

Fokus, Modus und die große Zahl: zur intonatorischen Indizierung des Fokus im Deutschen

Anton Batliner

Angaben zur Veröffentlichung / Publication details:

Batliner, Anton. 1989. "Fokus, Modus und die große Zahl: zur intonatorischen Indizierung des Fokus im Deutschen." In *Zur Intonation von Modus und Fokus im Deutschen*, edited by Hans Altmann, 21–70. Tübingen: Niemeyer.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



FOKUS, MODUS UND DIE GROSSE ZAHL. ZUR INTONATORISCHEN INDIZIERUNG DES FOKUS IM DEUTSCHEN.

Anton Batliner

1. Einleitung

1.1 Überblick

In diesem Beitrag wird anhand eines von mehreren Sprechern produzierten, kontrollierten Korpus die intonatorische Markierung des Satzfokus im Deutschen bei Fragen und Nicht-Fragen untersucht. Mit Hilfe eines statistischen Klassifikationsverfahrens (Diskriminanzanalyse) werden die für die Markierung relevanten intonatorischen Parameter und Transformationen der Parameterwerte bestimmt. Durch Prototypenbildung 'über die große Zahl' (Mittelwertbildung) und über Hörerurteile werden 'normale', prototypische Fälle von 'nicht-normalen', aber dennoch regulären Fällen abgegrenzt sowie die sprecherübergreifende und die sprecherspezifische Relevanz der einzelnen intonatorischen Merkmale untersucht.

1.2 Begriffsklärung, Stand der Forschung

Der Begriff 'Satzmodus' wird von uns im Sinne des Satzmodusmodells von Altmann (1987) gebraucht. 'Fokus' bezeichnet den semantisch 'wichtigsten' Teil des Satzes, der durch den Fokusakzent (FA) gekennzeichnet ist. Mit phonetischen Mitteln kann der FA **intonatorisch** durch Grundfrequenz (Fo), Dauer und Intensität indiziert werden, **nicht-intonatorisch** durch damit z.T. gekoppelte Phänomene, wie spektrale Unterschiede oder Pausensetzung; auf die nicht-intonatorischen Merkmale werden wir nicht weiter eingehen. Im folgenden verzichten wir normalerweise auf eine explizite Trennung der physikalischen Größen von ihren perzeptiven Korrelaten (Fo vs. Tonhöhe, Intensität vs. Lautstärke, etc.). In dieser Arbeit sollen keine unterschiedlichen Akzentarten gekennzeichnet

werden, wie Normalakzent, Kontrastakzent, emphatischer Akzent o.ä. Im normalen (nicht-kontrastiven) Fall kann der FA nur auf einer 'betonbaren' Silbe realisiert werden, auf der ein Wortakzent liegt. Wortakzent und FA können mit den gleichen intonatorischen Mitteln markiert werden; im einzelnen ist natürlich zu untersuchen, wie diese Mittel eingesetzt werden (vgl. dazu Bannert 1985 und Thorsen 1987).¹

Normalerweise wird angenommen, daß für die Akzentmarkierung Fo am wichtigsten ist, Dauer weniger wichtig und Intensität am unwichtigsten; vgl. dazu den Literaturüberblick in Beckman (1986), die aber fürs Englische der Intensität größere Bedeutung beimißt. Die intonatorische Markierung des Akzents ist z.B. fürs Englische besser untersucht als fürs Deutsche; noch Klein (1980:32) mußte feststellen: "Es scheint mir kein gutes Zeichen für den Stand der deutschen Phonetik, daß völlige Unklarheit darüber herrscht, was das signalphonetische Gegenstück des 'Akzents' ist." Seitdem sind einige Arbeiten zur Form des Akzents im Deutschen erschienen, vgl. etwa Adriaens (1984), Bannert (1985), sowie einige Arbeiten im Sammelband von Altmann (1988). Allerdings sind die Fragestellungen meist speziell, so daß Generalisierungen nur vorsichtig gezogen werden können.

1.3 Methodik

Die traditionelle Intonationsforschung zog in den Augen des Forschers 'repräsentative' Fälle nicht nur zur Illustration, sondern auch zur Begründung von Regeln heran; zur Kritik an diesem Vorgehen vgl. Klein (1980, 1982). Diese 'qualitative' Vorgehensweise wurde in der neueren Forschung durch eine 'quantitative' abgelöst, bei der angegeben wird, auf der Grundlage welchen Materials Aussagen gemacht werden. Problematisch bleibt dabei noch immer der Schluß vom untersuchten Material auf die Gesamtheit (Schluß von der Stichprobe auf die Population), also etwa auf 'die intonatorische Fokusmarkierung im Deutschen'. Wir wollen nun drei in diesem Zusammenhang wichtige Punkte aufzeigen und skizzieren, welche Strategie wir jeweils einschlagen werden.

¹ Wir haben in diesem Abschnitt unsere Sprachregelung skizziert, ohne daß damit impliziert sein soll, daß wir sie für 'besser' halten als andere Regelungen, wie etwa eine Einschränkung des Begriffs 'Intonation' auf Grundfrequenz/Tonverlauf, die Verwendung von 'Satzakzent' statt 'Fokusakzent' o.ä.

1.3.1 Sprecherauswahl

Sprecherauswahl und damit der Umfang des Materials sind zuallererst ein ökonomisches Problem; man kann nicht versuchen, alle Sprecher des Deutschen (das es als Einheit letztlich gar nicht gibt) zu repräsentieren, sondern muß sich auf bestimmte Varietäten beschränken. Die von uns untersuchte Varietät ist die süddeutsch/bairisch gefärbte Standardsprache. Alle im folgenden gemachten Aussagen beziehen sich also nur auf diese Varietät. Es wurden sechs Sprecher (3 weibliche, 3 männliche) gewählt. Damit können schon vorsichtige Schlüsse z.B. über die Repräsentativität der Produktionen einzelner Sprecher gezogen werden.² Die Sprecher werden nicht nur interpretativ verglichen (via Inspektion von Kurvenverläufen o.ä.), sondern mit Hilfe eines statistischen Verfahrens. Damit lassen sich objektive, in Zahlen ausdrückbare Angaben machen.

1.3.2 Sprecherinstruktion

Oft werden die Sprecher angewiesen, auf welcher Silbe der Akzent gesetzt werden soll, welche Emphasestufe verwendet werden soll, etc. Auf diese Weise sollen 'Störvariablen' ausgeschaltet werden. Eine solche Instruktion entspricht allerdings nicht der Forderung nach größtmöglicher Nähe zur natürlichsprachlichen Kommunikation; man begibt sich dabei auch der Möglichkeit, auf interessante Abweichungen zu stoßen: Wenn man z.B. vorgibt, daß Fragen immer mit steigendem Tonverlauf produziert werden sollen, so wird man nie herausfinden, wann und in welchem Umfang Fragen einen fallenden Tonverlauf aufweisen. Wir gaben deshalb unseren Sprechern keine expliziten Instruktionen, sondern nur 'implizite', d.h. über eine Kontextvorgabe gesteuerte; im einzelnen vgl. dazu Oppenrieder (1988b) und Batliner/Oppenrieder (1989).

1.3.3 Konvergierende, sich stützende Strategien

So wie man bei Perzeptionsexperimenten mit manipulierten und damit per se unnatürlichen Stimuli nie sicher sein kann, ob man überhaupt das 'Richtige' oder

² Die Forderung der theoretischen Statistik nach zufälliger und repräsentativer Auswahl der Stichprobe aus der Population wird in der Praxis so gut wie nie erfüllt. Unser Begriff der 'Repräsentativität' ist denn auch bescheidener: Bei einem Sprecher hat man eine Kasuistik, bei zwei Sprechern zwei Kasuistiken. Bei sechs Sprechern könnte man sogar mit einfachen statistischen Tests (z.B. Chi-Quadrat) signifikante Unterschiede feststellen, wenn sich ein Sprecher von allen anderen unterscheidet. Ob die Stichprobe insgesamt repräsentativ ist, bleibt ein inhaltliches Problem und kann nur durch Replikation der Experimente mit anderen und größeren Stichproben gelöst werden.

doch nicht nur experimentelle Artefakte untersucht, so kann man bei Produktionsdaten nicht sicher sein, ob im Einzelfall 'richtig' produziert wurde, und welches perzeptive Korrelat das Gemessene hat (vgl. die Fehleranalyse in Teil 9). Alle unsere Äußerungen wurden deshalb den weiter unten beschriebenen Hörtests unterzogen.

1.4 Statistisches Modell

Das von uns gewählte Verfahren der Diskriminanzanalyse gehört zu den in der Mustererkennung (und damit auch in der automatischen Spracherkennung) üblichen Verfahren für die Klassifikation von Fällen in **bekannte Gruppen** anhand möglichst relevanter Merkmale (**Prädiktorvariablen**); es handelt sich um einen 'Klassifikator mit überwachtem Lernen' (vgl. Niemann 1974, Ruske 1988). Die Diskriminanzanalyse funktioniert ähnlich wie die multiple Regressionsanalyse: Es werden lineare Kombinationen der unabhängigen (Prädiktor-) Variablen gebildet, die möglichst optimal zwischen den Gruppen unterscheiden können. Jede Prädiktorvariable erhält einen Gewichtungskoeffizienten, der anhand der vorgegebenen Daten und Gruppenzugehörigkeit der Fälle so geschätzt wird, daß die resultierende Diskriminanzfunktion zwischen den Gruppen möglichst stark differiert; anders gesagt: die Variabilität **zwischen** den Gruppen soll im Verhältnis zur Variabilität **innerhalb** der Gruppen möglichst groß sein.

In der Mustererkennung sind für die Wahl des Verfahrens zum einen praktische Gründe (Verfügbarkeit, Zeitbedarf), zum anderen die Effizienz entscheidend, also die Güte der Klassifikation. In der Phonetik sollte man darüber hinaus das Modell wählen, das am ehesten die Prozesse im 'natürlichen Sprecher/Hörer' abbildet. Letztlich dürften beide Auswahlverfahren konvergieren, also das Modell am effizientesten sein, das die menschliche Fähigkeit am besten abbildet.³ Auch wenn man keine zu weitreichenden Schlüsse ziehen sollte, so wird sich doch zeigen, daß die Diskriminanzanalyse sowohl unsere Produktions- als auch unsere Perzeptionsdaten sehr gut abbildet. Die folgenden Eigenschaften lassen die Diskriminanzanalyse für unsere Zwecke geeignet erscheinen (zum Verfahren im einzelnen vgl. Norusis 1986 und Klecka 1980; eine Anwendung auf die Prädiktion des Modus zeigen Batliner 1988 und Batliner et al. 1989):

³ Diese Aussage produzierte bisher jedesmal entweder Widerspruch oder den Wunsch nach eingehender Begründung und im weiteren Verlauf Diskussionen über statistische Verteilungen, Künstliche Intelligenz und das Wesen des Menschen. Eine Begründung kann hier nicht gegeben werden, einen Beweis gibt es nicht, genausowenig wie für die gegenteilige Ansicht.

1. Die Hörerurteile, die wir unserer FA-Zuweisung zugrundelegen, erfüllen nicht unbedingt die Verteilungsvoraussetzungen für regressionsanalytische Verfahren (vgl. unten Fig.1), die deshalb weniger geeignet sind als Verfahren, bei denen die Fälle bekannten Klassen zugeordnet werden.
2. Die Umwandlung von kontinuierlichen Betonungsabstufungen in 'betont/unbetont' dürfte eine wichtige Stufe in der natürlichsprachlichen Kommunikation widerspiegeln: Der Mensch kann nur zwischen einer sehr begrenzten Zahl von Akzentuierungsstufen unterscheiden, und er wird sich ziemlich bald für eine relativ grobe Repräsentation des Gehörten in 'betont=wichtig' vs. 'unbetont=unwichtig' entscheiden, um nicht zuviel Information speichern zu müssen. (Im bisher einzigen Experiment zur Frage, wieviele Akzentstufen der Mensch konsistent mit alleiniger Hilfe intonatorischer Merkmale unterscheiden kann, kommt Lieberman 1965 auf zwei Stufen: 'betont' vs. 'unbetont'. Zur Umkodierung von Äußerungen bzw. Sätzen im Verstehensprozeß vgl. Hörmann 1978:460ff.)
3. Das Verfahren kann die für die Klassenzuweisung relevanten Merkmale (Prädiktorvariablen) auswählen und sie nach ihrer Relevanz einstufen. Es bietet die Möglichkeit, einzelne Variablen (univariate Analyse) und mehrere oder alle Variablen gleichzeitig (multivariate Analyse) zu testen.
4. Es können beliebige Teilstichproben (**Lernstichproben**) zum Trainieren der Klassenzuweisung anhand der Prädiktorvariablen sowie andere Teilstichproben zum Klassifizieren (**Prüfstichproben**) ausgewählt werden. Wir können systematisch einzelne Sprecher(-gruppen) zu anderen Sprechern in Beziehung setzen und Gemeinsamkeiten und Unterschiede herausarbeiten; anders gesagt: man kann Sprecherunabhängigkeit simulieren und damit das nachbilden, was der Mensch auch kann, nämlich 'richtige' Akzenturteile bei Äußerungen eines ihm bis dato unbekannten Sprechers abgeben (im einzelnen vgl. Teil 4).
5. Das Verfahren liefert quantifizierte Gütemaße, insbesondere Prozentwerte für die Zahl der richtig klassifizierten Fälle.

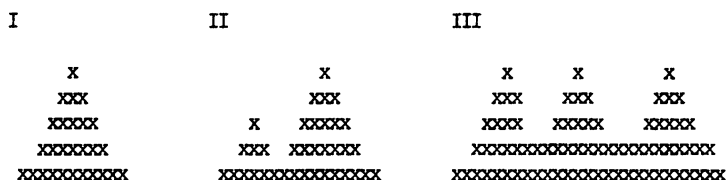
1.5 Das Prototypenkonzept

Medin/Barsalou (1987) weisen auf einen entscheidenden Unterschied zwischen zwei Auffassungen von Kategorien hin: Einmal sind Kategorien definiert durch Regeln, das andere Mal durch typische Vertreter, die viele bzw. charakteristische Merkmale der Kategorie aufweisen; die Verteilung der Merkmale ist aber nicht einfach durch Regeln definierbar: Die Kategorien sind **prototypisch** strukturiert. Wir nehmen analog zur zweiten Auffassung an, daß es (proto-) typische Realisierungen bestimmter Akzentstrukturen gibt, und daß eine Definition dieser Realisierungen ausschließlich durch Regeln zumindest auf Schwierigkeiten stößt.

