo-Anstiegs (wie gerade gezeigt) oder (c) die größere Fo-Veränderung (vgl. Batliner/Schiefer 1987) an zweiter Stelle stand.

6. DISKUSSION

Es konnte nachgewiesen werden, daß der Ordnungseffekt einen entscheidenden Einfluß auf die Berechnung der Kategorialität im strikten Paradigma ausübt. Wir wollen nun drei mögliche, unterschiedlich weitreichende Folgerungen aus diesem Befund diskutieren:

(1) **Das strikte Paradigma der Kategorialen Wahrnehmung ist abzulehnen.** - Man könnte dagegen einwenden, daß das Paradigma nicht für die Wahrnehmung von Tonverläufen, sondern für die Wahrnehmung von Segmenten, insbesondere Plosiven konzipiert wurde. Der Ordnungseffekt läßt sich aber auch im segmentellen Bereich nachweisen, vgl. Chuang/Wang 1978a. Wir fanden ihn auch bei VOT- und Amplitudenwahrnehmung (Schiefer, bisher unveröffentlicht). U.E. blieb der Ordnungseffekt im Rahmen des Paradigmas nicht deswegen unbeobachtet, weil es ihn nicht gibt, sondern weil er bei den üblichen Testdesigns, z.B. dem ABX-Test, vgl. oben Teil 2, nicht so deutlich sichtbar wurde.

(2) **Der Ordnungseffekt wird in das Paradigma integriert.** - Es gibt natürlich Wege, den Ordnungseffekt in das Paradigma zu integrieren, etwa indem man ihn wie die Schrittgröße einbaut - damit wird er zum heuristischen Problem. Wir können uns allerdings keine befriedigenden Kriterien vorstellen, anhand derer er in das strikte Paradigma integrierbar wäre. Man müßte also einen anderen Kriterienkatalog aufstellen. Es bleibt zu fragen, ob es sich dann noch um das gleiche Paradigma handelt. Damit kommen wir zur dritten, am weitesten reichenden Folgerung:

(3) **Das globale Paradigma der Kategorialen Wahrnehmung ist grundsätzlich abzulehnen.** - Egal, welche Kriterien ein solches Paradigma ansetzt: Mit dieser Folgerung wird in Frage gestellt, daß man mit einem einzigen Paradigma sinnvoll alle lautsprachlichen Phänomene bearbeiten kann. Prima vista mag der Ordnungseffekt ein zu bescheidenes Phänomen sein, um eine solche Schlußfolgerung motivieren zu können. Wir wollen deshalb einige weitere relevante Punkte zumindest anführen, auch wenn wir sie nicht im Detail diskutieren können:

Die Kategoriengrenze: Der Ordnungseffekt tangiert insbesondere die Kriterien, mit denen die Beschaffenheit der Kategoriengrenze bestimmt wird. Über dieses innerhalb des Paradigmas doch grundlegende Konzept schreibt Repp (1984:320):

One true shortcoming of the categorical perception paradigm is that it has overemphasized the importance of the boundaries between phonetic categories. [...] The boundaries [...] are a mere epiphenomenon, apparent only in a particular experimental situation. [...] beyond the realm of artificial speech continua, the boundary concept has little to offer.

Nun ist die Kategoriengrenze unmittelbar oder mittelbar an allen der vier eingangs vorgestellten Kriterien beteiligt. Welchen Wert hat aber ein Paradigma, das sich wesentlich auf dieses 'Epiphänomen' stützt, wobei das Epiphänomen als solches schon durch einen anderen systematischen Faktor, nämlich den Ordnungseffekt - egal, ob es sich bei diesem um ein 'echtes' Phänomen oder auch nur um ein Epiphänomen handelt - in Frage gestellt ist?

Der Begriff 'Kategorialität': Ausgangspunkt für das Paradigma war ein experimentelles Ergebnis: synthetisierte Plosivkontinua wurden diskontinuierlich, also kategorial wahrgenommen (Liberman et al. 1957; zur Geschichte des Paradigmas vgl. wieder Repp 1984). Da dies anfangs nur bei solchen **sprachlichen** Phänomenen zu beobachten war, folgerte man daraus: 'speech is special', die Wahrnehmung von Sprache ist kategorial, nicht kontinuierlich strukturiert. Im Laufe der Jahre mußte die Anfangsaussage zwar modifiziert werden - Vokale werden wenig bis gar nicht kategorial wahrgenommen, es gibt eine kategoriale Wahrnehmung auch im nicht-sprachlichen Bereich - , der Grundgedanke blieb aber attraktiv: Es gibt einfach sehr viele Kategorien (mit distinktivem Zeichencharakter) in der Sprache, und das Paradigma schien ein geeignetes Instrument zu sein, mit dem sich solche Kategorien im phonetischen Bereich beschreiben lassen. Der Begriff 'Kategorialität' ist also grundlegend, aber zugleich schillernd. Es gibt ihn, strikt definiert, im Paradigma, es gibt ihn in allen möglichen Varianten in der Linguistik, und es gibt ihn im normalen Sprachgebrauch als die Eigenschaft von jemand oder von etwas, unterschiedlichen Gruppen anzugehören. In der Praxis kann das dazu führen, daß man den Begriff, der innerhalb des Modells definiert ist, in einer weniger strikten oder gar umgangssprachlichen Bedeutung verwendet. Diese Verwendung ist problematisch, wenn stillschweigend der weniger strikte Begriff angewandt, aber doch die mit dem strikten Modell verbundene theoretische Aussage damit gemeint ist. Eine solche Begriffscamouflage zeigt sich z.B. bei Lindsay/Ainsworth (1985): einerseits eine explizite Bezugnahme auf das Modell, andererseits eine unzureichende Anwendung der Kriterien. Im Einzelfall läßt sich das natürlich einfach als unzulässiges Vorgehen kritisieren. Insgesamt ist es aber ein Phänomen der normativen und zugleich verwirrenden Kraft des faktischen Wissenschaftsdiskurses: Der Begriff wird in so unterschiedlichen Bedeutungen verwendet, daß er

bedeutungsleer wird. Dies beruht u.E. letztlich auf der inhaltlich inadäquaten globalen Verwendung dieses Begriffs, wie wir im folgenden Punkt skizzieren wollen.

Die Adäquatheit des Begriffs 'Kategorialität': Modelle haben die Tendenz, sich zu verselbstständigen. D.h. man bleibt zu sehr dem experimentellen Mikrokosmos verhaftet und vernachlässigt, daß das Modell ja nur ein Hilfsmittel sein soll, um die natürlichsprachliche Kommunikation besser verstehen zu können. (Vgl. auch das Schlußwort in Repp 1984:322: "It is to be expected [...] that the traditional methodology will eventually give way to new approaches that more directly address the important theoretical and practical problems raised by **communication in the real world.**" [unsere Hervorhebung]) Nun ist die Modellierung der *communication in the real world* in Experimenten unterschiedlich lebensnah: Was den segmentalen Bereich betrifft, so ist die übliche Aufgabe, etwa innerhalb eines Kontinuums von [p] nach [b] zu differenzieren, per se unnatürlich. Der Mensch differenziert normalerweise keine Laute, sondern zumindest Wörter - etwa *Pein* vs. *Bein* - wenn nicht doch immer Satz- oder Textbedeutungen. Entscheidend bleibt aber, daß es in diesem Fall wirklich nur zwei Kategorien gibt, zwischen denen differenziert werden kann, eben Fortis und Lenis. Damit ist nicht gemeint, daß der Hörer keine Zwischenstufen zwischen [p] und [b] hören und ihnen keinen Sinn beimessen kann, etwa als Indikator der regionalen Herkunft des Sprechers. Solche Differenzierungen laufen aber auf einer anderen Ebene: sie haben **Anzeichen-**, aber keine **Zeichenfunktion**. Bei der Intonation ist die Aufgabe, etwa zwischen Frage und Exklamativ zu unterscheiden, eher 'aus dem Leben gegriffen'. Man mag einwenden, daß diese Aufgabe viel komplexer ist als die Unterscheidung zwischen /p/ und /b/. Im Sinne eines *ecological approach*, also eines Ansatzes, der die Bedingungen der natürlichsprachlichen Kommunikation stärker ins Kalkül zieht, ist sie angemessener: Die Frage, ob es sich um ein [p] oder [b] handelt, beschäftigt den Phonetiker, aber nicht den 'normalen' Menschen. Die Frage, ob es sich um eine Frage oder einen Exklamativ handelt, hat dagegen jeden 'normalen' Menschen schon öfter beschäftigt. Allerdings sind die Kategorien Frage und Exklamativ weniger eindeutig unterschieden als etwa /p/ und /b/. Es gibt zwar auch eindeutige, prototypische Fragen versus eindeutige Nicht-Fragen (selen das nun Exklamative oder Aussagen; zu dieser Differenzierung vgl. Batliner 1988). Es gibt aber ebenfalls mehrere Abstufungen der Fragehaltigkeit, mit der eine Frage gestellt wird (vgl. dazu Batliner 1987). Wieder etwas anders liegt es bei der Akzentuierung, wie in unserem Testsatz *Kommen Sie*. Hier gibt es im Regelfall nur eine alternative Fokussierung - damit wird natürlich eine kategoriale Wahrnehmung begünstigt. Bei anderen Konstellationen ist aber auch eine

Doppelfokussierung möglich (wobei noch dahingestellt bleiben muß, ob sie sich auch in zwei gleichwertigen Akzentuierungen niederschlägt.)

Es gibt also ein ganzes Spektrum unterschiedlich strukturierter Kategorien ('echt' binäre mit scharfen Grenzen, binäre mit breiten Übergangsbereichen, durch den Kontext mehr oder weniger beeinflussbare, etc.), und es ist mehr als fraglich, ob sie sich gemeinsam mit ein und demselben Instrumentarium und ein und demselben Begriff 'Kategorialität' beschreiben lassen. Aus all diesen Gründen halten wir eine Abkehr vom strikten Paradigma besonders bei der Betrachtung von Intonationsphänomenen und eine vorsichtig dosierte Verwendung des Begriffs "kategorial" für wünschenswert.

LITERATUR:

Batliner, A. (1987): Kategorialität und Kontexteffekte bei Frage- und Exklamativmodus im Deutschen. Perzeptionsexperimente zur Rolle des Fo-Verlaufs. Ms.

Batliner, A. (1988): Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des Fo-Gipfels. (In diesem Band).

Batliner, A. / Schiefer, L. (1987): Stimulus category, reaction time, and order effect - an experiment on pitch discrimination. Proceedings XIth ICPhS, Vol. 5: 46-49.

Chuang, C.-K. / Wang, W.S.-Y. (1978a): The time-order error in judgement of prosodic features: The pitch, the intensity, and the duration. J. Acoust. Soc. Am. 62: S48(A). (Zitiert nach Chuang/Wang 1978b).

Chuang, C.-K. / Wang, W.S.-Y. (1978b): Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. J. Acoust. Soc. Am. 64: 1004-1014.

Cutting, J. E. (1982): Plucks and bows are categorically perceived, sometimes. Perception and Psychophysics 31: 462-476.

Fechner, G. Th. (1964): Elemente der Psychophysik. Amsterdam, E.J. Bonset. (Nachdruck der Ausgabe Leipzig 1860).

Hadersbeck, M. (1987): A new program for manipulation of natural speech - interpolation between two natural utterances. Proceedings XIth ICPhS, Vol. 5: 35-38.

Lieberman, A.M. / Harris, K.S. / Hoffman, H.S. / Griffith, B.C. (1957): The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54: 358-368.

- Lindsay, D. / Ainsworth, W. A. (1985): Two models of nuclear intonation. *Journal of Phonetics* 13: 163-173.
- Luce, R. D. (1959): *Individual choice behavior*. New York u.a., John Wiley & Sons.
- Repp, B. H. (1984): Categorical perception: Issues, methods, findings. In: Lass, N. J. (Hg.): *Speech and Language*. Orlando u.a., Academic Press: 243-335.
- Repp, B.H. / Healy, A.F. / Crowder, R.G. (1979): Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance* 5: 129-145.
- Rosen, S.M. (1977): The effect of fundamental frequency patterns on perceived duration. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1: 17-30.
- Simon, Th. (1983): Manipulation of natural speech signals according to the speech parameters of different speakers. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 17: 233-245.
- Smith, M.R. (1976): An investigation of changes in categorical perception of vowels. Unpubl. Manuskript. Department of Linguistics, University of Connecticut. (Zitiert nach Repp et al. 1979).
- Stott, L.H. (1933): The discrimination of short tonal duration. Ph.D. diss. University of Illinois Library. (Zitiert nach Woodrow 1951).
- Stott, L.H. (1935): Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology* 18: 741-766.
- Wickelgren, W.A. (1968): Unidimensional strength theory and component analysis of noise in absolute and comparative judgements. *Journal of Mathematical Psychology* 5: 102-122.
- Woodrow, H. (1935): The effect of practice upon time-order errors in the comparison of temporal intervals. *The Psychol. Rev.* 42: 127-152.
- Woodrow, H. (1951): Time perception. in: Stevens, S.S. (Hg.): *Handbook of Experimental Psychology*. New York u.a., John Wiley & Sons: 1224-1236.
- Woodrow, H. - Stott, L.H. (1936): The effect of practice on positive time-order errors. *Journal of Experimental Psychology* 19: 694-705.
- Woodworth, R.S. (1950²): *Experimental psychology*. New York, Holt, Rinehart, and Winston. (1. Auflage 1938).

Anmerkung: Die folgende umfangreiche Aufsatzsammlung erschien erst nach Fertigstellung dieses Aufsatzes und konnte deshalb nicht berücksichtigt werden:

Harnad, S. (Hg.) (1987): *Categorical perception. The groundwork of cognition*. Cambridge u.a., Cambridge University Press.

Experimente zur Zeitstruktur von synthetischen Phrasierungseinheiten.

Doris Rittinghaus, Bochum

0. Abstract

Es werden die Ergebnisse von Tests vorgestellt, die die Rolle der Zeitstruktur bei der Manifestation von Phrasierungseinheiten zum Gegenstand haben; vorerst beschränkt sich das Testmaterial auf synthetische Silbenfolgen.

Die prosodischen Parameter, die für die Wahrnehmung von Gliederungen als relevant postuliert werden, insbesondere die Zeitstruktur, werden gesteuerter Veränderung unterworfen. Ausgangsmaterial für die Modifikationen ist eine Kette aus synthetisch generierten Silben, die aus zwei alternierenden Lauten, dem Konsonanten /m/ und dem Vokal /a/, bestehen.

Die zentrale Hypothese besagt, daß die Wahrnehmung einer Lautkette als in zwei aufeinanderfolgende Phrasierungseinheiten gegliedert nicht nur auf dem Einfügen von Pausen in die ungegliederte Struktur beruht, sondern auch auf der Einheit inhärenten prosodischen Merkmalen. Diese Merkmale werden hauptsächlich in der Zeitstruktur vermutet. Eines dieser Merkmale ist das sogenannte "phrase final lengthening", d.h. die Dehnung des direkt vor der postulierten Phrasierungsgrenze befindlichen Vokals.

Um die Frage zu beantworten, ob die Phrasierung nur ein Phänomen des Grenzbereichs, repräsentiert durch eine Pause oder eine Verlagerung der Pause in das vorhergehende Segment durch die Vokaldehnung, ist, oder ob die Phrasierungseinheit eine eigene Zeitstruktur besitzt, wurden Sequenzen generiert, in denen die Vokaldehnung durch eine Verkürzung des dem Vokal vorausgehenden Konsonanten kompensiert wird. Vorbild für dieses Verfahren ist das Konzept des 'momentary tempo' von VENCOV 1981. Die Ergebnisse dieser Tests können in dem Sinne interpretiert werden, daß die Länge des Vokals und nicht die der ganzen Silbe für die Wahrnehmung einer Phrasierungsgrenze zwischen Phrasierungseinheiten verantwortlich ist.

In anderen Testsequenzen wurde die Dehnung über alle Silben der Einheit verteilt. Die Ergebnisse zeigen, daß der Kontrasteffekt an der Grenze den ständigen Anstieg der Grenzsignale über die gesamte Phrasierungseinheit hinweg übertrifft.

1. Einleitung

Am Lehrstuhl für Akustik und allgemeine Elektrotechnik der Ruhruniversität Bochum wurden mit Hilfe eines an einen Microcomputer angeschlossenen Phonemeditors synthetische Silbenfolgen hergestellt, im Zeitbereich manipuliert und zu Testreihen zusammengestellt.

Dieses synthetische Material hat gegenüber natürlichsprachlichem den Vorteil, daß die Hypothesen in Isolation, unabhängig von semantischen Faktoren untersucht werden können, wobei allerdings zu überprüfen bleibt, ob sich die Ergebnisse ohne weiteres auf natürlichsprachliches Material übertragen lassen.

2. Testaufbau

Die synthetischen Silbenfolgen bestehen aus jeweils 8 Silben /ma/, wobei das /m/ jeweils 135 ms, das /a/ jeweils 140 ms lang ist; der Abstand zwischen den aufeinanderfolgenden einzelnen Silben beträgt 27 ms.

Die Testreihen setzen sich aus drei bis fünf verschiedenen solcher Silbenfolgen in Zufallsreihenfolge zusammen, wobei jede Folge bis zu 10mal vorkommt, je nach Anzahl der verschiedenen Stimuli. Da die Tests auf die Versuchspersonen leicht enervierend wirken, erweist es sich als günstig, in einer Testreihe eine Anzahl von maximal 30 Stimuli nicht zu überschreiten. Zwischen den einzelnen Stimuli wurde eine Pause von 2,1 sec. eingefügt.

Die Aufgabe der Versuchspersonen bestand darin zu entscheiden, ob die gehörte Silbenfolge als in sich gegliedert oder als ungegliedert interpretiert wurde. Diese Entscheidung wurde von den Versuchspersonen auf einem Testbogen angekreuzt, auf dem auf der einen Seite die Silbenfolgen graphisch als ungegliedert, auf der anderen als gegliedert dargestellt wurden:

mamamama,mamamama 0 ... 0 mamamamamamama

An den Tests nahmen jeweils zwischen 5 und 10, teilweise auch 15 Versuchspersonen, größtenteils Studenten unterschiedlicher Fachrichtungen sowie Mitarbeiter, teil. Die Tests wurden überwiegend in einer geräuscharmen Kammer durchgeführt.

3. Tests

3.1. Pausenlänge

Die ersten beiden Tests sollten untersuchen, von welcher Länge an Pausen zwischen den synthetischen Silben wahrgenommen und als Einschnitte interpretiert werden.

Der erste Test bestand aus dem ungegliederten Original und 3 Folgen, in die nach der 4. Silbe Pausen von 40 ms, 80 ms und 200 ms Länge eingefügt wurden. Die Folgen mit den Pausen von 80 ms und 200 ms Länge wurden deutlich als gegliedert erkannt, während bei einer Pausenlänge von 40 ms die Ergebnisse nicht so eindeutig waren (Ergebnisse der Bewertungen als "gegliedert" in %):

1. Serie		
1.1. Original	0 %	(0)
1.2. 40 ms Pause	60 %	(24)
1.3. 80 ms Pause	95 %	(38)
1.4. 200 ms Pause	100 %	(40)

Im zweiten Test wurde die ungegliederte Folge mit Folgen mit Pausenlängen von 20 ms, 30 ms, 40 ms und 50 ms gemischt den Versuchspersonen präsentiert. Die Ergebnisse zeigen, daß auch Pausen von 20 ms und 30 ms Länge schon wahrgenommen werden. Zudem wird deutlich, daß die Eindeutigkeit der Bewertung der Folgen als gegliedert mit zunehmender Pausenlänge wächst, wobei ein deutlicher Sprung zwischen 30 und 40 ms Pausenlänge liegt.