Wir wollen zwei unterschiedliche Arten der Prototypengewinnung miteinander vergleichen: einmal die Gewinnung über 'die große Zahl', d.h. Bildung über Mittelwerte der relevanten Parameter; zum anderen die Gewinnung über 'typische' Exemplare, die bei Hörtests eine besonders gute Bewertung erhielten.⁴

⁴ Die Art der Gewinnung von Prototypen hat natürlich etwas mit dem zugrundegelegten Prototypenkonzept zu tun. Zu diesem Punkt machen wir keine

Eine Gewinnung über die große Zahl setzt voraus, daß Prototypen etwas mit den häufigsten Fällen zu tun haben – im Idealfall, daß sie unimodal und normalverteilt sind. Diese Annahme ist natürlich zu einfach. Es seien anhand der Verteilung einer einzigen Variablen (eines einzigen Merkmals) drei mögliche Verteilungen diskutiert:



Der 'ideale' Fall der Verteilung einer Variablen ist eine unimodale, symmetrische – die Normalverteilung; so könnte man annehmen, daß Fragen sich immer durch einen hohen finalen Fo-Wert (Offset) auszeichnen (als 'Zielton'), der von den Sprechern angestrebt und mehr oder weniger genau getroffen wird; die Verteilung um den Mittelwert ist dann zufällig (I). Die Verteilung des Offsets bei Fragen entspricht aber eher (II), vgl. Batliner et al. (1989): Der Offset ist in den meisten Fällen hoch, kann aber auch tief sein. Weitere Verteilungen sind vorstellbar, von denen (III) einen möglichen Fall darstellt: eine trimodale Verteilung. Grundsätzlich kann man annehmen, daß sich bei einer genügend großen Stichprobe die Verteilung der Normalverteilung annähert; bei den Fragen muß man dann zwei Normalverteilungen, eine mit hohem, eine andere mit tiefem Offset annehmen, wobei die mit dem hohen Offset den häufigsten Fall repräsentiert. Man kann also über die große Zahl repräsentative prototypische Fälle ermitteln. Auf der anderen Seite ist damit nicht gesagt, daß nur solche repräsentativen Fälle prototypische Fälle sind in dem Sinn, daß sie voll akzeptabel sind. Natürlich ist es auch eine Frage der 'Klassifikationstiefe': Je nachdem wieviele Gruppen (*Clusters*) angesetzt werden, und wieviele Fälle in die Analyse eingehen, bekommt man ein mehr oder weniger differenziertes Bild der möglichen Merkmalsausprägungen bestimmter Klassen.

1.6 Prototypen und Sprecheridiosynkrasien

Bei den Fällen, die nicht dem häufigsten Fall, den wir als **Normal-Fall** bezeichnen, zuzurechnen sind, sollte unterschieden werden zwischen solchen, die bei jedem Sprecher auftreten können, und solchen, die auf Sprechereigenheiten zurückzuführen sind. Eindeutige Beispiele sind auf der einen Seite Fragen mit hohem vs. Fragen mit tiefem Offset; jeder Sprecher kann beides äußern. Auf der anderen Seite stehen Stimmqualitäten wie behauchte Stimme oder *creaky voice* (Laryngalisierung), die zumindest im Deutschen sprecherspezifisch sind. Bei den intonatorischen Merkmalen, die zur Kennzeichnung des Akzents eingesetzt werden, ist noch nicht klar, was obligatorisch, was fakultativ ist, oder anders gesagt, was kennzeichnend für die untersuchte Varietät (in unserem Fall die süddeutsche/bairische Standardsprache) und was kennzeichnend für einen Idiolekt ist. In vielen Fällen dürften sich Varietät und Idiolekt nicht klar voneinander trennen lassen, sondern nur in dem Sinn, daß einzelne Sprecher gewisse Formen häufiger oder systematischer gebrauchen als andere Sprecher.

1.7 Unterschiedliche Gewinnung der Prototypen

In einem ersten Schritt sollen die für die Fokussierung **relevanten intonatorischen Parameter** und ihr Stellenwert ermittelt werden (Teil 3 und 4). Die Diskriminanzanalyse hat dabei den Vorteil, daß sie exakte (wenn auch nicht immer leicht zu interpretierende) Maße liefert; sie hat den praktischen Nachteil, daß eine Vielzahl von solchen Maßen zwar exakt, aber doch recht unanschaulich ist. Im zweiten Schritt wird deshalb versucht, über die häufigen, repräsentativen Fälle (gewonnen über eine **Mittelwertbildung**) **Prototypen** als **Kerntypen** darzustellen (Teil 7). Dabei arbeiten wir interpretativ anhand von anschaulichen Abbildungen, aber immer gestützt auf die exakten Maßzahlen des statistischen Verfahrens. Kontrastiv und ergänzend dazu sollen sprecherübergreifende und sprecherspezifische Merkmalausprägungen ermittelt werden (Teil 5). Dazu dient auch im dritten Schritt die **Gewinnung von Prototypen über Hörerurteile** (Teil 8) sowie der Vergleich der im zweiten (Kerntypen) und im dritten (Kerntypen und Randtypen) Schritt gewonnenen Prototypen.

2. Material

Tab. 1 zeigt die verwendeten Testsätze sowie die indizierten Satzmodi. Da wir uns in diesem Beitrag nur mit der Position des FA beschäftigen, wie sie von Versuchspersonen bestimmt wurde, gehen wir nicht weiter auf die Konstruktion

der Kontextsätze und Testsätze und auf die intendierten unterschiedlichen Fokusstrukturen wie einfacher Fokus, Fokusprojektion, und geteilter Fokus ein; (vgl. dazu Oppenrieder 1989 sowie die Korpusbeschreibung in Batliner/Oppenrieder 1989).

Tab.1: Testsätze und Satzmodi:

Matrixsatz	1. Phrase	2. Phrase	3. Phrase	
1. <i>Sie läßt</i>	<i>die Nina</i>	<i>das Leinen</i>	<i>weben</i>	?/.
2. <i>Lassen Sie</i>	<i>den Manni</i>	<i>die Bohnen</i>	<i>schneiden</i>	?/!
3. <i>Lassen wir</i>	<i>den Leo</i>	<i>die Blumen</i>	<i>düngen</i>	?/!

Fragesätze

1. Assertiver Fragesatz
2. V-1-Fragesatz
3. V-1-Fragesatz

vs.
vs.
vs.

Nicht-Fragesätze

- Aussagesatz
Imperativsatz
Adhortativsatz

2.1 Produktionsexperimente

Sprecher waren StudentInnen der Germanistischen Linguistik in den Anfangssemestern ohne eine genaue Kenntnis des zugrundegelegten Satzmodussystems; sie wurden für ihre Teilnahme bezahlt. Sechs Sprecher (Sp1, Sp2 und Sp3 weiblich, Sp4, Sp5 und Sp6 männlich) produzierten die Testsätze im schallarmen Raum des Instituts für Phonetik in München jeweils zwei- bis viermal; insgesamt erhielten wir 360 Realisationen, pro Sprecher also ca. 60 (vgl. Batliner/Oppenrieder 1989). Von jeder Äußerung wurde ein Mingogramm mit Fo-Verlauf, Intensitätsverlauf und Zeitsignal erstellt. Da später alle Äußerungen zusätzlich in einer digitalisierten Kopie vorlagen, konnten die daran ermittelten Intensitätswerte sowie die Gesamtsteigung der Äußerungen (*all point regression line*) mit in die endgültige Auswertung eingehen. An den Mingogrammen wurden neben den üblichen Parameterwerten (vgl. Batliner/Oppenrieder 1989) zusätzliche speziell für dieses Korpus extrahiert. Eine Beschreibung der extrahierten Parameterwerte findet sich in Batliner/Oppenrieder (1989); vgl. auch Tab.2. Hier möchten wir deshalb nur auf zwei Punkte eingehen:

1. Als zu beschreibende Einheiten wurden die Phrasen 1 bis 3, vgl. Tab.1, festgesetzt, nicht die Nukleussilben innerhalb dieser Phrasen. Damit ließ sich das Segmentierungsproblem z.B. beim Vokal-Nasal-Vokal-Übergang bei *die Nina* oder *das Leinen* umgehen. Dieses Vorgehen schien gerechtfertigt, da die segmentale Struktur der drei Testsätze sehr ähnlich ist. Die 1. Phrase wird im weiteren nicht mehr erwähnt; sie wurde zwar von

einigen Hörern ab und an auch als akzentuiert bezeichnet, aber doch so selten, daß eine statistische Analyse darüber wenig sinnvoll wäre.

2. Beim Messen wurde zwar auf phonetisches, nicht aber auf phonologisches Wissen rekurriert; mikroprosodische Kontexteffekte (starker Fo-Abfall nach stimmlosem Plosiv etc.) wurden – als für die Perzeption irrelevant – nicht mitgemessen. Es wurden aber nicht phonologische Sequenzen hypostasiert (etwa 'Hochton, gefolgt von Tieftton') und dann die phonetische auf diese phonologische Struktur abgebildet, sondern es wurde quasi-automatisch gemessen. (Man kann sich einen automatischen Algorithmus vorstellen, der starke Fo-Veränderungen nach stimmlosen Konsonanten unberücksichtigt läßt. Eine automatische Extraktion phonologischer Sequenzen scheint uns nicht möglich zu sein – zumindest kennen wir keine.) Im Zweifelsfall entschieden wir uns dafür, Bewegungen zu berücksichtigen, auch wenn unklar ist, ob sie perzeptiv auffällig sind. So zeigt sich oft bei der unbetonten Endsilbe der Verben und grundsätzlich steigendem Fo-Verlauf ein mehr oder weniger ausgeprägter Fo-Abfall am Ende. Da kein Kriterium gefunden werden konnte, wie man einen perzeptiv irrelevanten Fo-Abfall (*relaxation gesture*) von einem deutlich intendierten, perzeptiv auffälligen trennen kann, wurde er immer mitberücksichtigt (s. auch Teil 7).

Von den gemessenen Variablen sind in unserem Zusammenhang die folgenden relevant:

Gesamtäußerung: Die Höhe des Grundfrequenzwertes am Äußerungsende (Fo-Offset), gemessen in Hz; die Fo-Steigung als Regressionskoeffizient für alle ermittelten Fo-Werte; die Dauer in Centisekunden.

2. und 3. Phrase: Fo-Maximum und Fo-Minimum in Hz ⁵; die Positionen von Phrasenanfang, Phrasenende (und damit die Phrasendauer), Fo-Maximum und Fo-Minimum auf der Zeitachse in Centisekunden; Intensitätsmaximum und Intensitätsmittelwert. (Da nicht absolute Intensitätswerte, sondern die Energieunterschiede zwischen verschiedenen Bereichen des Sprachsignals interessierten, wurden die berechneten Millibelwerte linear in der Weise transformiert, daß sie als 16 Bit-Werte speichergünstig darstellbar waren; zur genauen Berechnung vgl. Nöth 1989. Die Werte lagen für drei Bänder vor: 0-5000 Hz, 100-3400 Hz und 300-2300 Hz.)

2.2 Hörtests

Es wurden vier Hörtests durchgeführt, an denen jeweils durchschnittlich 12 Hörer teilnahmen. Versuchspersonen waren StudentInnen der Phonetik oder Germanistischen Linguistik, die für ihre Teilnahme bezahlt wurden. In einem **Natür-**

⁵ Die Entscheidung für die Messung prominenter Punkte (Maxima und Minima) ist primär auf praktische Gründe zurückzuführen: Sie können relativ genau bestimmt werden, wogegen Fo-Verläufe und damit zusammenhängende Größen wie die Steigung Schwierigkeiten bereiten. Zur Frage der perzeptiven Relevanz vgl. auch Batliner (1989b).

lichkeitstest wurden die Äußerungen im Kontext (vgl. Batliner/Oppenrieder 1989) vorgegeben; die Hörer mußten auf einer Ratingskala von 1 bis 5 (1 für 'paßt sehr gut zusammen' und 5 für 'paßt überhaupt nicht zusammen') angeben, wie gut der Testsatz in den Kontext paßt. Als 'Natürlichkeitsmaß' erhalten wir den Mittelwert aus diesen Beurteilungen. In einem **Kategorisierungstest** wurden die Äußerungen ohne Kontext dargeboten; die Versuchspersonen mußten den Äußerungstyp (Frage, Aussage, Exklamativ, Imperativ oder Wunsch) bestimmen. Wenn etwa bei einem Fragesatz von 14 Hörern 12 auf Frage und 2 auf Aussage entschieden, so ist die Trefferquote für diese Äußerung $12/(12+2)$, also 0.86; im einzelnen vgl. Oppenrieder (1988a).

Im 1. **Akzenttest** wurden die Äußerungen wie beim Kategorisierungstest zusammen mit anderen Äußerungen (vgl. Batliner/Oppenrieder 1989) ohne Kontext dargeboten. Die Hörer mußten sich für eine Silbe als die am stärksten betonte entscheiden. Da in einigen wenigen Fällen die Hörer den Hauptakzent auf dem Matrixsatz, aber nicht, wie intendiert, auf der 2. oder 3. Phrase hörten, wurde zusätzlich ein 2. **Akzenttest** durchgeführt, der nur die uns interessierenden 360 Äußerungen enthielt. Auf den 1. Akzenttest werden wir nur in Teil 9.2 bei der Fehleranalyse eingehen. Der 2. Akzenttest wurde hauptsächlich in Hinblick auf andere als hier interessierende Fragestellungen durchgeführt (phonetisches Korrelat der Fokusprojektion etc; vgl. Oppenrieder 1989). In unserem Zusammenhang lieferte er die Kenngröße für die Stärke des Akzents auf der 2. und der 3. Phrase und damit für die Zuweisung des FA.