Test C		
2.1. Original	5 %	(4)
2.2. 20 ms Pause	75 %	(60)
2.3. 30 ms Pause	79 %	(63)
2.4. 40 ms Pause	95 %	(76)
2.5. 50 ms Pause	96 %	(77)

Da in den folgenden Tests einige der Folgen noch wiederholt einer Bewertung unterzogen wurden, ist es interessant, die Ergebnisse aller Pausenlängen, die getestet wurden, im Vergleich zu betrachten. Auch hier wächst die Eindeutigkeit der Bewertung mit zunehmender Pausenlänge, wobei allerdings der Sprung zwischen 40 und 50 ms liegt.

3.1. 20 ms Pause	40 %
3.2. 30 ms Pause	79 %
3.3. 40 ms Pause	69 %
3.4. 50 ms Pause	96 %
3.5. 80 ms Pause	97 %
3.6. 200 ms Pause	100 %

3.2. Dehnung

Die nächsten Tests dienten der Untersuchung, ob die Dehnung eines Vokals ebenfalls als Gliederungssignal bewertet wird und, wenn das zutrifft, von welcher Länge an die Dehnung wahrgenommen wird.

Der erste Versuch dieser Reihe enthielt Dehnungen des vierten Vokals von 40 ms, 80 ms und 140 ms sowie das ungegliederte Original. Alle Dehnungen wurden als gegliedert bewertet, die Dehnungen von 80 ms und 140 ms mit 100% der Antworten, aber auch die 40ms-Dehnung mit 75%.

Serie 2		
4.1. Original		0 %
4.2. 4./a/ + 40 ms		75 %
4.3. 4./a/ + 80 ms		100 %
4.4. 4./a/ + 140 ms		100 %

Ein zweiter Test mit Dehnungen von 20 ms, 30 ms, 40 ms und 60 ms erbrachte lediglich bei der Dehnung von 60 ms eine eindeutige Bewertung für die Gliederung, die anderen Folgen wurden als ungegliedert interpretiert, wobei auch die Dehnung von 40 ms ein wesentlich schwächeres Ergebnis zeigte als im vorangegangenen Test.

Test E		
5.1. Original		6 %
5.2. 4./a/ + 20 ms		24 %
5.3. 4./a/ + 30 ms		17 %
5.4. 4./a/ + 40 ms		17 %
5.5. 4./a/ + 60 ms		85 %

Die 40ms-Dehnung kam in mehreren der darauffolgenden Test vor, dabei liegt das Gesamtergebnis höher, vgl. die Zusammenstellung der Ergebnisse aller Folgen mit Dehnung an der postulierten Gliederungsgrenze:

6.1. 4./a/ + 20 ms	20 %
6.2. 4./a/ + 30 ms	17 %
6.3. 4./a/ + 40 ms	49 %
6.4. 4./a/ + 60 ms	86 %
6.5. 4./a/ + 80 ms	94 %
6.6. 4./a/ + 140 ms	100 %

Man sieht also deutlich, daß eine kurze Dehnung insgesamt schlechter als eine gleichlange Pause wahrgenommen wird.

Die Ergebnisse sind aber mit Vorsicht zu betrachten, denn der Aufbau einer Testreihe scheint die Ergebnisse der einzelnen Folgen stark zu beeinflussen. Wir bemühten uns in jedem Fall darum, daß in jedem Test sowohl ein eindeutig gegliederter als auch ein eindeutig ungegliederter Stimulus vorkam. Bemerkenswert ist dabei, daß der ungegliederte Originalstimulus bei 100%iger Sicherheit als solcher erkannt wurde.

3.3. Vergleich Pause-Dehnung

In der nächsten Testreihe wurden Folgen mit Pausen und Dehnungen gemischt, und zwar Folgen mit 40 ms und 80 ms Pause und Folgen mit 40 ms- bzw. 80 ms-Dehnung des vierten Vokals sowie das Original.

Während bei 80 ms sowohl die Pause als auch die Dehnung eindeutig als Gliederungssignal interpretiert wurden, wurde bei 40 ms die Pause deutlich stärker als die Dehnung als gegliedert bewertet:

Test 1		
7.1. Original		4 %
7.2. 40 ms Pause		64 %
7.3. 4./a/ + 40 ms		52 %
7.4. 80 ms Pause		100 %
7.5. 4./a/ + 80 ms		100 %

Das bestätigt sich, wenn man die Pausenlängen und die gleichlangen Dehnungen aus verschiedenen Tests vergleicht:

20 ms Pause	40 %	4./a/ + 20 ms	20 %
30 ms Pause	79 %	4./a/ + 30 ms	17 %
40 ms Pause	69 %	4./a/ + 40 ms	49 %
80 ms Pause	97 %	4./a/ + 80 ms	94 %

3.4. Verteilte Dehnung

Der nächste Schritt bestand in der Untersuchung der Frage, ob für die Wahrnehmung der Dehnung die Länge des Segments unmittelbar vor der postulierten Gliederungsgrenze als Kriterium dient, oder ob sich die Dehnung auch auf mehrere Silben verteilen läßt.

Der erste Test dieses Untersuchungsabschnitts setzte sich zusammen aus dem ungegliederte Original, einer Folge mit 40 ms Pause nach der vierten Silbe, einer Folge mit Dehnung des vierten Vokals um 40 ms sowie einer Folge, in der diese Vokaldehnung auf die dritte und vierte Silbe mit jeweils 20 ms verteilt wurde. Während die Folge mit einfacher Dehnung und die mit der Pause als gegliedert interpretiert wurden, wurde die verteilte Dehnung überwiegend als ungegliedert bewertet:

Test 2	
8.1. Original	4 %
8.2. 40 ms Pause	84 %
8.3. 4./a/ + 40 ms	68 %
8.4. 3.+4./a/+20 ms	28 %

Dieses Ergebnis scheint dafür zu sprechen, daß lediglich das letzte Segment als Bewertungsmaßstab für die Gliederung dient, die Dehnung der vierten Silbe um 20 ms allein erzielte ähnliche Ergebnisse, vgl. 5.2..

Da aber die Dehnung von 40 ms schon vorher einen kritischen Grenzfall darstellte, wurde ein ähnlicher Test durchgeführt, in dem jedoch die Dehnung 60 ms betrug. Diese Dehnungsfolge wurde zusammen mit dem Original mit verteilten Dehnungen gemischt, und zwar wurde die Dehnung wie folgt auf die vier ersten Silben verteilt:

	1.Silbe	2.Silbe	3.Silbe	4.Silbe
1. Folge	+ 0ms	+ 10ms	+ 20ms	+ 30ms
2. Folge	+ 0ms	+ 20ms	+ 20ms	+ 20ms
3. Folge	+ 0ms	+ 0ms	+ 30ms	+ 30ms
4. Folge	+ 0ms	+ 0ms	+ 0ms	+ 60ms

Alle drei Folgen mit verteilter Dehnung wurden als nicht gegliedert bewertet, die Ergebnisse lagen noch niedriger als bei den Folgen mit Dehnung der vierten Silbe um 20 ms bzw. 30 ms. Die Folge mit 3x20 ms wurde am wenigsten als gegliedert interpretiert, ihr Ergebnis stimmt mit dem des ungegliederten Originals überein:

Test B	
9.1. Original	9 %
9.2. 4./a/ + 60 ms	90 %
9.3. 2./a/ + 10 ms,	
3./a/ + 20 ms,	
4./a/ + 30 ms	20 %
9.4. 2.,3.+4./a/+20ms	9 %
9.5. 3.+4./a/+30 ms	26 %

Bei der Bewertung der Länge scheint also lediglich die Silbe unmittelbar vor der Gliederungsgrenze als Maßstab genommen zu werden.

3.5. Kompensierte Dehnung

Diese Testreihe sollte dazu dienen festzustellen, ob die Vokaldehnung durch eine entsprechende Kürzung des vorhergehenden Konsonanten ausgeglichen werden kann.

Im ersten Test wurden zwei Paare gemeinsam mit dem ungegliederten Original den Versuchspersonen präsentiert: Eine Folge, in der das vierte /a/ um 40 ms gedehnt wurde, eine zweite, in der diese Vokaldehnung durch Kürzung des vierten /m/ um 40 ms kompensiert wurde und zwei Folgen, in denen entsprechende Veränderungen um 80 ms vorgenommen wurden:

Test A

10.1. Original		7,5 %
10.2. 4./a/ + 40 ms		57,5 %
10.3. 4./m/ - 40 ms,		
4./a/ + 40 ms		42,5 %
10.4. 4./a/ + 80 ms		85 %
10.5. 4./m/ - 80 ms		
4./a/ + 80 ms		85 %

Ein weiterer Test dieser Reihe wurde mit der Dehnung von 60 ms durchgeführt, wobei die Folge mit der kompensierten Dehnung noch häufiger als gegliedert bewertet wurde als die mit einfacher Dehnung:

Test D

11.1. Original		5 %
11.2. 4./a/ + 60 ms		80 %
11.3. 4./m/ - 60 ms,		
4./a/ + 60 ms		92,5 %

Diese Ergebnisse lassen sich als Evidenz für die Hypothese bewerten, daß im Falle der Gliederung nicht die Länge der Silbe im Grenzbereich den Ausschlag gibt, sondern die Länge des Vokals. Die Kontrastwirkung im Grenzbereich setzt sich gegenüber dem kontinuierlichen Anwachsen der Grenzsignale durch.

4. Diskussion

Alle diese Tests wurden im Rechenzentrum der RUB einer statistischen Auswertung unterzogen. Dabei wurde deutlich, daß die einzelnen Ergebnisse stark vom Testaufbau abhängig sind. So erreichte z.B. die in mehreren Tests auftretende Folge mit einer Dehnung des 4. /a/ um 40 ms Werte zwischen 17,14% (Test E, 4.4.) und 75 % (Serie 2, 3.2.). Ein Grund hierfür ist sicherlich der niedrige Anteil der gegliedert-Bewertungen in Test E (29,71%) gegenüber Serie 2 (68,75 %), vgl. auch Tab. 2.