Die Versuchspersonen durften den Matrixsatz nicht als akzentuiert bewerten und mußten für die Phrasen 1 bis 3 zwei Urteile vergeben: Hauptakzent und zweitstärkster (Neben-)Akzent. Wenn also bei einer Äußerung von 12 Hörern einer den Hauptakzent auf der 1. Phrase, acht auf der 2. und drei auf der 3. hörten, so wurde daraus für die uns interessierende Beziehung der 2. zur 3. Phrase ein Kennwert gebildet; in unserem Beispiel wäre er: $(8-3)/(1+8+3)$. Die Zahl der Akzentzuweisungen auf die 3. Phrase wird also von der Zahl der Zuweisungen auf die 2. Phrase abgezogen, und das Ergebnis durch die Anzahl der Versuchspersonen dividiert. Der Wert liegt zwischen +1.0 und -1.0, er ist positiv, wenn der Hauptakzent auf der zweiten, und negativ, wenn er auf der dritten Phrase gehört wurde. In acht Fällen wurden 2. und 3. Phrase als gleich stark akzentuiert bewertet, der Kennwert war also '0.0'. Diese Fälle werden in der Auswertung nicht berücksichtigt.⁶

⁶ Fälle mit fehlender Klassenzugehörigkeit können von der Diskriminanzanalyse in der Lernphase nicht berücksichtigt werden, sie werden aber klassifiziert. Wie erwartet, wurden die acht Fälle je nach Konstellation und Prädiktorvariablen mal der einen und mal der anderen Gruppe zugeschlagen, mit

3. Auswahl der Merkmale

Es herrscht bei weitem keine Einigkeit darüber, welche der intonatorischen Merkmale F_0 , Dauer und Intensität in welcher Form (Umrechnung) die relevanten für die Akzentmarkierung sind. Wenn man nun nicht wie im Perzeptionsexperiment mit manipuliertem Material nur einen Parameter variiert und die anderen konstant läßt, also etwa Hz- und Halbton- (Ht-) Unterschiede beurteilen läßt, so steigt der Aufwand durch die Kombinationsmöglichkeiten exponentiell an und ist nicht mehr zu bewältigen. Man muß sich daher i.a. auf ein heuristisches Verfahren beschränken, das auf plausiblen Annahmen beruht. Eine ausführliche Darstellung der von uns vorgenommenen Berechnungen würde über den Rahmen dieses Beitrags hinausgehen. Wir wollen uns deshalb auf einen Überblick beschränken und dabei die wichtigsten Teilergebnisse anführen.

3.1 F_0 -Werte

Wenn man Sprecher mit unterschiedlichen Stimmlagen (Frauen und Männer) analysieren möchte, so ist der Hz-Rohwert keine optimale Skalierung. (Das mag bei großen Stichproben weniger auffällig sein, ist aber sehr relevant, wenn z.B. von einem einzigen Sprecher auf andere geschlossen wird.) Eine 'gehörsadäquate' Transformation in Ht-Werte nach der Formel $Ht = 17.31 * LN(Hz)$ zur **Normierung des Stimmumfangs** verbesserte die Prädiktionsgüte nicht entscheidend (vgl. dazu auch Batliner et al. 1989). Entscheidend ist eine **Normierung der Stimmlage** durch einen Bezug zu einem Wert, der entweder für den Sprecher oder für die Äußerung 'repräsentativ' ist. Dieser Wert kann der **sprecherspezifische Basiswert** sein (der tiefste vom jeweiligen Sprecher erreichbare Wert) oder der **Äußerungsmittelwert** (den man sogar grob aus wenigen F_0 -Werten berechnen kann, vgl. Nöth et al. 1987, Batliner 1988a).

Die Unterschiede in der Klassifikationsgüte zwischen normierten Hz- und Ht-Werten bzw. zwischen Normierung zum Basis- oder zum Mittelwert waren nicht auffällig. Eine Entscheidung zwischen diesen Transformationen kann also fürs erste aus praktischen Gesichtspunkten erfolgen. Ein automatisches Verfahren wird dann, wenn es sprecherunabhängig arbeiten soll, natürlich den Äußerungsmittelwert nehmen. Wir entschieden uns dafür, im weiteren mit der Ht-Transformation zum Basiswert zu arbeiten, da sie im Schnitt etwas bessere Ergebnisse lieferte bzw. bei anderen, aber ähnlichen Konstellationen geliefert hat,

einer Tendenz zur Gleichverteilung. Es handelt sich also um Grenzfälle; vgl. dazu die Fehleranalyse in Teil 9.2.

vgl. Batliner (1988). Diese Transformation ist in der phonetischen Literatur auch üblich und deshalb eher vergleichbar.

3.2 Deklination

Es ist bekannt, daß die zeitliche Positionierung der akzentuierten Phrase in der Äußerung eine Rolle spielt: Am Ende einer Äußerung genügt eine geringere Ausprägung der Parameterwerte als zu Beginn, um den Fokus zu markieren (vgl. auch Batliner 1989a). Für die Berechnung der Deklinationslinie als Bezugsgröße gibt es kein anerkanntes Standardverfahren (vgl. dazu die Kontroverse zwischen Lieberman et al. 1985, 't Hart 1986, Lieberman 1986). Problematisch ist weiter, daß Untersuchungen zur Deklination normalerweise an relativ konsistenten Korpora mit Aussagesätzen durchgeführt wurden. Es ist unklar, welche Bezugswerte man für einen Fragesatz annehmen soll: den aktuellen Deklinationsverlauf der Äußerung oder einen 'unmarkierten' Referenzverlauf? Man kann natürlich versuchen, die Fo-Werte zu einer Deklinationslinie in Beziehung zu setzen, um damit, ähnlich wie im nächsten Abschnitt bei den Verhältniswerten, die Perzeption des Hörers nachzubilden. Diese Deklinationslinie kann eine abstrakte Referenzlinie sein oder aus den jeweils konkret gegebenen Fo-Werten der Äußerung berechnet werden. Solche Werte führten bei uns zu inkonsistenten und insgesamt schlechteren Ergebnissen. Damit ist nicht nachgewiesen, daß die Deklination keine Rolle spielt - es mag einfach daran liegen, daß wir nicht die richtige Deklinationslinie als Bezugsgröße ansetzten. (Wir errechneten eine **abstrakte** ('neutrale') Deklinationslinie pro Sprecher aus den Nicht-Fragesätzen mit 'unmarkierter' Deklination bzw. eine **konkrete** pro Äußerung als eine *all point regression line*.)

Grundsätzlich ist u.E. die Deklination immer noch eher ein Untersuchungsgegenstand an sich als eine sinnvolle, normierte Bezugsgröße. Die Auswirkung der Deklination - daß die intonatorischen Merkmale späterer Akzente weniger stark ausgeprägt sein müssen als früherer, um den Fokus zu markieren - geht sowieso in das Klassifikationsverfahren ein, da die Diskriminanzanalyse ja genau diese Verhältnisse in der Analysephase bei der Lernstichprobe vorfindet und sich 'merkt': Das Verfahren betrachtet keine absoluten Größen, sondern Verteilungen (vgl. weiter unten Fig.2).

Aus all diesen Gründen wurden die Fo-Werte nicht zu einer Deklinationslinie in Beziehung gesetzt.

3.3 Dauerwerte

Die Positionen von Fo-Maximum und Fo-Minimum wurden nicht transformiert. Bei der Dauer der 2. und 3. Phrase brachte die folgende Divisions- und Multiplikations-Umrechnung immer bessere Ergebnisse als die Dauer-Rohwerte oder eine additive Umrechnung:

Dauer der Phrase relativ zur mittleren Dauer der Phrase pro Sprecher und zur mittleren Silbendauer der Gesamäußerung; Formel:

$$\frac{\text{Dauer der Phrase}}{\text{mittl. Phrasendauer}} * \frac{\text{Dauer der Phrase}}{\text{Äußerungsdauer / Silbenzahl}}$$

Dieses Ergebnis war zu erwarten, da Länge und Dauer als Maß nur relativ zur unmittelbaren Umgebung bzw. zur globalen Umgebung (Sprechtempo) sinnvoll sind. Die ebenfalls relevante intrinsische Dauer besonders der Vokale wurde nicht berücksichtigt; das schien bei unserem konsistenten Material vertretbar.

3.4 Intensitätswerte

Die Maxima waren grundsätzlich bessere Prädiktoren als die Mittelwerte. Von uns vorgenommene Transformationen (zum Mittelwert wie bei den Hz-Werten, vgl. oben, oder zur intrinsischen Lautstärke anhand eines Bezugskorpus) ergaben keine systematische Verbesserung. Wir entschieden uns deshalb dafür, die Intensitätsmaxima im Bereich von 0 bis 5000 Hz für die endgültige Analyse zu übernehmen, die im Schnitt etwas bessere Ergebnisse brachten als die beiden anderen Hz-Bänder.

Man beachte, daß die Intensitätswerte für die ganzen Phrasen ermittelt wurden, nicht für die betonten Silben. Es gehen also auch die Intensitäten der Frikative und Plosive in die Berechnung der Intensität ein, nicht nur die der stimmhaften Passagen. Bestimmte Umrechnungen, die z.B. Beckman (1986) vorschlägt, konnten deshalb nicht durchgeführt werden. Die Bezugswerte zur intrinsischen Lautstärke der Vokale stammen von anderen Sprechern; man weiß, daß diese Werte von Sprecher zu Sprecher differieren können. Aus diesen Gründen spiegelt sich in der Unmöglichkeit, die Intensitätswerte durch Transformationen aufzuwerten, sowie in der relativ geringen Relevanz der Intensitätswerte für die Klassifikation (vgl. Teil 4) zwar die allgemeine Ansicht über die geringe Bedeutung der Intensität bei der Markierung des Akzents wider, ein Nachweis dieser geringen Bedeutung ist damit aber nicht gegeben. (Vgl. auch die Diskus-

sion zur unterschiedlichen Rolle der Intensität bei den einzelnen Sprechern in Teil 5.)

3.5 Verhältniswerte

Der Hörer beurteilt natürlich bei der Akzentzuweisung die Parameterwerte nicht absolut, sondern jeweils in Relation zur Umgebung. Wir können also auch in unserer Analyse nicht die absoluten Werte der 2. und der 3. Phrase zugrundelegen, sondern z.B. nur die Differenz der Werte. Solche Verhältniswerte haben den rechentechnischen Vorteil, daß sie die Zahl der Prädiktorvariablen um die Hälfte reduzieren; gleichzeitig erhalten Extremwerte ein geringeres Gewicht, die Verteilung wird regulärer und nähert sich schneller der Verteilung der Population an. Auf der anderen Seite geht Information verloren. Grundsätzlich bleibt zu untersuchen, ob und in welcher Form solche Verhältniswerte psychophysisch adäquater sind als die Rohwerte.

Taylor/Wales (1987) haben für das australische Englisch vergleichbare Untersuchungen durchgeführt. Sie haben allerdings nur mit Verhältniswerten gearbeitet, ohne auf die Frage einzugehen, ob Rohwerte nicht doch genauso gut oder besser geeignet sind. Aus den folgenden Gründen ist ihre Untersuchung nicht direkt mit der unseren vergleichbar: Ihr Material stammt von nur einer Sprecherin; das Problem der Normierung der unterschiedlichen Sprechlagen ergab sich also bei ihnen nicht. Sie gaben die Akzentposition explizit vor. Ihr Material variiert etwas mehr als unseres. Bei der Akzentbewertung gaben sie eine Rating-skala von 1 bis 5 vor, keine einfache 'betont-unbetont'-Entscheidung. Die von ihnen extrahierten Parameter sind mit unseren zwar nicht immer identisch, aber durchaus vergleichbar; Fo-Werte wurden bei ihnen nicht transformiert. Bei der Bildung der Verhältniswerte setzten sie die akzentuierte mit der unakzentuierten Phrase in Beziehung; drei Werte wurden gebildet (Taylor/Wales 1987:239):

"Division ratio = a/u .

Subtraction ratio = $a-u$.

Michaelson Contrast ratio = $(a-u)/(a+u)$.

(a = accented, u = unaccented)"

Die Autoren erzielen mit dem Kontrast-Verhältnis (*Contrast-ratio*) bei den Regressionsanalysen deutlich bessere Ergebnisse als mit den beiden anderen: Durchschnittlich ist R^2 als das Maß der 'erklärten Varianz' bei Subtraktion und Division .29 bzw. .15, beim Kontrast-Verhältnis .85. Sie folgern deshalb:

"Although contrast ratios do not appear to have been used in prosody perception, [our results show] contrast-based features to be consistently

superior in predicting accent rating than any of the other types of combinations of ratios investigated." (Taylor/Wales 1987:243).

Wir haben für die von uns extrahierten Rohwerte und die daraus ermittelten transformierten Werte Regressionsanalysen analog zu Taylor/Wales (1987) sowie Diskriminanzanalysen durchgeführt. Allerdings wurde dabei nicht die akzentuierte Phrase zur unakzentuierten in Bezug gesetzt, sondern immer die zweite zur dritten. Dadurch ändern sich die Verteilungen der Werte, aber natürlich nicht die grundsätzliche Beziehung. Das kann also nicht erklären, warum wir nie einen so ausgeprägten Unterschied zugunsten des Kontrast-Verhältnisses gefunden haben. Beim Vergleich der Verhältniswerte mit den Rohwerten waren die Verhältniswerte bei kleinen Lernstichproben etwas besser als die Rohwerte, bei großen Lernstichproben war es umgekehrt; dafür dürften die o.a. rechentechnischen Unterschiede verantwortlich sein.

Man muß bedauern, daß Taylor/Wales (1987) die Rohwerte überhaupt nicht berücksichtigt haben. Bei einem Vergleich der Verhältniswerte untereinander ergaben sich für unsere Daten keine so großen Differenzen, die Subtraktion war im Schnitt etwas besser als das Kontrast-Verhältnis, und beide besser als die Division. Dies gilt für alle Werte, untransformierte und transformierte, also auch für die Verhältniswerte der Hz-Ausgangswerte. Es kann daher nicht daran liegen, daß durch Transformation und nachfolgende Verhältnisbildung eine 'Übertransformation' stattfand.

Dieses Ergebnis bedeutet für unser weiteres Vorgehen, daß wir keine Verhältniswerte, sondern die Werte der 2. und der 3. Phrase als Prädiktoren annehmen. Für den großen Unterschied zwischen den Ergebnissen von Taylor/Wales (1987) und unseren können wir keine plausible Erklärung finden. Natürlich sind das australische Englisch und das Deutsche sowie das experimentelle Design nicht direkt miteinander vergleichbar; es ist aber unwahrscheinlich, daß sich die Psychophysik der Sprecher des australischen Englisch so sehr von der deutscher Sprecher unterscheidet. Unsere Ergebnisse lassen also die Schlußfrage von Taylor/Wales (1987:245), "[...] whether a neural analogue can be found for contrast-based acoustic-features detectors" in einem etwas dunkleren Licht erscheinen.

4. Klassifikationsphase

Tab.2 zeigt alle bei der endgültigen Analyse verwendeten Variablen, ihre Kurzbezeichnungen, sowie die an ihnen vorgenommenen Transformationen. In der ersten Spalte stehen die Prädiktorvariablen, die Werte aus den Hörtests sowie die zu prädizierenden Klassen. In der zweiten Spalte folgen die im weiteren

verwendeten Kurzbezeichnungen, in der dritten ist angegeben, ob die Variable für die Moduszuweisung (M) und/oder die FA-Zuweisung (F) verwendet wurde. In der letzten Spalte ist die Art der Berechnung und der Transformation angegeben.

Tab.2: Variablen der Klassifikationsphase

VARIABLE	KURZB.	M/F	BERECHNUNGEN/TRANSFORMATIONEN
Prädiktorvariablen:			
Steigung	<i>Stg</i>	M	Ht-Werte (17.31*LN(Hz))
Offset	<i>Off</i>	M	Ht-Werte, transformiert zum sprecherspezifischen Basiswert
Fo-Maximum, 2. und 3. Phrase	<i>Max2</i> <i>Max3</i>	M/F M/F	Ht-Werte, transformiert zum sprecher- spezifischen Basiswert in Halbtönen
Fo-Minimum, 2. und 3. Phrase	<i>Min2</i> <i>Min3</i>	M/F M/F	
Rel. Position von Fo-Max./Min. (2.u.3.Phrase)	<i>Pos2</i> <i>Pos3</i>	M/F	Differenz der Werte von Fo-Maximum und Fo-Minimum auf der Zeitachse; positiver Wert, wenn Maximum vor Minimum, sonst negativer Wert.
Dauer der 2. und 3. Phrase	<i>Dau2</i>	F	Dauer/mittl. Phrasendauer * Dauer / (Äußerungsdauer/Silbenzahl)
	<i>Dau3</i>	F	
Intensität (2. und 3. Phrase)	<i>Int2</i>	F	relative Millibelwerte im Bereich 0-5000 Hz
	<i>Int3</i>	F	
Hörerurteile:			
Moduszuweisung	MOD		Prozent Fragekategorisierung, errech- net aus den Kategorisierungstest
FA-Zuweisung	FOK		errechnet aus dem 2. Akzenttest (Werte zwischen +1.0 und -1.0)
Zu prädizierende Klassen:			
Satzmodus	MODUS		aus MOD errechnete Dichotomisierung: FRAGE/NICHT-FRAGE
Satzfokus	FOKUS		aus FOK errechnete Dichotomisierung: 2. / 3. Phrase fokussiert

Off und *Stg* werden nur für die Moduszuweisung, *Dau2*, *Dau3*, *Int2* und *Int3* nur für die FA-Zuweisung analysiert; alle anderen Variablen lagen sowohl der Modus- als auch der FA-Zuweisung zugrunde. Die Dauer- und die Intensitäts-

werte ergaben bei MODUS eine sehr schlechte Prädiktion. Bei FOKUS ist es intuitiv nicht einsehbar, was z.B. ein hoher Offset mit der FA-Zuweisung zu tun haben soll – es sei denn, er liegt in der Nähe des Fo-Maximums auf dieser Silbe oder ist damit identisch. Eine Reduzierung der Prädiktorvariablen ist auch aus rechentechnischen Gründen erwünscht.

Tab.3 zeigt die Korrelation der Prädiktorvariablen für das gesamte Korpus, Tab.4 getrennt nach Fragen (untere Dreiecksmatrix) und Nicht-Fragen (obere Dreiecksmatrix). Zur einfacheren Darstellung sind die Korrelationskoeffizienten mit 100 multipliziert. Wie erwartet, sind die Werte z.T. hoch miteinander korreliert. Das kann triviale Gründe haben, etwa wenn *Off* mit *Max3* korreliert: Z.T. sind die Werte identisch, zum Teil interdependent. Dauer- und Intensitätswerte sind untereinander sowie mit den Ht-Werten weniger stark korreliert als die Ht-Werte untereinander. Aus Tab.4 ist ersichtlich, daß die Korrelationen bei Fragen und bei Nicht-Fragen doch sehr unterschiedlich ausfallen – ein Hinweis darauf, daß die Parameter bei Fragen und bei Nicht-Fragen unterschiedlich eingesetzt werden. Wir werden deshalb im folgenden die Prädiktion von FOKUS auf drei verschiedene Art und Weisen berechnen:

- | | |
|---------------------------------------|-----------|
| (a) Fragen und Nicht-Fragen zusammen: | FOKUS |
| (b) nur Fragen: | FOKUS-F |
| (c) nur Nicht-Fragen: | FOKUS-NF. |

Tab.3: Korrelation (*100) der Prädiktorvariablen untereinander.

	<i>Off</i>	<i>Stg</i>	<i>Max2</i>	<i>Max3</i>	<i>Min2</i>	<i>Min3</i>	<i>Pos2</i>	<i>Pos3</i>	<i>Dau2</i>	<i>Dau3</i>	<i>Int2</i>	<i>Int3</i>
<i>Off</i>	—											
<i>Stg</i>	68	—										
<i>Max2</i>	39	44	—									
<i>Max3</i>	91	75	42	—								
<i>Min2</i>	34	37	30	40	—							
<i>Min3</i>	70	80	41	67	38	—						
<i>Pos2</i>	-44	-66	-24	-50	-30	-73	—					
<i>Pos3</i>	-74	-44	-15	-58	-14	-43	25	—				
<i>Dau2</i>	-11	05	10	-13	-25	14	-22	09	—			
<i>Dau3</i>	-17	-21	-06	-03	06	-37	29	15	-38	—		
<i>Int2</i>	09	-04	09	06	-07	03	03	08	05	-03	—	
<i>Int3</i>	34	24	15	35	17	26	-14	-08	-11	04	63	—

Tab.4: Korrelation (*100) der Prädiktorvariablen untereinander. Fragen (F): untere Dreiecksmatrix; Nicht-Fragen (NF): obere Dreiecksmatrix.

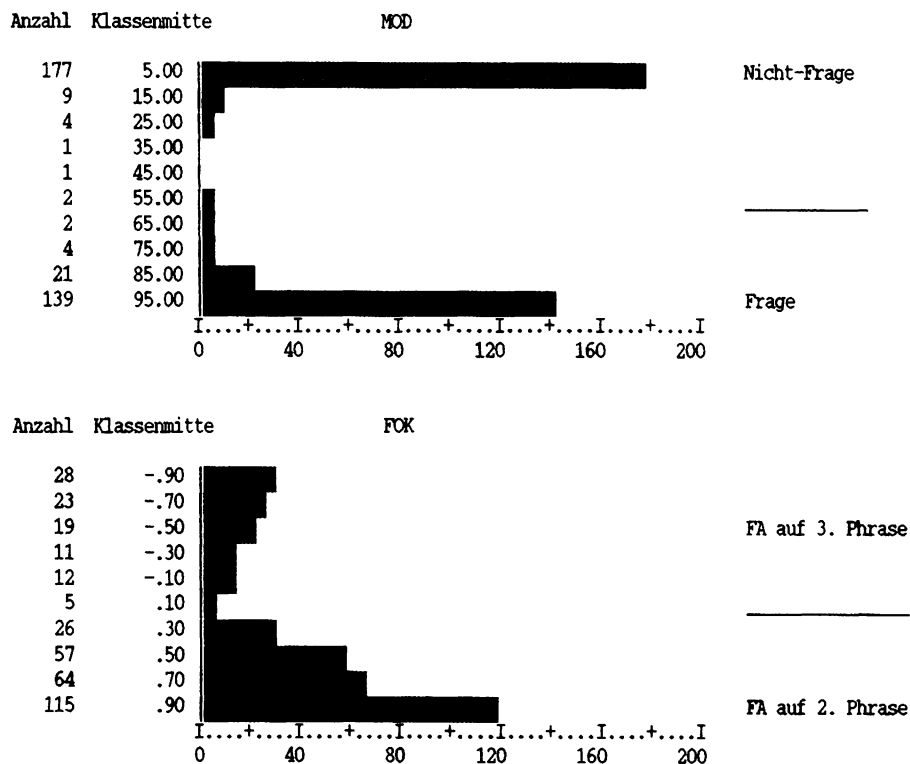
F\NF	Off	Stg	Max2	Max3	Min2	Min3	Pos2	Pos3	Dau2	Dau3	Int2	Int3
Off	—	20	37	22	21	69	-15	-13	09	17	-03	-17
Stg	11	—	33	46	22	24	-19	-06	-05	24	-08	07
Max2	-06	09	—	26	31	32	02	13	10	06	10	06
Max3	89	23	-05	—	39	34	-22	23	-16	56	-07	13
Min2	17	25	09	19	—	31	-26	10	-14	23	-14	07
Min3	29	73	15	25	26	—	-18	16	03	24	18	09
Pos2	06	-62	-04	-03	-12	-71	—	08	-17	-06	-07	-07
Pos3	-84	00	13	-65	-01	-17	-17	—	-00	11	37	37
Dau2	-11	34	20	-04	-32	39	-42	10	—	-42	07	-17
Dau3	-04	-36	-00	-00	02	-57	44	-05	-41	—	-04	21
Int2	22	-12	08	19	-02	-02	14	14	03	00	—	65
Int3	45	11	04	45	15	17	03	-24	-01	-05	63	—

Fig.1 zeigt die Verteilung der Variablen MOD und FOK, also der Hörerurteile bei der Modus- und FA-Zuweisung. Die Hörer waren sich offensichtlich bei der Moduszuweisung viel sicherer als bei der FA-Zuweisung. Man kann darüber spekulieren, ob die Unterscheidung Frage/Nicht-Frage kommunikativ wichtiger und damit vom Sprecher eindeutiger signalisiert und vom Hörer eindeutiger bestimmbar ist als die FA-Zuweisung. In unserem Korpus hat sicher auch eine Rolle gespielt, daß unser Material zu einem Drittel Kontexte enthielt, die Doppelfokus indizieren; damit mag die FA-Indizierung oft weniger eindeutig ausfallen als bei einfachem Fokus. Es ist auch bekannt, daß die Akzent- bzw. FA-Zuweisung grundsätzlich keine leichte Aufgabe ist (vgl. Thorsen 1987:87ff).

Die Variablen MOD und FOK erfüllen nicht unbedingt die Voraussetzungen für eine parametrische Statistik, etwa für eine Korrelation mit Pearsons r zwischen ihnen und den Prädiktorvariablen. Es wurde deshalb Goodmans und Kruskals Gamma-Koeffizient als Assoziationsmaß berechnet; in Tab.5 sind die Werte analog zu Tab.3 und 4 mit 100 multipliziert.⁷ Es zeigt sich, daß MOD stärker mit den Prädiktorvariablen assoziiert ist als FOK, daher auch möglicherweise stärker bzw. eindeutiger die Gruppenzuweisung indiziert.

⁷ Die Berechnung des Gammakoeffizienten ist ähnlich der von Kendalls τ , dessen Berechnung wiederum ähnlich der Berechnung des bekannteren nichtparametrischen Korrelationskoeffizienten r von Spearman. Die Gamma- und τ -Werte fallen i.a. kleiner aus als Spearmans r , allen liegen aber die gleichen Ordnungsrelationen zugrunde. (Vgl. im einzelnen zur unterschiedlichen Berechnung Hildebrand et al. 1977 und Norusis 1986).

Fig.1: Verteilung der Hörerurteile MOD (Moduszuweisung: Frage/Nichtfrage) und FOK (FA-Zuweisung: 2./3. Phrase betont).



Tab.5: Assoziation (*100) der Prädiktorvariablen mit den Hörerurteilen MOD (Moduszuweisung) und FOK (FA-Zuweisung) aufgeteilt nach Fragen (FOK-F) und Nicht-Fragen (FOK-NF). Goodmans und Kruskals Gamma-Koeffizient.⁸

	MOD	FOK-F	FOK-NF
<i>Off</i>	81		
<i>Stg</i>	*		
<i>Max2</i>	39	36	14
<i>Max3</i>	79	-17	-31
<i>Min2</i>	33	-04	-20
<i>Min3</i>	62	12	02
<i>Pos2</i>	-39	-19	10
<i>Pos3</i>	-54	22	-07
<i>Dau2</i>		19	33
<i>Dau3</i>		-17	-29
<i>Int2</i>		03	17
<i>Int3</i>		-04	-02

⁸ Die Assoziation mit *Stg* konnte aus rechentechnischen Gründen (mangelnder Speicherplatz) nicht ermittelt werden.

Bei den im folgenden besprochenen Analysen, deren Ergebnisse in Tab.6 bis 8 aufgelistet sind, gingen immer alle Fälle in die Lern- und in die Prüfstichprobe ein (vgl. unten die Konstellation (3) in Tab.9).⁹ Tab.6 zeigt die Ergebnisse der univariaten Analyse mit jeweils nur einer Variablen als Prädiktorvariable, der 'bivariaten' (paarweise werden Werte für die 2. und die 3. Phrase zugrundegelegt), sowie einen Vergleich mit der multivariaten Analyse. Wie man sieht, ermöglicht bei der univariaten Analyse *Max3* bei MODUS und FOKUS-NF schon eine sehr gute Prädiktion; bei FOKUS-F ist die Verbesserung von uni- bzw. bivariater zur multivariaten Analyse am größten; die multivariate ist aber immer besser. (Dieses Ergebnis ist nicht trivial, da eine Hinzunahme von irrelevanten Prädiktorvariablen oft das Ergebnis eher verschlechtert als verbessert.)

Um den Beitrag der einzelnen Prädiktorvariablen abzuschätzen, kann man entweder die Diskriminanzfunktionskoeffizienten der Variablen oder die Korrelation zwischen Variable und Diskriminanzfunktion interpretieren; das letztere ist in unserem Fall vorzuziehen, da die Variablen z.T. miteinander korrelieren (vgl. dazu im einzelnen Klecka 1980:33f).

Tab.7 zeigt deshalb die Korrelation der Prädiktorvariablen mit der Diskriminanzfunktion; die Koeffizienten sind wieder mit 100 multipliziert. Da die Richtung der Vorzeichen pro Spalte arbiträr ist, werden hier wie auch im weiteren die Vorzeichen der Werte zur Vereinfachung immer so wiedergegeben, daß der Wert dann positiv ist, wenn ein hoher und/oder positiver Wert eine Frage bzw. einen FA auf der 2. Phrase indizieren, und vice versa ein niedriger und/oder negativer Wert eine Nichtfrage bzw. einen FA auf der 3. Phrase. Mit '*' gekennzeichnete Variablen passieren den Toleranztest nicht, da sie lineare Kombinationen aus anderen Variablen sind. Sie gehen deshalb nicht in die Diskriminanzfunktion ein. Verantwortlich dafür sind Korrelationen der Variablen untereinander; es ist damit aber nicht gesagt, daß gerade diese ausgeschiedenen Variablen perzeptiv irrelevant sind (zum Toleranztest vgl. Norusis 1986:B-18).