Bei einer Dehnung von 40 ms Länge scheint eine kritische Grenze zu liegen (vgl. den Durchschnittswert, den diese Folge erzielte (49,2 % bei insgesamt 315 Bewertungen). Bei der entsprechend langen Pause von 40 ms dagegen sind die Schwankungen in den Bewertungen geringer (zwischen 45 % (Test I (40,75 % gegliedert-Bewertungen), 10.3.) und 95 % (Test C (70,0% gegliedert-Bewertungen), 2.4.), wenngleich ebenfalls sehr hoch.

Insgesamt werden Pausen deutlicher als Dehnungen als Gliederungssignale empfunden, vgl. hierzu auch die oben dargestellte Gegenüberstellung unter 3.3.; selbst eine Pause von 50 ms liegt mit 96,25 % in der Bewertung als Gliederungssignal um einiges höher als eine Dehnung um 60 ms (85,78 %). Vergleicht man die Folgen

4.2.	4./a/ + 20 ms	23,81 %
6.4.	3.+4./a/ + 20 ms	28,00 %
7.4.	2.+3.+4./a/ + 20 ms	8,75 % ,

so zeigt sich an den geringen Unterschieden deutlich, daß bei der Bewertung der Gliederung lediglich die Silbe unmittelbar vor der postulierten Grenze als Bewertungsmaßstab bei der Wahrnehmung herangezogen wird.

Innerhalb dieser letzten Silbe unmittelbar vor der Grenze wiederum gibt offensichtlich nur der Vokal den Ausschlag, wie ein Vergleich von Folgen mit bloßer Dehnung des Vokals und Folgen, in denen die Dehnung durch eine entsprechende Kürzung des Konsonanten kompensiert wird, zeigt, vgl. Test A und Test D, wobei sich keine Unterschiede von größerer Signifikanz ergaben.

Da bei ersten Tests mit Frequenzveränderungen die Frequenzabsenkungen sehr hoch angesetzt worden waren und bereits ohne zusätzliche Pausen oder Dehnungen überwiegend als eindeutig gegliedert wahrgenommen wurden, konnten leider die Kombinationen mit Pausen und Dehnungen kaum andere Ergebnisse erzielen. Es ist aber vorgesehen, in diesem Bereich weitere Tests zu erstellen und durchzuführen.

Test III

12.1.	Original		11,25 %	(9)
12.2.	f0 122 -	100 Hz	92,5 %	(74)
12.3.	f0 122 -	89 Hz	98,75 %	(79)
12.4.	f0 122 -	78 Hz	97,5 %	(78)

Test II

13.1.	4./a/ +	20 ms	10,16 %	(11)
13.2.	f0 122 -	100 Hz	98,95 %	(94)
13.3.	f0 122 -	100 Hz,	98,95 %	(94)
	20 ms Pause			
13.4.	f0 122 -	100 Hz,	100 %	(95)
	4./a/ +	20 ms		
13.5.	20 ms Pause		10,53 %	(10)

5. Tabellen:

Tabelle 1: Getestete Stimuli

1.	Original	635	6,92 %	(44)
2.	20 ms Pause	175	40 %	(70)
3.	30 ms Pause	80	78,75 %	(63)
4.	40 ms Pause	250	69,2 %	(173)
5.	50 ms Pause	80	96,25 %	(77)
6.	80 ms Pause	65	97 %	(63)
7.	200 ms Pause	40	100 %	(40)
8.	4./a/ +	280	20 %	(56)
9.	4./a/ +	105	17,14 %	(18)
10.	4./a/ +	315	49,2 %	(155)
11.	4./a/ +	225	85,78 %	(193)
12.	4./a/ +	105	94,29 %	(99)
13.	4./a/ +	40	100 %	(40)
14.	3./a/ +	25	28 %	(7)
	4./a/ +			
15.	2./a/ +	80	20 %	(16)
	3./a/ +			
	4./a/ +			
16.	2./a/ +	80	8,75 %	(7)
	3./a/ +			
	4./a/ +			

17.	3./a/ + 30 ms,	80	26,25 %	(21)
	4./a/ + 30 ms			
18.	4./m/ - 40 ms,	40	42,5 %	(17)
	4./a/ + 40 ms			
19.	4./m/ - 60 ms,	40	92,5 %	(37)
	4./a/ + 60 ms			
20.	4./m/ - 80 ms,	40	85 %	(34)
	4./a/ + 80 ms			
21.	4./a/ + 20 ms,	80	52,5 %	(42)
	20 ms Pause			
22.	f0 122 - 100 Hz (-22 Hz)	175	96 %	(168)
23.	f0 122 - 89 Hz (-33 Hz)	80	98,75 %	(79)
24.	f0 122 - 78 Hz (-44 Hz)	80	97,5 %	(78)
25.	f0 122 - 100 Hz (-22 Hz),	95	98,95 %	(94)
	20 ms Pause			
26.	f0 122 - 100 Hz (-22 Hz),	95	100 %	(95)
	4./a/ + 20 ms			
		-----	-----	-----
		3285	54,36 %	(1786)

Tabelle 2: Bewertungen als +gegliedert in den einzelnen Tests

Serie 1	40x4=160	102	63,75 %
Test C	80x5=400	280	70,00 %
Serie 2	40x4=160	110	68,75 %
Test E	105x5=525	156	29,71 %
Test 1	25x5=125	80	64,00 %
Test 2	25x4=100	46	46,00 %
Test B	80x5=400	123	30,75 %
Test A	40x5=200	111	55,5 %
Test D	40x3=120	71	60,8 %
Test I	80x5=400	163	40,75 %
Test III	80x4=320	240	75,0 %
Test II	95x5=475	304	64,0 %

6. Literatur

- COOPER, W.E., DANLY, M. (1981): "Segmental and Temporal Aspects of Utterance Final Lengthening". *Phonetica* 38, 106-110.
- POMPINO-MARSCHALL, B. u.a. (1982): "Does the Closed Syllable Determine the Perception of 'Momentary Tempo'?". *Phonetica* 39, 358-367.
- POMPINO-MARSCHALL, B. u.a. (1984): "'Koartikulation' und 'Steuerung' in der Wahrnehmung des 'momentanen Tempos'". *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 19, 306-314.
- TILLMANN, H.G., MANSELL, P. (1980): *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. München: Klett-Cotta.
- VENCOV, A.V. (1981): "Temporal Information Processing in Speech Perception". *Phonetica* 38, 193-203.

WECHSELBEZIEHUNG ZWISCHEN SATZMODALITÄT UND AKZENTUIERUNG IN SATZFINALER POSITION BEI DER REALISIERUNG VON INTONATIONSKONTUREN

Thomas P. Goldbeck, Walter F. Sendlmeier, (Gießen/Nijmegen)

ZUSAMMENFASSUNG

In der vorliegenden Arbeit wird untersucht, inwieweit zwischen den intonatorischen Mitteln, die einerseits zur Kennzeichnung der Satzmodalität und andererseits zur Akzentuierung bestimmter Wortteile in satzfinaler Position verwendet werden, Wechselbeziehungen bestehen. Die akustische Analyse der hier zugrunde gelegten Daten belegt, daß in deklarativen Sätzen der bedeutungsdifferenzierende Wortakzent auch in finalen Wörtern primär durch die relative Tonhöhe realisiert wird. Die Funktionen Modalisierung und Akzentuierung beeinflussen sich gegenseitig nicht. Für interrogative Sätze ergibt sich hingegen, daß sich die Opposition "betont vs. unbetont" durch die Dauer und/oder den Zeitpunkt des F₀-Anstiegs manifestiert, wobei unterschiedliche Verlaufsmuster zwischen Zweisilbern gegenüber Drei- und Viersilbern festzustellen sind. Darüberhinaus wurde auch die perzeptive Relevanz dieser Parameter untersucht, in dem die akustisch analysierten Parameter modifiziert und resynthetisiert Hörern in Identifikationsaufgaben zur Beurteilung dargeboten wurden.

1. EINLEITUNG

In der phonetischen und linguistischen Forschung wurde der Untersuchung der Intonation in den letzten Jahren verstärkte Aufmerksamkeit gewidmet. Dabei

geht man allgemein davon aus, daß von den akustischen Parametern, die für die Intonation verantwortlich sind, der Signalfrequenz primäre Relevanz zukommt. Gleichwohl stellt bereits Denes (1959) fest, daß sich die von der Intonation übertragene Information auf der akustischen Ebene durch ein komplexes Zusammenspiel der Parameter Signalfrequenz, Intensität und Dauer manifestiert (vgl. auch Fry, 1968).

Zur linguistischen Funktion der Satzintonation einer Sprache gehört zum einen die Steuerung der Satzmodalität, d.h., daß durch bestimmte suprasegmentale Verlaufsmuster Äußerungen als terminal, progredient oder interrogativ gekennzeichnet werden. Zum anderen hat die Intonation gleichzeitig die Funktion, einzelne Wort- oder Satzteile hervorzuheben. Inwieweit zwischen den intonatorischen Mitteln, die einerseits zur Kennzeichnung der Satzmodalität und andererseits zur Akzentuierung bestimmter Wortteile in satzfinaler Position verwendet werden, Wechselbeziehungen bestehen, soll in der vorliegenden Arbeit untersucht werden. Daß solche Wechselbeziehungen bestehen, wird einsichtig an Wörtern, die bedeutungsdifferenzierenden Wortakzent haben und in finaler Satzposition stehen. Spricht man etwa den Satz 'Sie erkennen den Tenor' mit der Betonung auf der ersten Silbe des Wortes 'Tenor' (Ténor) bzw. mit Endbetonung im Wort 'Tenor' (Tenór) jeweils als Aussage und als Frage, so wird deutlich, daß bei diesen vier Mustersätzen die Realisierung der Hervorhebung im Intonationsverlauf variiert wird.

2. METHODE

2.1 HERSTELLUNG DES TESTMATERIALS

Folgende 16 Sätze liegen der Untersuchung zugrunde:

1. Das wird módern!
2. Das wird módern?
3. Das wird modérn!
4. Das wird modérn?

5. Sie erkennen den Ténor!
6. Sie erkennen den Ténor?
7. Sie erkennen den Tenor!
8. Sie erkennen den Tenor?

9. Sie wollten das Hindernis umfahren!
10. Sie wollten das Hindernis umfahren?
11. Sie wollten das Hindernis umfahren!
12. Sie wollten das Hindernis umfahren?