⁹ 'Realistischer' wäre natürlich das 'leave-one-out'-Verfahren, bei dem reihum ein Sprecher beim Lernen ausgelassen und dann allein klassifiziert wird. Die Werte in Tab.6 und 7 sind also alle etwas zu optimistische Schätzungen (vgl. zu dieser Überadaptation bei Identität von Lern- und Prüfstichprobe Ruske 1988 und Teil 9). Es ist aber unwahrscheinlich, daß die Relation der Werte *zueinander* sich stark ändern würde. Wir haben deshalb in dieser Phase davon abgesehen, Lern- und Prüfstichprobe zu trennen.

Tab.6: Prozent richtige Klassifikation bei univariater Analyse, 'bivariater' (paarweiser) sowie multivariater Analyse: Frage/Nicht-Frage (MODUS), FA gesamt (FOKUS), FA bei Frage (FOKUS-F) und FA bei Nicht-Frage (FOKUS-NF).

	MODUS		FOKUS		FOKUS-F		FOKUS-NF	
	uni.	bi.	uni.	bi.	uni.	bi.	uni.	bi.
<i>Off</i>	93.0		—		—		—	
<i>Stg</i>	85.2		—		—		—	
<i>Max2</i>	72.8		59.9		79.6		53.5	
		> 93.8		> 77.6		> 80.8		> 91.7
<i>Max3</i>	93.5		59.8		62.9		88.4	
<i>Min2</i>	64.7		58.5		52.7		61.6	
		> 84.3		> 62.4		> 71.3		> 70.2
<i>Min3</i>	84.3		47.4		69.5		47.0	
<i>Pos2</i>	76.1		51.1		77.8		69.2	
		> 84.6		> 69.3		> 82.0		> 51.9
<i>Pos3</i>	77.5		51.1		62.3		48.1	
<i>Dau2</i>	—		66.1		59.9		69.6	
		>		> 73.5		> 71.3		> 81.5
<i>Dau3</i>	—		71.8		70.1		81.5	
<i>Int2</i>	—		58.2		52.1		61.5	
		>		> 65.6		> 55.7		> 70.3
<i>Int3</i>	—		54.7		50.9		53.3	
<hr/>								
multiv.		97.1		92.7		94.6		95.5
<hr/>								
Mindest- verbes- serung	3.6	3.3	26.6	15.1	15.0	12.6	7.1	3.8

Tab.7: Korrelation (*100) zwischen den Prädiktorvariablen und der Diskriminanzfunktion bei multivariater Analyse für Frage/Nicht-Frage (F/NF), FA gesamt (FOKUS), FA bei Frage (FOKUS-F) und FA bei Nicht-Frage (FOKUS-NF).

	F/NF	FOKUS	FOKUS-F	FOKUS-NF
<i>Off</i>	82	—	—	—
<i>Stg</i>	55	—	—	—
<i>Max2</i>	24	25	43	12
<i>Max3</i>	78	-28	-24	-61
<i>Min2</i>	15	-11	-02	-15*
<i>Min3</i>	46	13	28*	-00
<i>Pos2</i>	-31	-16	-43	02*
<i>Pos3</i>	-35	12	23*	-02*
<i>Dau2</i>	—	31*	25	32
<i>Dau3</i>	—	-37	-26	-41
<i>Int2</i>	—	12	03	18
<i>Int3</i>	—	-10	-06	-08

* Variablen, die das Toleranzkriterium nicht passiert haben (lineare Kombinationen aus anderen Variablen).

Die univariaten Analysen können mit solchen Perzeptionstests verglichen werden, bei denen jeweils nur ein Parameter manipuliert wird, die multivariaten mit natürlicher Sprache, bei der alle Variablen zur Modus- oder Fokusindizierung beitragen können. Genausowenig wie man direkt aus Perzeptionstests auf die Rolle des jeweiligen Parameters in der natürlichsprachlichen Kommunikation schließen kann, genauso wenig ist die multivariate Analyse einfach eine Addition aus den univariaten Analysen. Wie schon erwähnt, bedeutet die Korrelation der Variablen untereinander, daß man grundsätzlich etwas vorsichtig bei der Interpretation sein sollte. Um ein einfaches Beispiel zu nehmen: Die hohen Werte bei *Off* und *Max3* bedeuten nicht, daß beide Variablen gleichermaßen sehr wichtig sind, da sie bei Fragen oft nahe beieinanderliegen und in manchen Fällen sogar identisch sind. Wenn man nun aber Tab.5, 6 und 7 miteinander vergleicht, also die Assoziation der Prädiktorvariablen mit den Hörerurteilen, die Klassifikationsgüte bei der univariaten Analyse sowie die Korrelation der Prädiktorvariablen mit der Diskriminanzfunktion, so stimmen die Werte gut überein. Hohe Werte in der einen Tabelle entsprechen jeweils hohen in den beiden anderen und vice versa niedrige in der einen niedrigen in den beiden anderen. Um diese Übereinstimmung größenmäßig darzustellen, wurden die Werte der drei Tabellen miteinander korreliert, also eine Art 'Metakorrelationsanalyse' durchgeführt. (Dabei wurden beim Vergleich mit den Klassifikationsergebnissen der univariaten Analyse die Werte aus den anderen beiden Tabellen gleichgerichtet, d.h. negative in positive umgeformt.) Da der Korrelationskoeffizient eine zu optimistische Schätzung des Zusammenhangs wiedergibt, ist in Tab.8 das Quadrat des Korrelationskoeffizienten, die sog. 'erklärte Varianz' R^2 aufgeführt.

Tab.8: 'Erklärte Varianz' (R^2) aus den Korrelationen zwischen den Werten in Tab.5, 6 und 7.

'abhängig'	'unabhängig'	MODUS	FOKUS-F	FOKUS-N
Hörerurteile	univariate Analyse	.94	.66	.69
Hörerurteile	multiv. Analyse	.94	.92	.90
multiv. An.	univ. Analyse	.94	.92	.77

Auch wenn die Assoziationen der Hörerurteile mit den Prädiktorvariablen univariat gewonnen wurden, so stimmen die Hörerurteile doch weitaus besser mit der multivariaten als mit den univariaten Analysen überein: Die erklärte Vari-

anz liegt immer über .90. Es war zu erwarten, daß die univariaten Analysen nicht ganz so gut mit der multivariaten Analyse übereinstimmen.¹⁰

Tab.9 zeigt die Ergebnisse für unterschiedliche Konstellationen von Lern- und Prüfstichproben. Die Reihenfolge der Behandlung ergibt sich aus dem folgenden Bild:

Lern- = Prüfstichprobe		Lern- ≠ Prüfstichprobe	
ein Sprecher	(1)	(5)	fünf Sprecher
fünf Sprecher	(2)	(4)	ein Sprecher
alle Sprecher	(3)		

FOKUS ohne Trennung von Fragen und Nicht-Fragen soll hier nur kurz behandelt werden, da eine Trennung der Hauptmodi grundsätzlich bessere Ergebnisse bringt. (Natürlich könnte man auch den gegensätzlichen Gesichtspunkt betonen: Es ist erstaunlich, daß die Ergebnisse trotz der fehlenden Trennung noch recht gut sind.)

- (1) Reihum ist ein Sprecher Lern- und gleichzeitig Prüfstichprobe. Damit kann die Konsistenz der Sprecher in sich überprüft werden. Dieses Vorgehen ist dem häufigen Vorgehen in der Phonetik vergleichbar, bei dem Daten nur eines Sprechers interpretiert werden. Die Ergebnisse sind immer sehr gut und liegen zwischen 98.2 und 100%. Da wir hier Lern- und Prüfstichprobe wegen der relativ geringen Zahl von Fällen nicht trennen, ist eine Überadaptation wahrscheinlich; d.h. die Konsistenz der Sprecher in sich dürfte bei Trennung von Lern- und Prüfstichprobe ohne eine Erhöhung der Stichprobengröße nicht mehr so eindeutig sein.
- (2) Reihum werden fünf Sprecher als Lern- und Prüfstichprobe genommen. Bei MODUS ist das Ergebnis unwesentlich schlechter als in (1), bei FOKUS-F und FOKUS-NF ist es um ca. 5% schlechter.¹¹
- (3) Alle sechs Sprecher sind Lern- und zugleich Prüfstichprobe. Die Hinzunahme eines weiteren Sprechers verschlechtert offensichtlich die Prädiktion nicht, da die Unterschiede unwesentlich sind.
- (4) Reihum ist ein Sprecher Prüfstichprobe auf der Grundlage der anderen fünf Sprecher, s. (2). Dieses Vorgehen wird als 'leave one out-Verfahren' bezeichnet; damit läßt sich eine sprecherunabhängige automatische Spracherkennung simulieren. Zugleich stellt man das Verfahren vor eine Aufgabe, wie sie jeder Mensch lösen kann, wenn er einem unbekannten Sprecher zum ersten Mal begegnet. Bei FOKUS-NF ist das Ergebnis immer noch sehr gut, bei FOKUS-F bzw. MODUS ist es um 8% bzw. 5% schlechter.
- (5) Reihum sind fünf Sprecher Prüfstichprobe auf der Grundlage des jeweils sechsten Sprechers. Damit kann die Repräsentativität eines einzelnen

¹⁰ Dieses Ergebnis zeigt die eingeschränkte Relevanz von Perzeptionsexperimenten, bei denen immer nur ein Parameter variiert wird.

¹¹ Diese Konstellation wird beim 'leave-one-out'-Verfahren (Konstellation 4) automatisch mitberechnet. Sie wird mit aufgeführt, da man im Vergleich zu Konstellation (3) ein grobes Maß dafür erhält, welchen Einfluß die Hinzunahme eines weiteren Sprechers auf die Klassifikationsgüte hat.

Sprechers getestet werden. Bei FOKUS-F und FOKUS-NF verschlechtert sich das Ergebnis deutlich gegenüber (4), bei MODUS wird es dagegen besser. Wir werden in Teil 7 auf dieses prima facie unerwartete Ergebnis zurückkommen.

Tab.9: Lern- vs. Prüfstichproben für Frage/Nicht-Frage (MODUS), FA gesamt (FOKUS), FA bei Frage (FOKUS-F) und FA bei Nicht-Frage (FOKUS-NF).

Lern = Prüf		—>		Lern ≠ Prüf
MODUS	ein Sprecher	98.9	92.0	fünf Sprecher
	fünf Sprecher	97.2	91.7	ein Sprecher
	alle Sprecher	97.1		
FOKUS	ein Sprecher	97.0	77.5	fünf Sprecher
	fünf Sprecher	93.1	84.1	ein Sprecher
	alle Sprecher	92.7		
FOKUS-F	ein Sprecher	98.2	75.7	fünf Sprecher
	fünf Sprecher	94.6	86.2	ein Sprecher
	alle Sprecher	94.6		
FOKUS-NF	ein Sprecher	100.0	81.6	fünf Sprecher
	fünf Sprecher	93.4	93.8	ein Sprecher
	alle Sprecher	95.5		

5. Relevante Merkmale: sprecherübergreifend oder sprecherspezifisch?

Wenn man nur an einem Sprecher trainiert und auf die übrigen Sprecher schließt, so verschlechtert sich die Prädiktion des FA deutlich. Das läßt vermuten, daß einzelne Sprecher nicht-repräsentative Eigenheiten im Einsatz der Merkmale aufweisen. Tab.10 bis 12 zeigen deshalb die Korrelationen der Prädiktorvariablen für MODUS, FOKUS-F und FOKUS-NF mit der Diskriminanzfunktion für eine Analyse, bei der jeweils ein Sprecher Lern- und Prüfstichprobe zugleich ist, also für die Konstellation (1) aus Tab.9; die Korrelationskoeffizienten sind wieder mit 100 multipliziert. Die drei rechten Spalten zeigen den Mittelwert pro Zeile (\bar{x}), die Standardabweichung (s) sowie die Differenz zwischen dem maximalen und dem minimalen Wert pro Zeile (Spannweite), jeweils auf die ganze Zahl gerundet. Die Mittelwerte stimmen gut mit Tab.7, also den Werten für die multivariate Diskriminanzanalyse für alle Sprecher (Konstellation (3) in Tab.9) überein. Betrachtet man nun die Standardabweichungen und die Spannweiten, so finden sich manchmal recht hohe Werte ($s > 20$ und Spannweite > 50). Sucht

man unter den sechs Sprechern nach den 'Verantwortlichen', so sieht man, daß immer dann, wenn einzelne Sprecher von den anderen abweichen, es sich um Sp1 oder Sp6 handelt: bei MODUS Sp1 bei *Off* und Sp6 bei *Pos3*; bei FOKUS-F Sp1 bei *Pos3* und Sp6 bei *Max2*, *Min2*, *Min3*, *Pos2* und *Int3*; bei FOKUS-NF Sp6 bei *Pos2*. Die niedrigere Korrelation von *Off* mit der Diskriminanzfunktion bei MODUS bei Sp1 weist darauf hin, daß diese Sprecherin oft Fragen mit einem niedrigen Offset indiziert (vgl. unten Fig.5); bei Sp6 ist die Korrelation von *Pos3* mit der FA-Zuweisung auf der 2. Phrase zwar auch negativ, aber viel größer. Dieser Sprecher setzt also ein übliches Merkmal für Fokus auf der 3. Phrase (Fo-Minimum vor Fo-Maximum, also ein ausgeprägter 'Tiefton' im Ton-Sequenz-Ansatz von Pierrehumbert 1980) stärker ein als die anderen Sprecher. Die anderen Abweichungen können analog interpretiert werden. Wir warten aber damit bis zur Diskussion von Fig.2 bis 10, da die Interpretation anhand der Abbildungen eingängiger sein dürfte.

Standardabweichung und Spannweite sind bei MODUS und bei FOKUS-NF geringer als bei FOKUS-F; für FOKUS-F zeigt auch Tab.9 durchgehend eine schlechtere Klassifikationsgüte als für MODUS und FOKUS-NF. Man kann also daraus schließen, daß bei Fragen eine Überlagerung von modus- und von fokusspezifischen Parameterausprägungen die Klassifikation erschwert.

Tab.10: Korrelation (*100) der Prädiktorvariablen mit der Diskriminanzfunktion pro Sprecher bei MODUS

	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	\bar{x}	s	Spannweite
<i>Off</i>	31	88	83	86	80	75	74	22	57
<i>Stg</i>	55	29	38	23	23	42	35	13	32
<i>Max2</i>	26	02	23	06	12	36	18	13	34
<i>Max3</i>	68	74	66	26	60	56	58	17	48
<i>Min2</i>	01	08	16	09	09	10	09	05	15
<i>Min3</i>	32	45	43	17	18	29	31	12	28
<i>Pos2</i>	-21	-30	-48	-10	-09	-09	-21	16	39
<i>Pos3</i>	-03	-35	-18	-17	-16	-65	-26	22	62

Tab.11: Korrelation (*100) der Prädiktorvariablen mit der Diskriminanzfunktion pro Sprecher bei FOKUS-F

	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	\bar{x}	s	Spannweite
<i>Max2</i>	54	27	19	40	34	75	42	20	56
<i>Max3</i>	-16	-12	-07	-12	01	-33	-13	11	34
<i>Min2</i>	13	-03	-03	24	17	-32	03	20	56
<i>Min3</i>	32	28	23	66	48	-18	30	28	84
<i>Pos2</i>	-53	-41	-19	-55	-52	17	-34	28	72
<i>Pos3</i>	64	05	08	12	28	11	21	22	59
<i>Dau2</i>	08	15	11	15	18	05	12	05	13
<i>Dau3</i>	-30	-07	-07	-49	-35	14	-19	23	63
<i>Int2</i>	-19	02	03	-29	-01	10	-06	15	39
<i>Int3</i>	-06	-18	05	-02	-04	-47	-12	19	52

Tab.12: Korrelation (*100) der Prädiktorvariablen mit der Diskriminanzfunktion pro Sprecher bei FOKUS-NF

	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	\bar{x}	s	Spannweite
<i>Max2</i>	14	19	07	03	16	04	11	07	16
<i>Max3</i>	-52	-25	-44	-60	-43	-32	-43	13	35
<i>Min2</i>	07	-15	-20	-14	-10	-14	-11	09	27
<i>Min3</i>	-01	05	-12	-06	06	04	-01	07	18
<i>Pos2</i>	-09	-08	15	10	-02	56	11	24	65
<i>Pos3</i>	-02	-01	-16	-30	-04	-03	-09	12	29
<i>Dau2</i>	23	33	08	19	13	39	23	12	31
<i>Dau3</i>	-48	-19	-29	-17	-34	-34	-30	11	31
<i>Int2</i>	12	07	10	02	16	10	09	06	18
<i>Int3</i>	01	-23	-10	-09	-11	-27	-13	10	28

6. Abschließende Bemerkungen zur statistischen Analyse

Tab.7 zeigt, daß z.B. *Int2* und *Int3* bei FOKUS-F und FOKUS-NF sowie *Pos2* und *Pos3* bei FOKUS-NF praktisch nicht mit der Diskriminanzfunktion korrelieren, d.h. nur sehr wenig zur Trennung der beiden Klassen beitragen. Man könnte natürlich solche Variablen ausscheiden; das hätte rechentechnische Vorteile. Unser Ziel war aber nicht in erster Linie eine Minimierung der Prädiktorvariablen, sondern eine phonetische Analyse der sprecherübergreifenden und der sprecherspezifischen Relevanz der Variablen. Wir werden im folgenden sehen, daß auch solche Merkmale, die insgesamt wenig zur Fokusindizierung beitragen, bei einzelnen Sprechern sehr wohl relevant sein können.

Die Ergebnisse der statistischen Analyse sind exakt, aber unanschaulich; es ist zwar möglich, aus den Zahlen repräsentative Parameterausprägungen und damit Verläufe bzw. Größenkorrelationen zu rekonstruieren – das wäre aber recht mühsam. Es soll deshalb genügen, daß man aus den Zahlen die relative Wichtigkeit der einzelnen Prädiktorvariablen erschließen kann. Im weiteren wollen wir uns eine Eigenschaft unserer Daten zunutze machen, die sie vor anderen Daten auszeichnet: Eine Darstellung im zweidimensionalen Raum ist nicht nur anschaulich und möglicherweise phonetisch/kognitiv relevant, wie das etwa bei der Nonmetrischen Multidimensionalen Skalierung der Fall ist (vgl. Batliner 1989b); eine solche Darstellung ist außerdem sehr konkret und bei phonetischen Daten üblich. Wir wollen uns deshalb im nächsten Abschnitt anhand der Mittelwertsdarstellungen der Prädiktorvariablen ein Bild von den typischen Vertretern, eben den Prototypen machen. Diese Darstellungen gleichen denen von aktuellen Fo-Verläufen; zum Unterschied vgl. Teil 7. Wenn wir diese Mittelwerte interpretieren, dann machen wir zwar nicht das Gleiche, aber doch etwas sehr Ähnliches wie die Diskriminanzanalyse, die ja versucht, über Verteilungsmaße die Abstände zwischen den Gruppen zu maximieren: Der Mittelwert ist ein bedeutendes, wenn auch nicht das einzig relevante Verteilungsmaß. Wenn nun die Unterschiede in Fig.2 bis 10 ins Auge fallen, dann sprechen sie für sich; wenn sie gering sind, dann können wir uns immer mit einem Blick in die Tabellen vergewissern, wie groß bzw. wie relevant dieser Unterschied ist.

7. Gewinnung der Prototypen über die große Zahl

In Fig.2 bis 10 sind die Mittelwerte für die relevanten Parameter aufgetragen, und zwar in Fig.2 für alle Äußerungen, in Fig.3 für die 'eindeutigen' Fokussierungen, in Fig.4 für die 'uneindeutigen' (vgl. zu den Begriffen weiter unten) und in Fig.5 bis 10 für die einzelnen Sprecher. Eine genauere Aufgliederung nach Satzmodi (z.B. bei den Nicht-Fragen nach Aussagesatz, Imperativsatz und Adhortativsatz) hat sich bei unserem Material nicht als sinnvoll erwiesen, da dafür die Zahl der Fälle pro Sprecher zu gering war. Fragen und Nicht-Fragen sowie Fokus auf der 2. und Fokus auf der 3. Phrase werden getrennt dargestellt; in Klammern ist jeweils die Zahl der Fälle angegeben (n), über die gemittelt wurde.¹² Die x-Achse zeigt den Zeitbereich in Centisekunden, die y-Achse ist Ht-skaliert, wobei der sprecherspezifische Basiswert auf Null gesetzt ist.

Die Werte *Max2*, *Min2*, *Max3* und *Min3* werden als Volltonquadrate dargestellt. Anfang und Ende der Phrasen 2 und 3 sind durch Kreise gekennzeichnet. Da die relationalen Einheiten *Dau2* und *Dau3* nicht in Centisekunden angegeben werden können, sind die Mittelwerte der Positionen von Phrasenanfang und -ende angegeben; die Linie zwischen Anfang und Ende kennzeichnet die mittlere Dauer der Phrasen. Um auch die Intensitätsverhältnisse in der gleichen Abbildung wiedergeben zu können, dient die y-Achse als Hilfsachse für die Differenz zwischen *Int2* und *Int3*: Die Phrase mit der schwächeren Intensität (bis auf eine Ausnahme immer die 3. Phrase) wird bezüglich der Dauerkennzeichnung auf 1 gesetzt, die mit der höheren auf $1 + (Int2 - Int3)/100$. Auf der y-Achse läßt sich damit die linear transformierte Differenz zwischen der Intensität auf der 2. und der auf der 3. Phrase ablesen. Man beachte, daß es sich nicht um 'produzierbare' Konturen handelt! So zeigt etwa Fig.10 mit Sp6 bei FOKUS 3, FRAGEN auf der zweiten Phrase keinen abrupten Tonsprung, also kein Jodeln; man kann aber ersehen, daß die Position der Minima um die der Maxima oszilliert, daß also die Minima etwa zu gleichen Teilen vor und nach den Maxima liegen. Aus diesem Grund sind *Max2* und *Min2* bzw. *Max3* und *Min3* nicht durch Linien miteinander verbunden.

¹² Beim Aufsummieren über die Sprecher oder über die Konstellationen können sich Unstimmigkeiten ergeben, die dadurch begründet sind, daß für einige Äußerungen nicht alle Variablen ermittelt werden konnten.

Fig.2: Alle Äußerungen

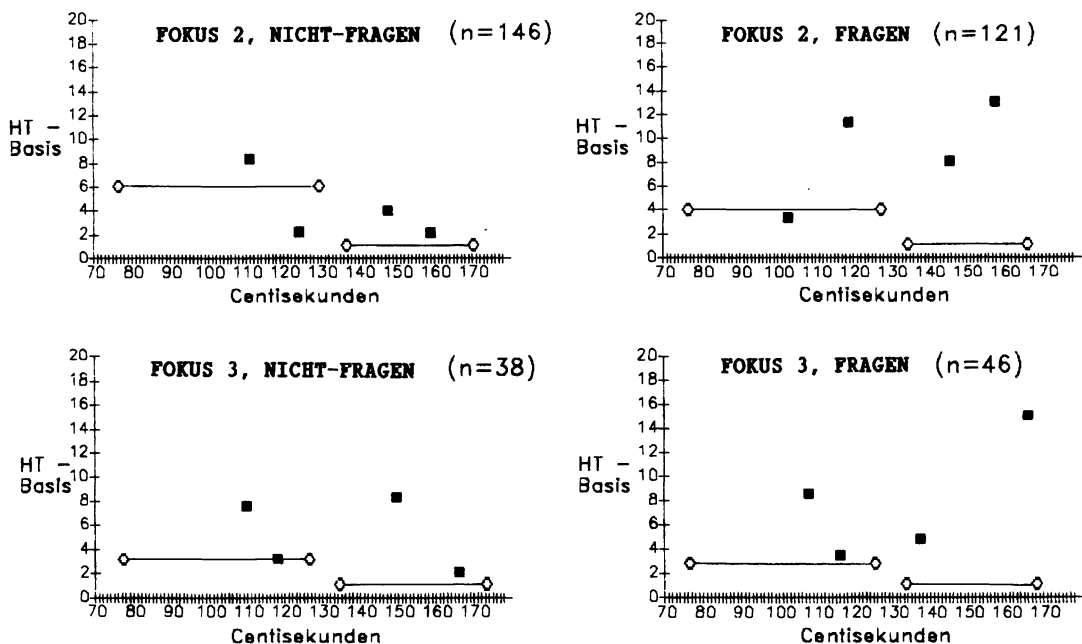


Fig.3: Eindeutige Fokussierungen

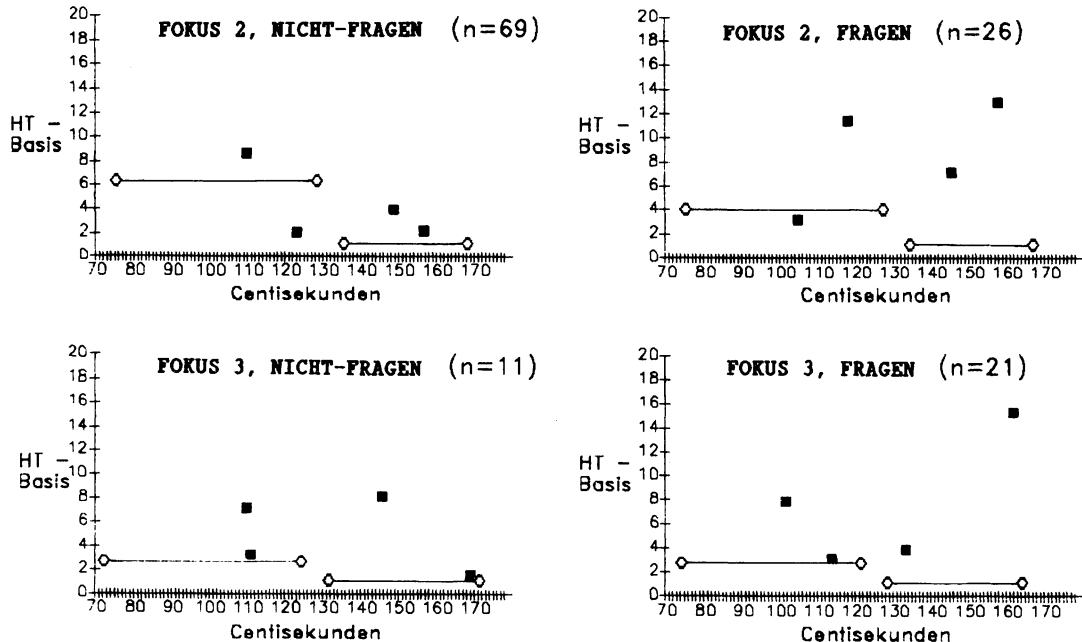
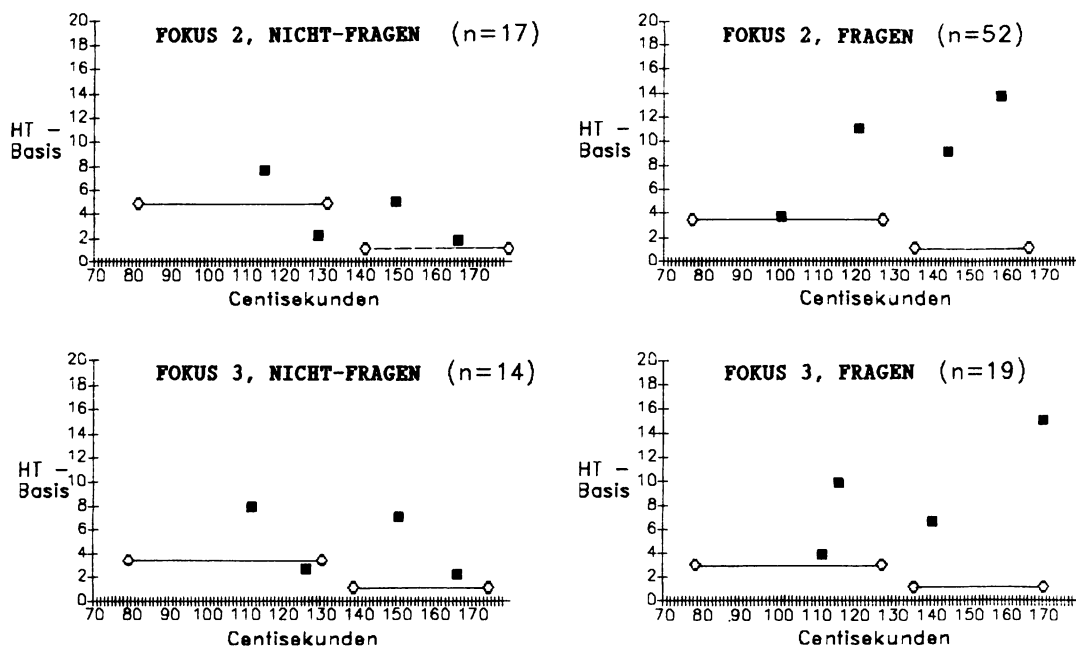


Fig.4: Uneindeutige Fokussierungen



Interpretiert werden können die Unterschiede zwischen den Werten der 2. und der 3. Phrase innerhalb einer Konstellation sowie die Unterschiede in den Spalten und in den Zeilen der Figuren zwischen jeweils den Werten der 2. oder der 3. Phrase. Natürlich hängt bei einer solchen Darstellung der sichtbare Unterschied auch von der Skalierung ab; wie schon erwähnt, kann man aber immer auf die statistischen Werte in den Tabellen zurückgreifen.