13. Sie können übersetzen!
14. Sie können übersetzen?
15. Sie können übersetzen!
16. Sie können übersetzen?

In jeweils vier Sätzen sind die Oppositionen des bedeutungsdifferenzierenden Akzentes und der Satzmodalität kombiniert. Diese Sätze wurden im Tonstudio des Institutes für Kommunikationsforschung und Phonetik der Universität Bonn von zwei Sprechern je viermal an verschiedenen Tagen realisiert, so daß sich insgesamt 128 Äußerungen ergaben.

2.2 AKUSTISCHE ANALYSEN

Die Analysen der akustischen Merkmale wurden an der Universität Gießen, Fachbereich Psychologie, mit einem Signalfrequenzanalyseprogramm auf LPC-Basis (ILS-STI; Rechner:PDP 11/35) durchgeführt. Sie führten zur Extraktion der F0-Werte, sowie zur Berechnung der Intensitäten und Dauern, die in einzelnen 'Frames' abgespeichert wurden. Aus den Realisierungen beider Sprecher wurden für jede Kategorie (Modalitätstyp und Akzentposition) die mittleren F0-Konturen berechnet. Hierzu wurden die Silben zunächst hinsichtlich ihrer Framezahl in Übereinstimmung gebracht; danach wurden zur Berechnung der Standardkonturen die zeitlich korrespondierenden Tonhöhenwerte gemittelt; die errechneten Parameter der in Frage stehenden Wörter wurden als Kurvenverläufe graphisch dargestellt. (Tektronix-Plotter). Diese so gewonnenen F0-Muster können als die Standardkonturen für die zwei Sprecher und die wiederholten Produktionen angesehen werden. Diese Standardkonturen wurden durch Schematisierung in eine abstrakte Form überführt, indem die F0-Variationen, die innerhalb einer Silbe weniger als 15 Hz betragen, als Level dargestellt werden. Kurvenverläufe mit einer größeren Variation - "Up- oder Downglides" - sind als lineare Steigungen bzw. Neigungen wiedergegeben (vgl. Rossi, 1971).

Für die Intensitätsverläufe wurden ebenfalls durch Mittelungen Standardkonturen abgeleitet. Die Berechnungen der mittleren Silbendauern sind implizit in den Angleichungen der Anzahl der Frames enthalten, die den Mittelungen der F0- und Intensitätsverläufe vorausgehen.

3. ERGEBNISSE

3.1 ZWEISILBER:

In deklarativen Sätzen bestätigt sich für die hier untersuchten Zweisilber (auch) in satzfinaler Position die allgemeine Ansicht, daß sich die Hervorhebung einer Silbe deutlich in der relativen Höhe ihrer F0-Kontur manifestiert. Allerdings zeigt sich ebenfalls, daß die betonten Silben eine längere Dauer aufweisen. Gleichzeitig wird die Satzmodalität durch einen Fall des F0-Verlaufs auf der finalen Silbe angezeigt (vgl. Abb. 1a). In interrogativen Sätzen hingegen finden sich erheblich tiefere Frequenzwerte für die betonten ersten Silben als in deklarativen Sätzen. Sie liegen etwa auf der Höhe der unbetonten Silben der deklarativen Sätze. Die wortinitialen unbetonten Silben in satzfinaler Position der interrogativen Sätze weisen gegenüber den entsprechenden betonten Silben keine wesentlichen Unterschiede im Grundfrequenzbereich auf. Dagegen liegen hier sehr ausgeprägte Dauerunterschiede vor. Dies gilt ebenfalls für die Unterscheidung von betonter und unbetonter Silbe in finaler Stellung. Die Kürze der unbetonten zweiten Silbe führt dazu, daß der zur Indizierung der Satzmodalität 'Frage' notwendige F0-Anstieg zu einem frühen Zeitpunkt in der Silbe beginnt (vgl. 't Hart, 1974). In der betonten zweiten Silbe steigt dagegen die Kurve erst später an (vgl. Abb. 1b). Betrachtet man den Intensitätsverlauf des Sprachsignals an dem Wendepunkt der F0-Kontur, so kann man feststellen, daß die Energie der betonten finalen Silbe genau dort ein relatives Maximum aufweist (vgl. Abb. 2).

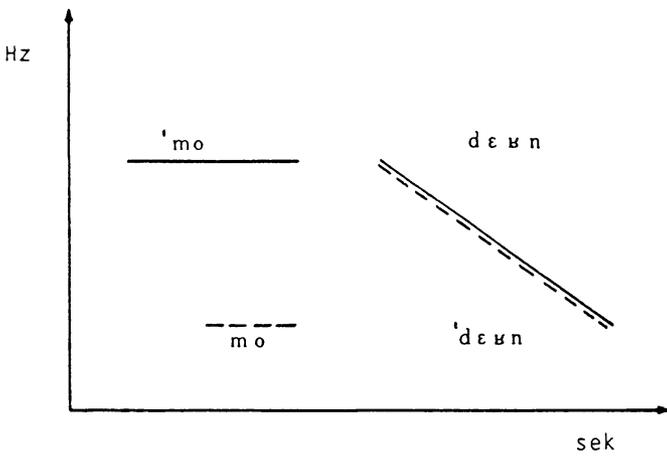


Abb. 1a: Schematisierte F0-Kontur in deklarativer Form

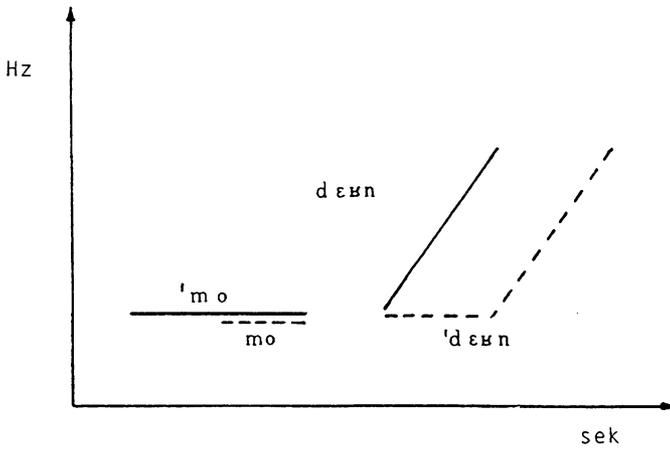


Abb. 1b: Schematisierte F0-Kontur in interrogativer Form

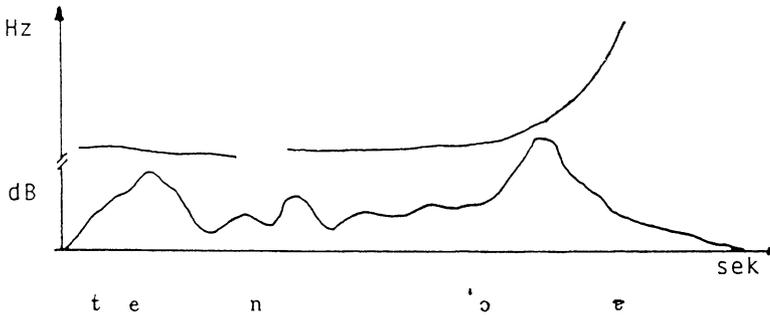


Abb. 2: Intensitätsmaximum am Wendepunkt der F0-Kontur bei betonter finaler Silbe in interrogativer Form

3.2 DREI- UND VIERSILBER

Für das dreisilbige Wort umfahren/umföhren ist in den deklarativen Sätzen wieder die Manifestation der Hervorhebung im Grundfrequenzbereich zu erkennen. Dauerunterschiede sind hier nicht festzustellen. Bei den interrogativen Sätzen zeigt sich, - wie bereits bei den Zweisilbern - daß unabhängig von der Betonung der initialen Silbe, keine unterschiedlichen F₀-Werte oder Dauern realisiert werden; d.h., daß auf der Silbe keine Parameterunterschiede aufzufinden sind. Betrachtet man die zweiten Silben, so stellt man fest, daß sich der F₀-Verlauf dieser Silbe, wenn sie betont ist, gegenüber der vorausgehenden Silbe nicht verändert, d.h. tief bleibt. Ist jedoch die erste Silbe betont, so ist auf der unbetonten zweiten Silbe ein Anstieg des F₀-Verlaufs zu beobachten und zwar unmittelbar nach dem Beginn der Silbe. Voraussetzung für die Signalisierung Satzmodalität 'Frage' ist die finale Steigung, die im Falle der betonten zweiten Silbe erst auf der finalen (dritten) Silbe beginnt, während im Falle der unbetonten zweiten Silbe dieser Anstieg der F₀-Kontur bereits in dieser Silbe beginnt und sich in der dritten Silbe fortsetzt. Da für die Unterscheidung der Betonung in der initialen Silbe keine Parameterunterschiede zur Verfügung stehen, liegt es nahe anzunehmen, daß die Betonung der ersten Silbe durch den Zeitpunkt des F₀-Anstiegs im weiteren Signalfrequenzverlauf gekennzeichnet ist. Folgt ein früher Anstieg zur Kennzeichnung der Satzmodalität 'Frage', so ist die erste Silbe betont (vgl. Abb. 3a und 3b).

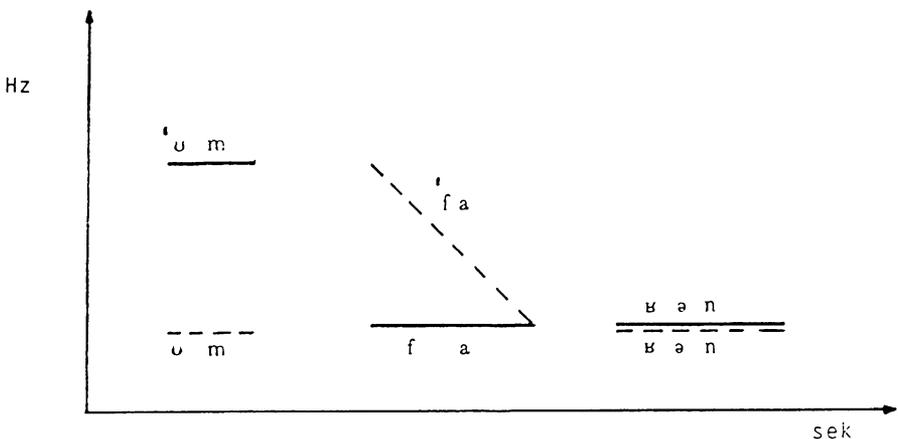


Abb. 3a: Schematisierte F₀-Kontur in deklarativer Form

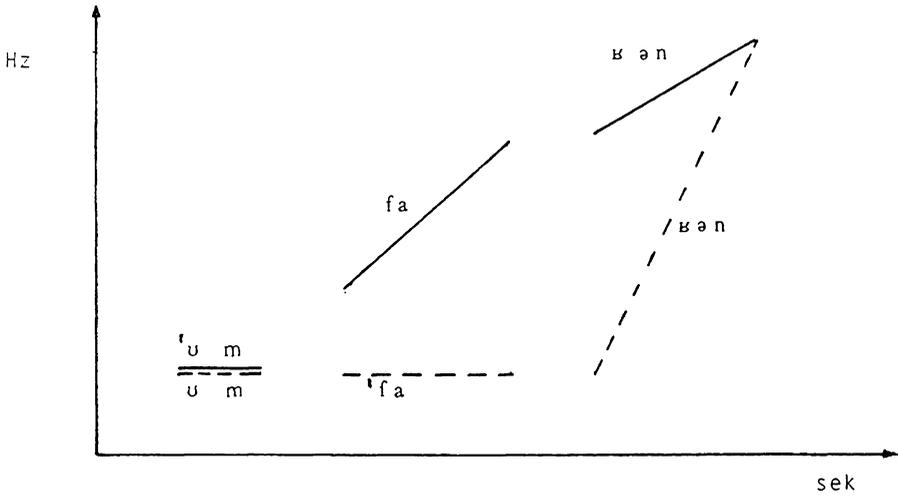


Abb. 3b: Schematisierte F0-Kontur in interrogativer Form

F0 als akzenttragender Parameter ist auch für die viersilbigen Wörter übersetzen/übersétzen in den deklarativen Formen zu identifizieren. Dauerunterschiede treten wie bereits bei den Dreisilbigen nicht mehr auf. In der interrogativen Form scheint sich die Tendenz zu bestätigen, daß die Akzentuierung auf der ersten Silbe durch einen unmittelbaren Anstieg auf der folgenden Silbe getragen wird. Genau wie im hier untersuchten Dreisilber ändert sich die F0-Struktur kaum, wenn die betonte Silbe die Pänultima ist; erst danach steigt der F0-Verlauf zur Kennzeichnung der Modalität 'Frage' an (vgl. Abb. 4a und 4b).

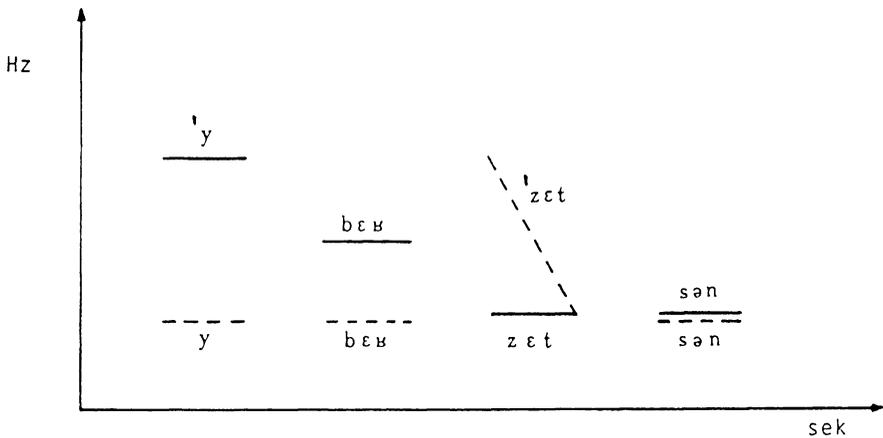


Abb. 4a: Schematisierte F0-Kontur in deklarativer Form

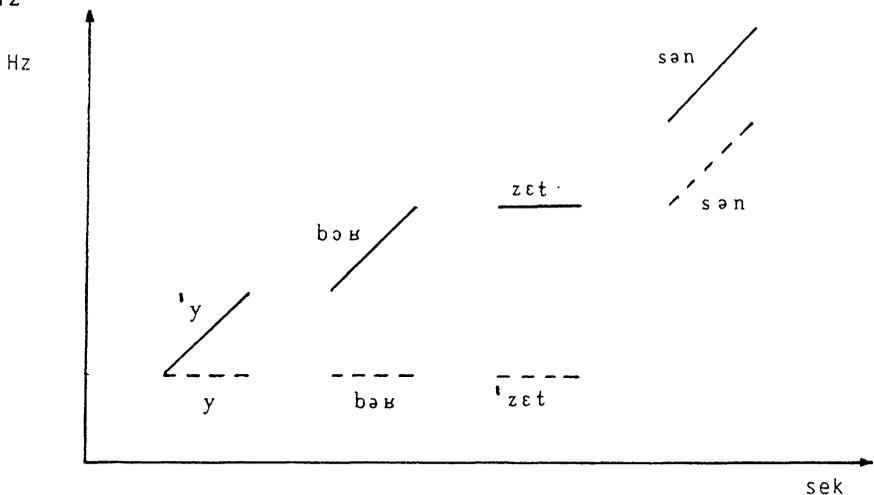


Abb. 4b: Schematisierte F0-Kontur in interrogativer Form

4. ZUSAMMENFASSUNG DER ERGEBNISSE DER AKUSTISCHEN ANALYSE

Die akustische Analyse der hier zugrundegelegten Daten belegt, daß in deklarativen Sätzen der bedeutungsdifferenzierende Wortakzent auch in finalen Wörtern primär durch die relative Tonhöhe realisiert wird. Die Funktionen Modalisierung und Akzentuierung beeinflussen sich gegenseitig nicht. Für interrogative Sätze ergibt sich hingegen, daß der Wortakzent in satzfinalen Wörtern nicht durch den Tonhöhenverlauf auf der betonten Silbe differenziert werden kann. Vielmehr zeigt sich hier, daß die Opposition betont vs. unbetont durch den Zeitpunkt des F0-Anstiegs (zur Signalisierung der Satzmodalität Frage) realisiert wird: ist die erste Silbe betont, so steigt F0 unmittelbar am Beginn der zweiten Silbe an; ist die zweite Silbe in Zweisilbern akzentuiert, so bleibt F0 fast bis an das Ende der zweiten Silbe tief und steigt erst dann steil an; folgt in Mehrsilbern der betonten nicht-ersten Silbe eine weitere Silbe, so bleibt F0 die gesamte Dauer der betonten Silbe tief und wird erst in der finalen Silbe angehoben. In allen Fällen ist der Beginn des Anstiegs durch ein Intensitätsmaximum gekennzeichnet.

Bei Zweisilbern fällt weiterhin besonders auf, daß mit der Unterscheidung betont vs. unbetont ein deutlicher Dauerunterschied zwischen der betonten und

unbetonten Silbe einhergeht. Da in dieser Untersuchung lediglich die akustische Manifestation der linguistischen Funktion Wortakzent in Abhängigkeit von der Satzmodalität betrachtet wurde, kann über die unterschiedliche Relevanz der Parameter Dauer bzw. Zeitpunkt der F0-Anstiegs, sowie insbesondere die Interaktion von Intensitätsgipfel und F0-Anstiegszeitpunkt für den Hörer noch keine Aussage gemacht werden. Um die perzeptive Relevanz dieser Parameter einschätzen zu können, wurden deshalb Wahrnehmungsexperimente durchgeführt.

5. PERZEPTIONSEXPERIMENTE

5.1 HERSTELLUNG DES STIMULSMATERIALS

Da die akustische Analyse der deklarativen Sätze eindeutig die relative Höhe des F0-Verlaufs als den mit der linguistischen Funktion variierenden Parameter auswies, wurden nur die Fragesätze berücksichtigt, denn für sie konnte die rein experimentell deskriptive Zugangsweise nur ungenügenden Einblick in die funktionale Bedeutung leisten.

Mit einem Computergraphikprogramm zur interaktiven Manipulation von Signalkennwerten wurden die analysierten Parameter der Wörter nach der Vorlage der jeweiligen Oppositionspartner einzeln nachgebildet (vgl. Abb. 5). So wurde zum Beispiel die Vertauschung der F0-Konturen der Wörter módern/modérn in der Weise durchgeführt, daß der Beginn des F0-Anstiegs innerhalb der zweiten Silbe von dem frühen Zeitpunkt an das Ende der Silbe verschoben wurde und entsprechend umgekehrt (vgl. Abb. 1b). Der Faktor Silbendauer wurde so verändert, daß die Finale lange Silbe des Wortes modérn gekürzt und die initiale kurze Silbe gleichzeitig nach den entsprechenden Werten des Oppositionspartners 'módern' gedehnt wurde. Der Austausch der Intensitätsgipfel wurde ebenfalls durch die Messwerte der RMSS-Maxima in dem Minimalpaarpartner bestimmt. Dabei ist zu berücksichtigen, daß die akustische Analyse für die Drei- und Viersilber keine Dauerunterschiede ergab; dennoch sind auch bei diesen Stimuli die Silbendauern durch Halbierung bzw.

Verdopplung geändert worden, um einen potentiellen Einfluß dieses experimentellen Faktors auch bei den Drei- und Viersilbern kontrollieren zu können. Auf diese Weise ergab sich ein varianzanalytisches 2x2x2 Design mit den Faktoren Silbendauer, Lage des Intensitätsmaximums und F0- Anstiegszeitpunkt mit jeweils einer zweifachen Ausprägung (Silbendauer: kurz/lang; Intensitätsmaximum: vorne/hinten; F0-Anstieg: früh/spät).