Wir nehmen an, daß man einen besonders ausgezeichneten intonatorischen Prototyp, nämlich den Normalfall (Kerntyp) über die große Zahl erhalten kann. Wir müssen dabei natürlich unterstellen, daß unsere Stichprobe eine gewisse Repräsentativität besitzt – nicht in dem Sinn, daß alle möglichen Kombinationen und damit auch Randtypen enthalten sind, aber in dem Sinn, daß eine andere Stichprobe die gleichen Normalfälle zeigen würde. Beim Betrachten von Fig.2 ist diese Annahme intuitiv einsehbar: Es zeigt sich, was man erwarten würde und was wir im folgenden erläutern.

Nicht-Fragen zeigen auf akzentuierter und unakzentuierter Phrase eine fallende Kontur ('Hochton' mit folgendem 'Tiefton' im Ton-Sequenz-Ansatz). Fokussierung wird durch einen höheren Tonumfang (*Range*) markiert, der hauptsächlich über die Maxima gesteuert wird. Die Prominenz der Maxima muß auch in Beziehung gesetzt werden zu ihrer Position auf der Zeitachse, da *Max3* bei FOKUS 3 nur unwesentlich höher ist als *Max2* im Gegensatz zum ausgeprägteren Unterschied von *Max2* und *Max3* bei FOKUS 2 (Einfluß der Deklinationslinie). Dauer und Intensitätsunterschiede sind nicht sehr ausgeprägt, sie markieren aber doch systematisch die fokussierte Phrase durch eine größere Ausprägung relativ zur nicht fokussierten Phrase.

Fragen zeigen auf der fokussierten Phrase eine steigende Kontur ('Tiefton' mit folgendem 'Hochton' im Ton-Sequenz-Ansatz), ebenfalls auf der nicht-fokussierten bei FOKUS 2 (bedingt durch den hohen Offset); bei FOKUS 3 hat die nicht-fokussierte zweite Phrase eine fallende Kontur. Dauer und Intensität markieren zwar auch die fokussierte Phrase durch eine größere Ausprägung, der Unterschied ist aber weniger augenfällig als bei den Nicht-Fragen. (Exakte Kennwerte dafür finden sich in Tab.9 bei der Korrelation der Prädiktorvariablen mit der Diskriminanzfunktion.)

Die Aufteilung in eindeutige und uneindeutige Fälle (Fig.3 und 4) wurde anhand von FOK, also anhand der Werte aus dem 2. Akzenttest vorgenommen. Es sollten dabei in beiden Gruppen in etwa gleich viele Fälle sein und nicht zu wenige, um Extremwerten kein zu großes Gewicht zu geben. Die Variable FOK hat einen Bereich zwischen -1.0 (alle Hörer setzten den Akzent auf die 3. Phrase) und +1.0 (alle Hörer setzten den Akzent auf die 2. Phrase). Wir definieren als 'eindeutig' die Fälle mit dem Wert $>+.8$ oder $<-.8$ ($n=127$), und als 'uneindeutig' die Fälle mit Werten zwischen .5 und 0.0 bzw. $-.5$ und 0.0 ($n=102$). Es zeigt sich, daß sowohl bei den eindeutigen wie bei den uneindeutigen Fällen die grundsätzlichen Relationen die gleichen sind; das kann als Argument für eine Prototypenbildung über die große Zahl genommen werden. Die Ausprägungen sind allerdings bei den uneindeutigen Fällen geringer.

Fig.5: Sprecher 1 (Sp1)

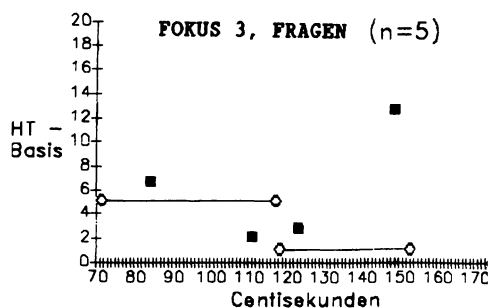
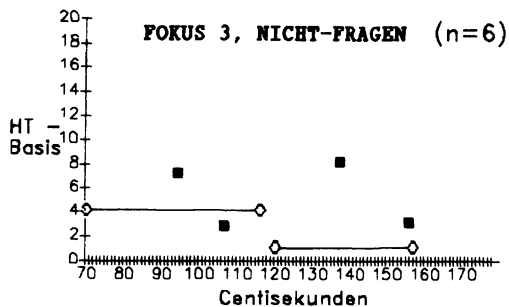
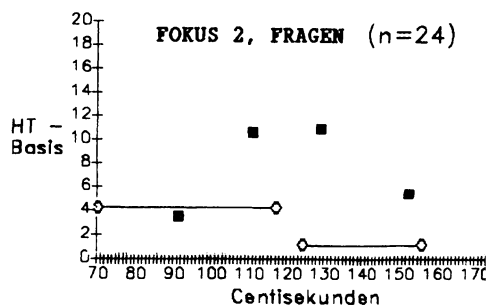
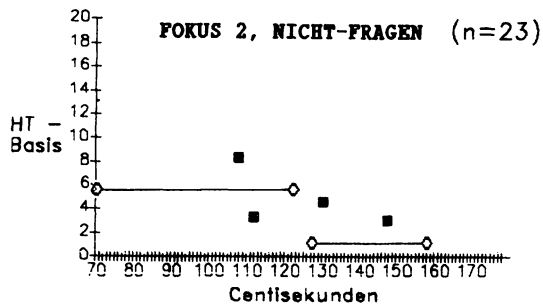


Fig.6: Sprecher 2 (Sp2)

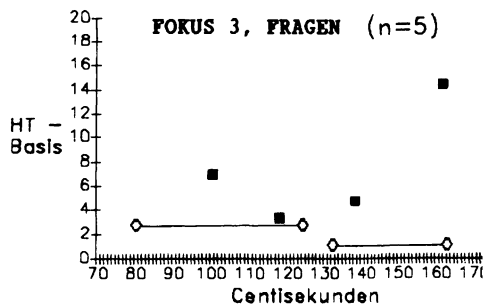
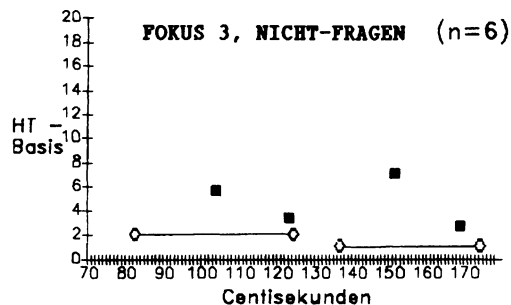
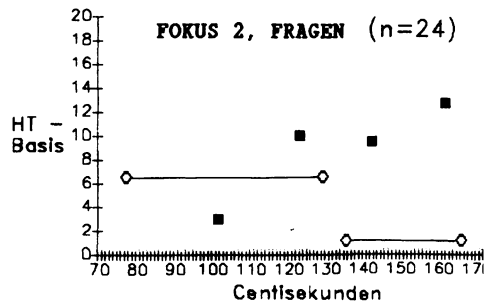
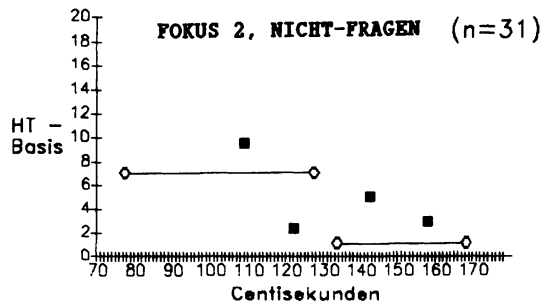


Fig.7: Sprecher 3 (Sp3)

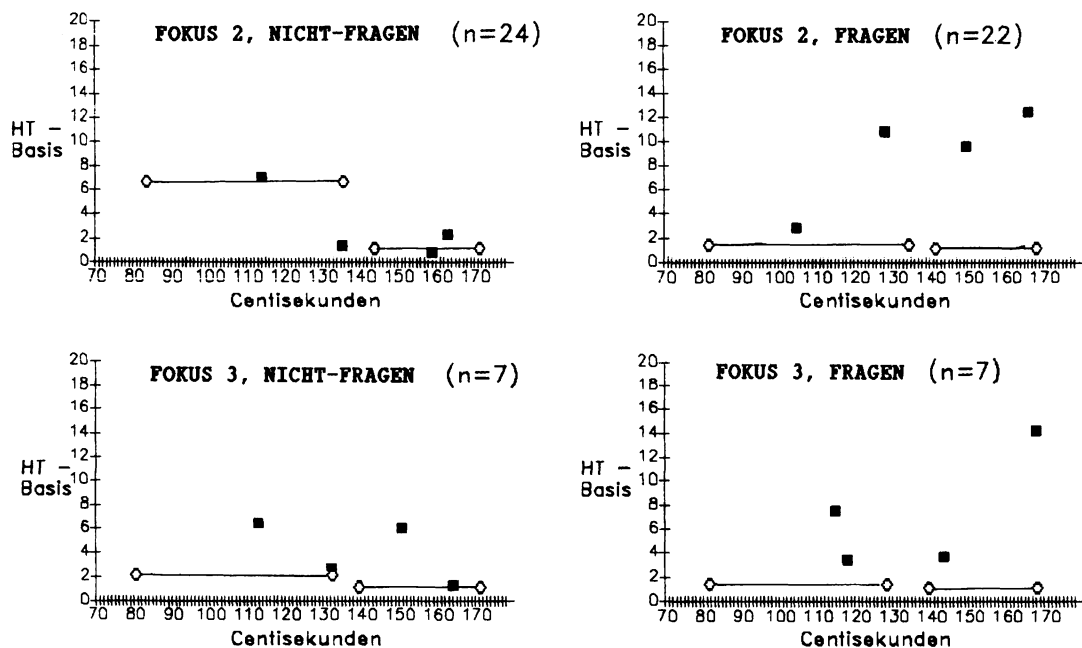


Fig.8: Sprecher 4 (Sp4)

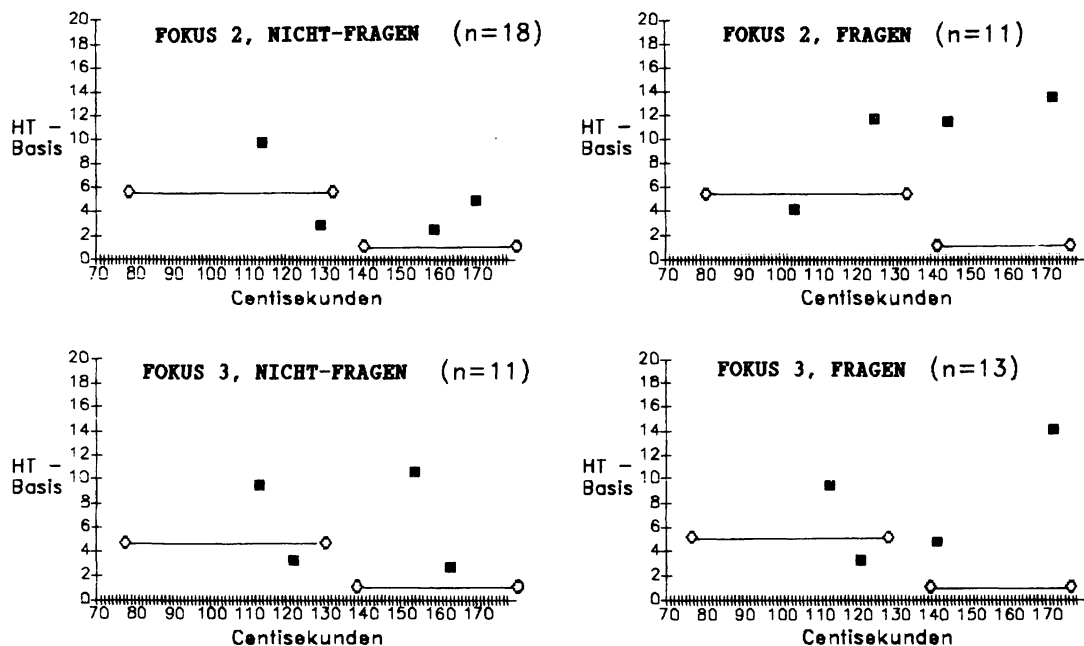


Fig.9: Sprecher 5 (Sp5)