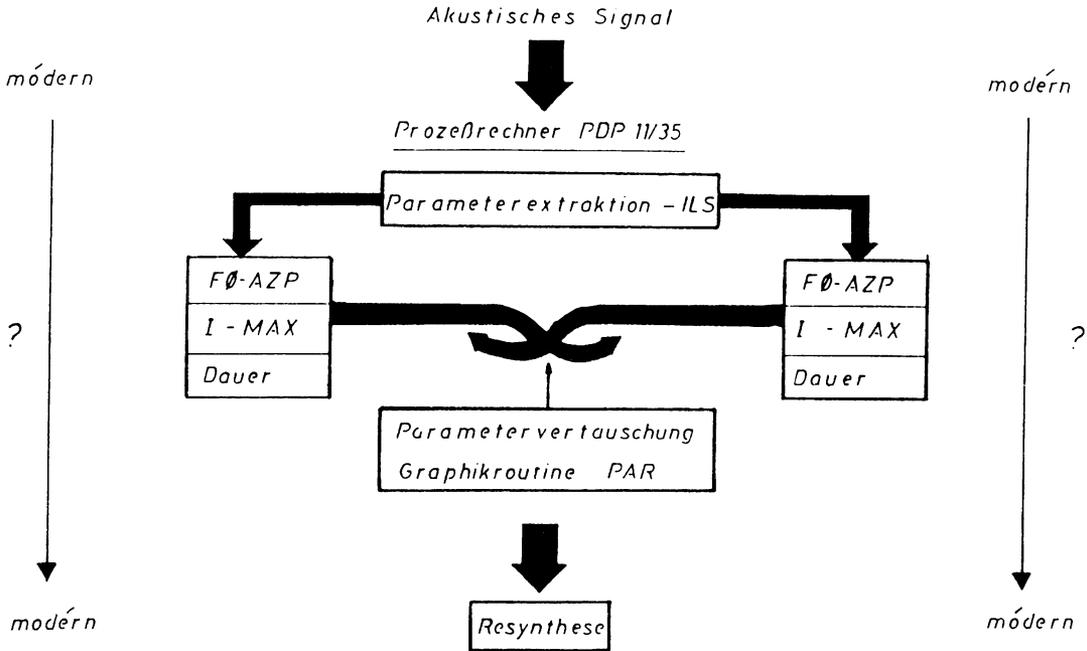


Abb. 5: Herstellung der Teststimuli für das Perceptionsexperiment
(F0-AZP = F0-Anstiegszeitpunkt, I-Max = Intensitätsmaximum)

5.2 VERSUCHSDURCHFÜHRUNG

Die veränderten und resynthetisierten Stimuli wurden Versuchspersonen in einer kategorialen Entscheidungsaufgabe zur Identifikation in einem reflexionsarmen Raum über einen Großlautsprecher dargeboten. An dem Versuch nahmen 21 Studenten (9 weiblich, 12 männlich) der Universität Bonn teil. Die Stimuli wurden in randomisierter Reihenfolge jeweils achtmal dargeboten.

6. ERGEBNISSE DES PERZEPTIONSEXPERIMENTES

Die varianzanalytischen Auswertungen der Wahrnehmungsdaten führte zu den in Tabelle 1 wiedergegebenen signifikanten Effekten.

Signifikanzniveau $p = .001$

	Haupteffekte			2x Interaktion			3x Interaktion
	Dauer //	F0 //	Int ///	D-F0 //	D-I //	F0-I ///	D-F0-I
MOdern	1516.03	93.03	98.46	27.30			
moDERn	847.78	119.45	221.32	157.34	183.		
TEnor	2355.21		73.76				
teNOR	433.46	222.31	106.74	44.12	50.21		
UMfah	1327.74	955.6	50.71	759.14		54.49	
umFAH	1318.87	35.36					
ÜBerse	220.33	117.36	56.23	44.82	78.		
überSE	268.88	110.99	121.08				

Tabelle 1: F-Werte der signifikanten Effekte

Der Faktor Dauer weist bei allen hier zugrundegelegten Wörtern den stärksten Haupteffekt für die Identifikationsleistung auf (vgl. Abb. 6a). Bei den drei Stimuli unfähren, ténor, módern führte die alleinige Veränderung der Silbendauern zu einem Umschlag in der wahrgenommenen Wortkategorie. Bei den Stimuli übersetzen, übersétzen, tenór, modérn wirkt sich die Veränderung der Zeitstruktur in der Weise aus, daß sich die Erkennungsleistung für die Ausgangsstimuli von einer klaren Identifikation hin zu einer zufälligen Einschätzung verschiebt. Bei dem Wort úmfahren zeigt sich, daß die Identifikation der Akzentposition aufgrund der ursprünglichen Dauerverhältnisse nur zufällig erfolgt, während eine Veränderung der Silbendauern deutlich zur Identifikation von umfähren, also zur Verschiebung der Akzentposition führt.

Für den Faktor F0-Anstiegszeitpunkt finden sich bis auf das Wort ténor ebenfalls signifikante Haupteffekte. Wenn auch damit ein Einfluß auf die Wahrnehmung der Akzentposition belegt ist, so führt er jedoch in keinem Fall zu einem kategorialen Umschlag (vgl. Abb. 6b). Eine Sonderstellung kommt aber auch hier wieder dem Wort úmfahren zu, bei dem Variation des

F₀-Anstiegszeitpunktes die Wahrnehmung der Hörer in ähnlicher Weise beeinflusst wie der Faktor Dauer. Die Veränderung der Intensität, die als dritter Faktor mit Ausnahme des Wortes umfährén ebenso signifikante Mittelwertsunterschiede bewirkt, führt zu vergleichbaren Verschiebungen in der Wahrnehmung wie die Variation des Zeitpunktes des F₀-Anstiegs (vgl. Abb. 6c). Ähnlich wie bei den beiden anderen Haupteffekten zeigt sich wieder für das Wort úmfahrén, daß die Position des Wortakzents im Ausgangsstimulus allein aufgrund des kontrollierten Faktors nicht von den Hörern bestimmt werden kann.

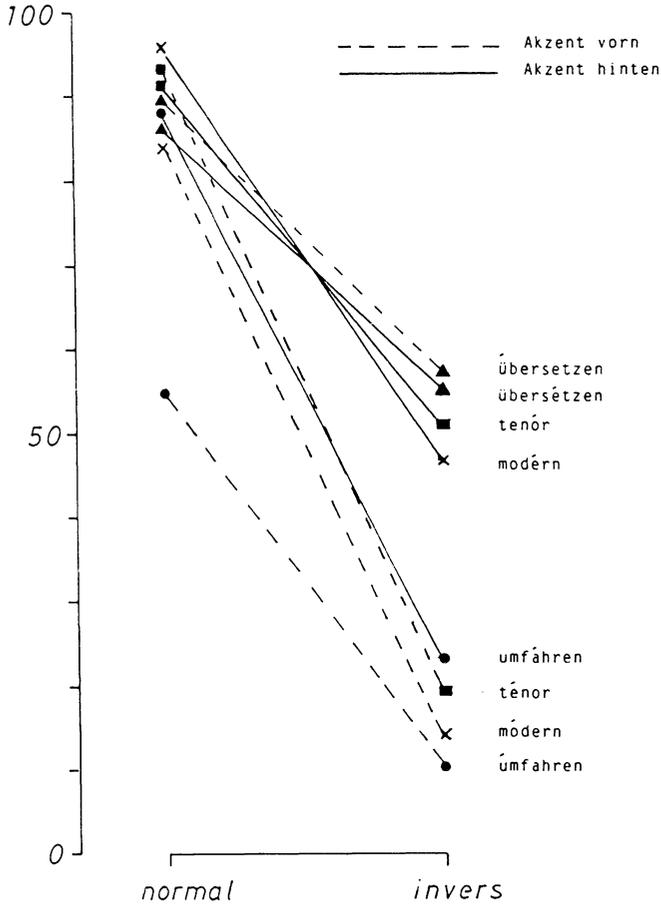


Abb. 6a: Mittelwerte für den Haupteffekt des Faktors Dauer

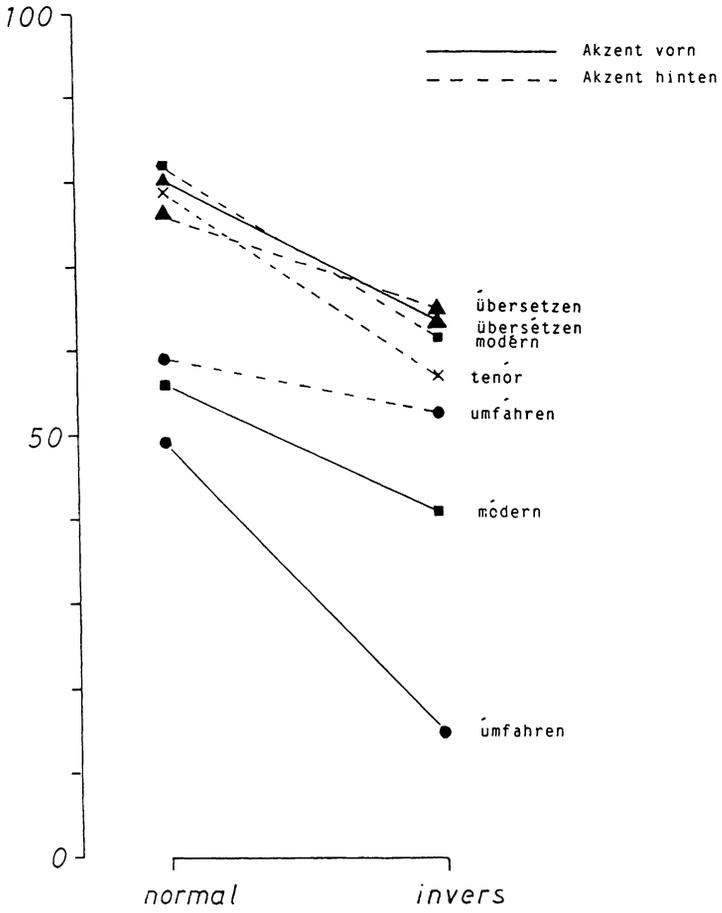


Abb. 6b: Mittelwerte für den Haupteffekt des Faktors F0-Anstiegszeitpunkt

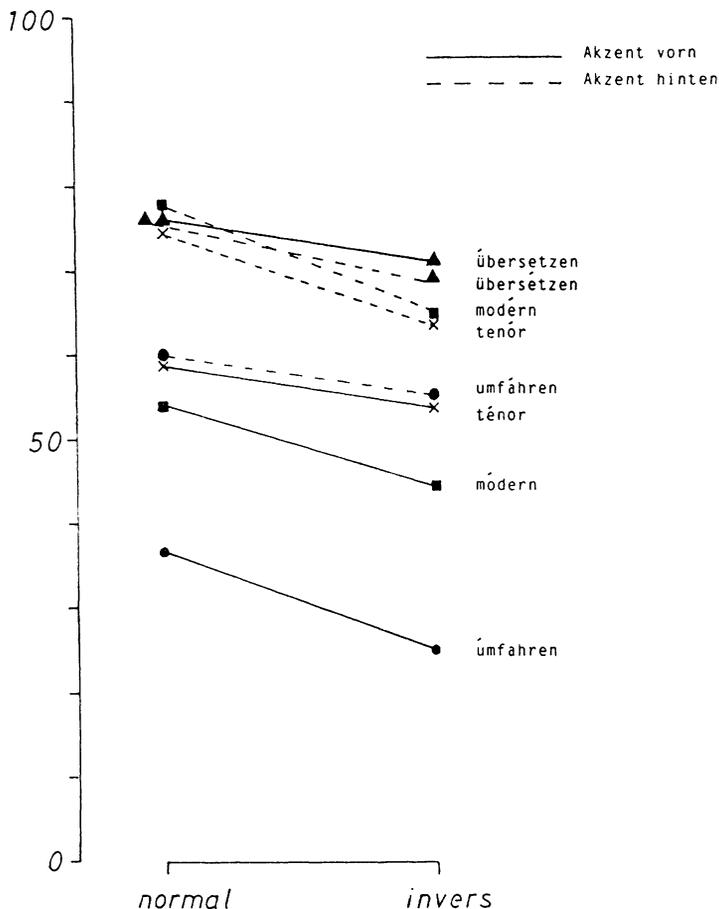


Abb. 6c: Mittelwerte für den Haupteffekt des Faktors Intensitätsmaximum

Neben den relativ schwachen Haupteffekten für das Wort umfahren findet sich eine deutliche Interaktion der experimentellen Faktoren Dauer und F₀-Anstiegszeitpunkt. Wenn der Parameter Dauer unmodifiziert ist, bewirkt die Verschiebung des Anstiegszeitpunktes von F₀ von früh nach spät einen kategorialen Umschlag der wahrgenommenen Wortbedeutung. Sind jedoch die Dauerverhältnisse der in Frage stehenden Silben jeweils invertiert, so wird unabhängig von der Position des F₀-Anstiegszeitpunktes die zweite Silbe als betont wahrgenommen (vgl. Abb. 7).

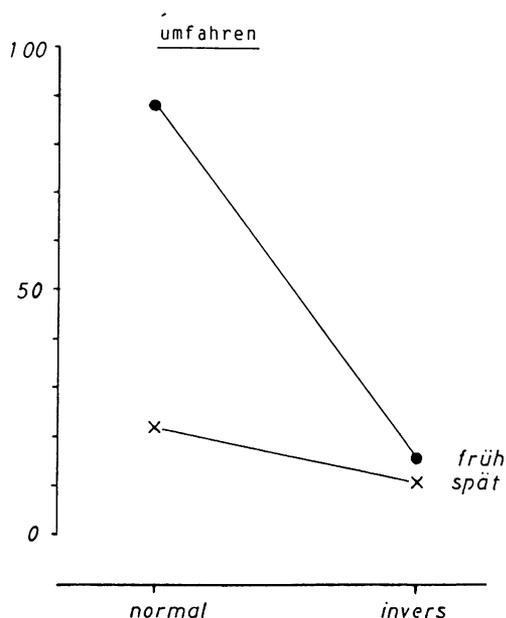


Abb. 7: Mittelwerte für die Interaktion der Faktoren Dauer und F0-Anstiegszeitpunkt

Für die Stimuli übersetzen, übersétzen, tenór, modérn führte keiner der signifikanten Haupteffekte zu einem Kategorienumschlag (vgl. Abb. 6a-c). Bei den beiden Zweisilbern modern, tenor fanden sich über die Haupteffekte hinaus signifikante Interaktionen zwischen den Faktoren Dauer und F0-Anstiegszeitpunkt bzw. Dauer und Intensitätsmaximum, die jeweils zu einer Veränderung der wahrgenommenen Akzentposition führten. Abbildung 8a veranschaulicht am Beispiel des Wortes modérn, daß die Modifikation des F0-Anstiegszeitpunktes bei unveränderter Silbendauer nur einen sehr schwachen Effekt aufweist; wird hingegen die Dauer verändert, so unterstützt die Verschiebung des F0-Anstiegszeitpunktes den Effekt auf die Wahrnehmung derart, daß die Lokalisierung der Akzentposition auf die erste Silbe verlegt wird. In gleicher Weise interagiert die Dauer mit dem Faktor Intensitätsmaximum (vgl. Abb. 8b). Lediglich bei den Viersilbern konnte auch durch das Zusammenwirken der verschiedenen Faktoren keine Veränderung der Akzentposition im Urteil der Hörer erreicht werden.

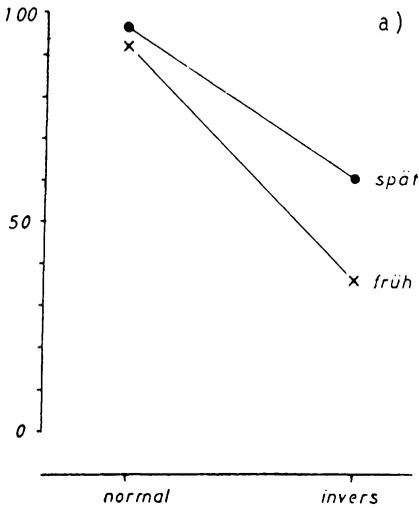


Abb. 8a

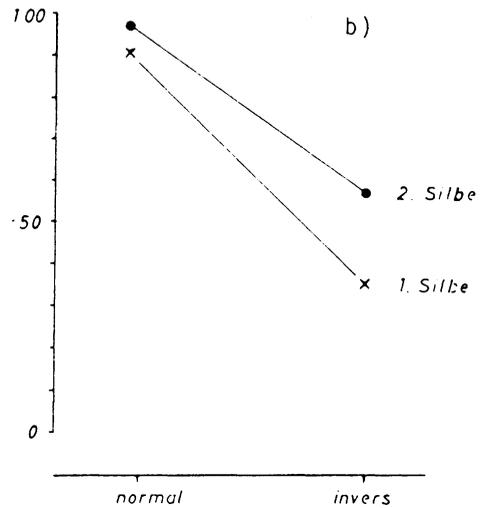


Abb. 8b

Mittelwerte der Interaktionen für die Faktoren Dauer - F0-Anstiegszeitpunkt (Abb. 8a) und Dauer - Intensitätsmaximum (Abb. 8b)

7. DISKUSSION

Die Perceptionsexperimente bestätigen die aus der akustischen Analyse abgeleitete Annahme, daß die Realisierung des distinktiven Wortakzents eines in satzfinaler Position stehenden Zweisilbers primär durch den Parameter Dauer bzw. Dauer in Kombination mit dem Faktor F0-Anstiegszeitpunkt erfolgt, wenn gleichzeitig interrogative Satzmodalität indiziert werden soll. Für die Dreisilber liefern die Daten ebenfalls Evidenz für die perzeptive Relevanz des Faktors Dauer; hierbei ist jedoch zu berücksichtigen, daß die akustische Analyse für die Drei- und Viersilber ursprünglich keine Dauerunterschiede der in Opposition stehenden Silben aufwies. Daher sollten die Faktoren F0-Anstiegszeitpunkt und Intensitätsmaximum bei der Interpretation dieser Stimuli in stärkerem Maße als relevante Parameter für die Wahrnehmung der Akzentposition betrachtet werden. Sieht man von dem Einfluß der Dauer ab, so

scheint zumindest für die Dreisilber der Faktor F0-Anstiegszeitpunkt für die Akzentuierung primär relevant zu sein. Die Ergebnisse dieser Untersuchung legen es nahe, die Gültigkeit der Annahme, daß die Grundfrequenz primär akzentindizierender Parameter sei, einzuschränken.

Die geringen Effekte für die Viersilber können dadurch erklärt werden, daß die beiden potentiell betonten Silben durch eine unbetonte Silbe getrennt sind und daß ihnen auch noch eine unbetonte Silbe folgt. Die Konfliktsituation für die gleichzeitige Realisierung der Satzmodalität und des Wortakzents ist folglich auch nicht in dem Maße wie bei den Zwei- und Dreisilbern gegeben.

LITERATUR

- Denes, Peter. 1959. A Preliminary Investigation of Certain Aspects of Intonation. In: *Language and Speech*, 2, S. 106-122.
- Fry, Dennis B. 1970. Prosodic Phenomena. In: B. Malmberg, *Manual of Phonetics*, Amsterdam, S. 365-410.
- 't Hart, Johan. 1974. Discriminability of the Size of Pitch Movements in Speech. In: *Annual Progress Report of the Institute for Perception Research* 11, Eindhoven. S. 56-63.
- Rossi, Mario. 1971. Le Seuil de Glissando ou Seuil de Perception des Variations Tonales pour les Sons de la Parole. In: *Phonetica*, Bd. 35, S. 11-40.
- Anmerkung: Teilergebnisse dieser Untersuchung sind bereits in der 'Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung', 40, S. 248-253 unter dem Titel "Wechselbeziehung zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen" publiziert.