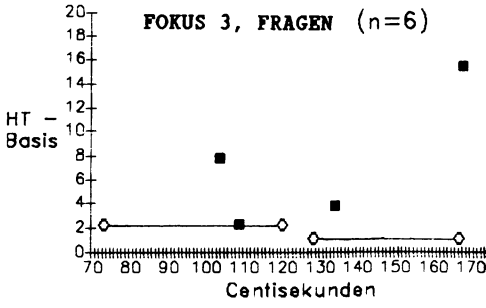
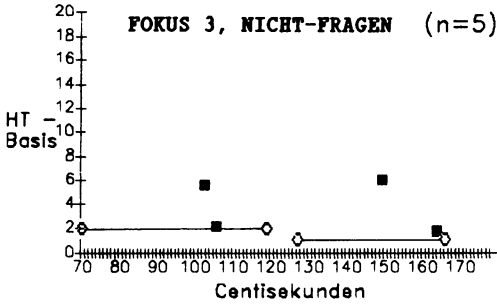
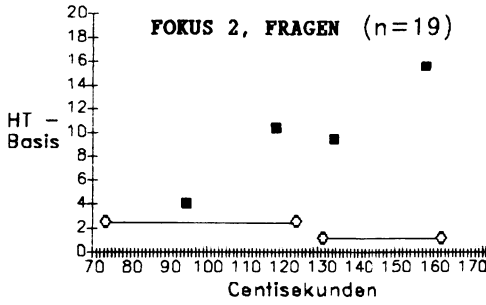
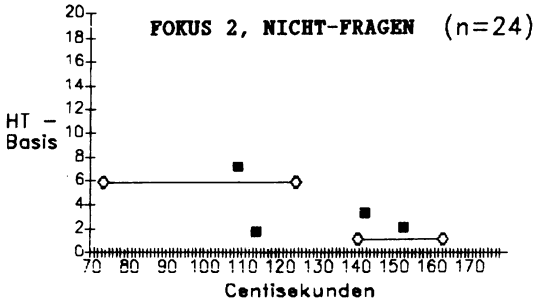
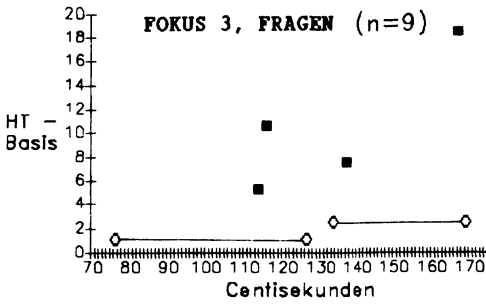
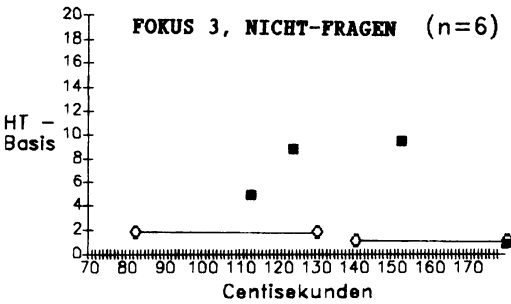
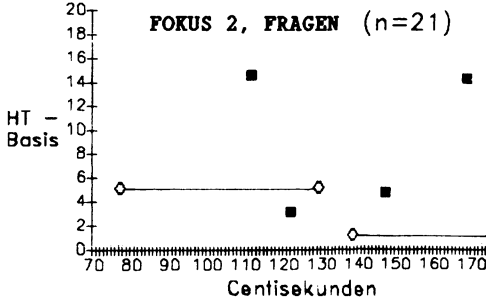
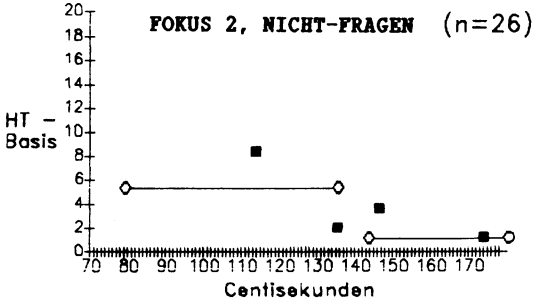


Fig.10: Sprecher 6 (Sp6)



In Fig.5 bis 10 ergibt sich bei den einzelnen Sprechern grosso modo das gleiche Bild, allerdings mit einigen interessanten Abweichungen, die nun besprochen werden.

- Sp1: Bei FOKUS 2, FRAGEN weist die dritte unbetonte Phrase keinen Fo-Anstieg, sondern einen Abfall von im Mittel 5 Htn auf. Der Abfall ist je nach Ausprägung perceptiv unterschiedlich auffällig: Mal wird eher ein ebener Verlauf wahrgenommen, mal ein leicht oder stark fallender. Immer unterscheiden sich die Äußerungen aber eindeutig von Äußerungen mit einem steigenden Verlauf. 20 Äußerungen wurden von Sp1 auf diese Weise produziert. Wir können damit das seltsam anmutende Ergebnis in Tab.9 erklären: Bei MODUS wurden bei der Konstellation (5), also bei einer Lernstichprobe von je einem Sprecher, bessere Ergebnisse erzielt als bei der 'leave-one-out'-Methode von Konstellation (4). Im letzten Fall 'verläßt' sich das Verfahren auf die fast immer entscheidende Rolle der Offset-Höhe bei der Klassifizierung von Fragen und Nicht-Fragen auch bei der Klassifizierung von Sp1; die 20 Fälle werden dabei fehlklassifiziert. Wenn das Verfahren aber von Sp1 ausgehend die anderen Sprecher klassifiziert, so 'verläßt' es sich auf die bei Sp1 relevanten Variablen *Stg* und *Max3* (vgl. Tab.10) und kann auf dieser Grundlage die anderen Sprecher gut klassifizieren.
- Sp2: Diese Sprecherin setzt die Intensität relativ stark ein (vgl. *Int3* in Tab.11 und 12).
- Sp3: Bei Nicht-Fragen wird die Intensität deutlich eingesetzt, bei Fragen kann man keinen Unterschied feststellen.
- Sp4: *Dau3* wird bei Nicht-Fragen von diesem Sprecher weniger eingesetzt als etwa von Sp1. Der Unterschied zwischen Sp4 und Sp1 bei den entsprechenden Zeileneinträgen von *Dau3* in Tab.12 ist gleich der Spannweite dieser Variable. Standardabweichung und Spannweite liegen im mittleren Wertebereich. Daraus kann man zum einen schließen, daß dieser Parameter relativ konsistent eingesetzt wird, daß aber auf der anderen Seite doch sprecherspezifische Unterschiede bestehen.
- Sp5: Bei Nicht-Fragen sehen wir deutliche Intensitäts- und Dauerunterschiede, bei den Fragen nur Dauer-, aber keine Intensitätsunterschiede.
- Sp6: Der Tonverlauf bei FOKUS 3, NICHT-FRAGEN ist bei der 2., unbetonten Phrase ein Anstieg, kein Abfall wie bei den anderen Sprechern ('Hut-Kontur' nach Cohen/'t Hart 1967; 'Brückenkontur' nach Wunderlich 1988). Ein solcher Verlauf ist – unterstellt man unserer Stichprobe eine gewisse Repräsentativität – zwar voll akzeptabel, aber nicht häufig und möglicherweise sprecherspezifisch. Intensität wird sowohl bei Frage als auch bei Nicht-Frage stark eingesetzt: nur bei diesem Sprecher kehrt sich bei FOKUS 3, FRAGEN die Relation von *Int2* und *Int3* um. Das könnte natürlich durch den sehr großen Range auf der 3. Phrase bei diesem Sprecher, bedingt sein; Sp5 hat aber einen größeren Range, ohne daß sich das Verhältnis bei der Intensität umkehrt. Bei FOKUS 2, FRAGE weist die fokussierte Phrase im Mittel einen starken Abfall auf, keinen Anstieg wie bei den Mittelwerten aller anderen Sprecher. Dieser Sprecher macht also eine Art 'Anleihe' im Akzentinventar der Nicht-Fragen, ohne daß die Äußerung deswegen unnatürlich ist (s. Teil 8). Ein Anstieg wie in Fig.2 ist also bei dieser

Konstellation zwar fragetypisch, aber das Gegenteil nicht irregulär. Wir können daher schließen, daß wir es hier mit einem Randtyp zu tun haben.

Die ins Auge fallenden und soeben aufgeführten Unterschiede zwischen den Sprechern, die man auch an den Extremwerten in Tab.11 und 12 festmachen kann, sind also auf die 'untypischen' Verläufe bei Sp1 und Sp6 zurückzuführen.

Wenn wir eine - vorsichtige - Gesamtinterpretation versuchen, so kann man sagen, daß bei allen Parametern eine gewisse Varianz tolerabel ist; diese Varianz ist für die normale Spannweite in Tab.11 und 12 verantwortlich. Bei den Tonverläufen gibt es aber größere Diskrepanzen, die darauf zurückzuführen sind, daß man hier auf ein unterschiedliches Markierungsinventar zurückgreifen kann. Diese Sachlage stützt ein Prototypenkonzept, das nicht mit Regeln, sondern mit typischen Merkmalsbündeln arbeitet: Eine gewisse Varianz kann auf Unschärfen in der phonetischen Realisierung der Regeln zurückgeführt werden. 'Gegenläufige', also stark unterschiedliche Realisierungen, wie wir sie vorfinden, könnten aber nur mit Regeln erfaßt werden, die mal die eine, mal die andere Realisierung erlauben. Eine Klassifikation anhand typischer, aber nicht immer obligatorischer Merkmale erscheint hier einleuchtender. (Vgl. dazu Medin/Barsalou 1987:463).¹³

Wir müssen im Augenblick dahingestellt lassen, inwieweit die Korrelationen der Prädiktorvariablen untereinander automatisch, also produktionsbedingt sind; ein ausgeprägter Fo-Anstieg bedingt z.T. auch eine größere Dauer als ein sehr geringer Fo-Anstieg, ein höheres Fo-Maximum erhöht die Intensität, usw. Auf der anderen Seite lassen unsere Daten doch den Schluß zu, daß die Parameter in einem gewissen Ausmaß auch willkürlich und voneinander unabhängig gesteuert werden können.

8. Gewinnung der Prototypen über Hörerurteile

Wir wollen nun einen anderen Weg der Gewinnung von Prototypen beschreiben, der uns zum gleichen Ziel führen wird. Man kann Prototypen als die Fälle auffassen, die in Hörexperimenten als die 'natürlichsten' bestimmt werden, wobei wir Natürlichkeit operational im Sinne unseres 'Natürlichkeitstests' (vgl. Teil 2.2) definieren. (Zu einem ähnlichen Vorgehen, allerdings mit anderen Schwel-

¹³ In einigen Fällen kann man eventuell pragmatische oder textlinguistische Verwendungsbedingungen festlegen, die unterschiedliche Realisationen möglich bzw. regulär machen; Beispiele wären die 'Fragehaltigkeit' oder die 'intonatorische Kongruenz' in Batliner (1989b). Bei unserem Material wurden aber die Randbedingungen - also auch die pragmatischen Verwendungsbedingungen - bei allen sechs Sprechern konstant gehalten.

lenwerten, vgl. Oppenrieder 1988b und Batliner 1988). Wir wollen die Schwelle nun sehr hoch ansetzen und definieren als Prototypen die Fälle, die bei |FOK| einen Wert von 1.0 (100%ige Übereinstimmung), bei MOD einen Wert von 80% für Fragen bzw. 20% für Nichtfragen (sehr gute Kategorisierung) und im Natürlichkeitstest Werte unter 2.0 (passend bis 'sehr passend') erreichen.¹⁴ Auf diese Weise erhalten wir gut 6% der Fälle (24 Äußerungen). Mit fünf Ausnahmen entsprechen die Verläufe denen aus Fig.2, also den über die große Zahl gewonnenen Prototypen. Fig.11 bis 14 sind typische Vertreter für je eine der vier Modus- und Fokuskonstellationen; sie illustrieren an konkreten Äußerungen die Mittelwerte von Fig.2. Von oben nach unten finden sich in diesen Abbildungen das Zeitsignal, die Intensität, der Verlauf der Ht-transformierten Fo-Werte sowie der Verlauf der Fo-Werte.¹⁵

Für drei der fünf Ausnahmen (Randtypen) steht Fig.15, also der für Sp6 typische Verlauf bei FOKUS 2, FRAGEN. Die beiden anderen sind von Sp5 produziert, der also in seinem Inventar über beide Fokussierungsarten verfügt (vgl. Fig.9, die verdeutlicht, daß für diesen Sprecher die Mittelwerte bei FOKUS 2, FRAGEN dem Kerntyp entsprechen). Fig.16 zeigt den für Sp6 typischen Verlauf bei FOKUS 3, NICHT-FRAGEN (vgl. die Mittelwerte in Fig.10), Fig.17 den für Sp1 typischen Verlauf bei FOKUS 2, FRAGE (vgl. die Mittelwerte in Fig.5).

¹⁴ Die exakten Schwellenwerte sind natürlich arbiträr. Man könnte untersuchen, wie sich eine schrittweise Veränderung dieser Werte im einzelnen auswirkt.

¹⁵ Bei den Fo-Werten wird für je drei Frames ein Wert dargestellt, wobei ein Frame 12,5 msec umfaßt. Die Skalierung der Ht-Achse ergibt sich aus $(Ht \cdot 100)$; zur Skalierung der Intensitätswerte vgl. Teil 2.1. Geringfügige fehlende Übereinstimmungen zwischen Zeitsignal und Fo-Verlauf sind auf die automatische Stimmhaft-Stimmlos-Entscheidung zurückzuführen, die der Fo-Extraktion vorausging.

Fig.11: FOKUS 2, NICHT-FRAGE (Sp5)

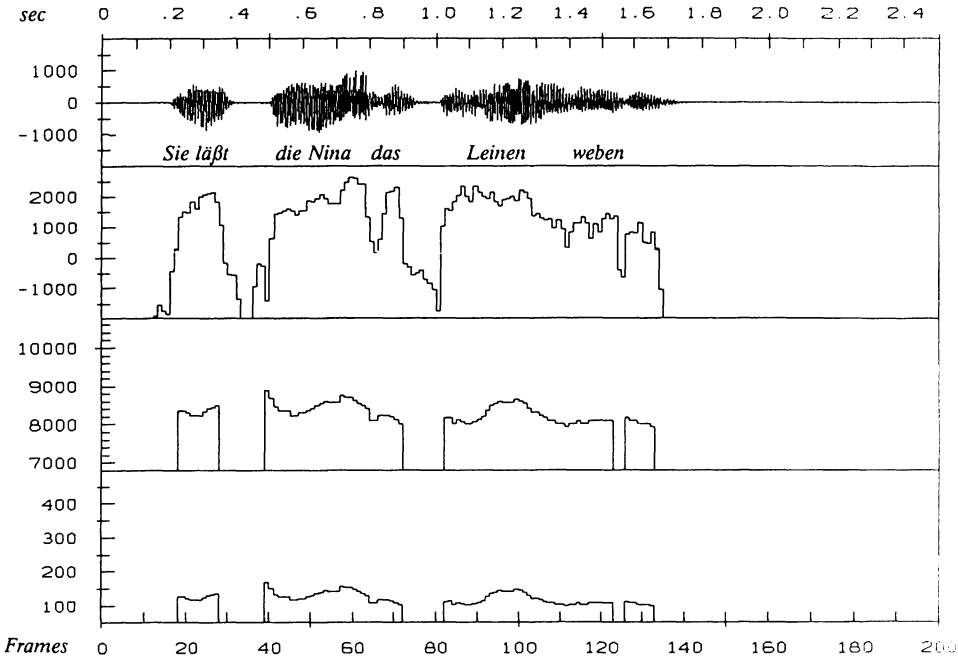


Fig.12: FOKUS 3, NICHT-FRAGE (Sp5)

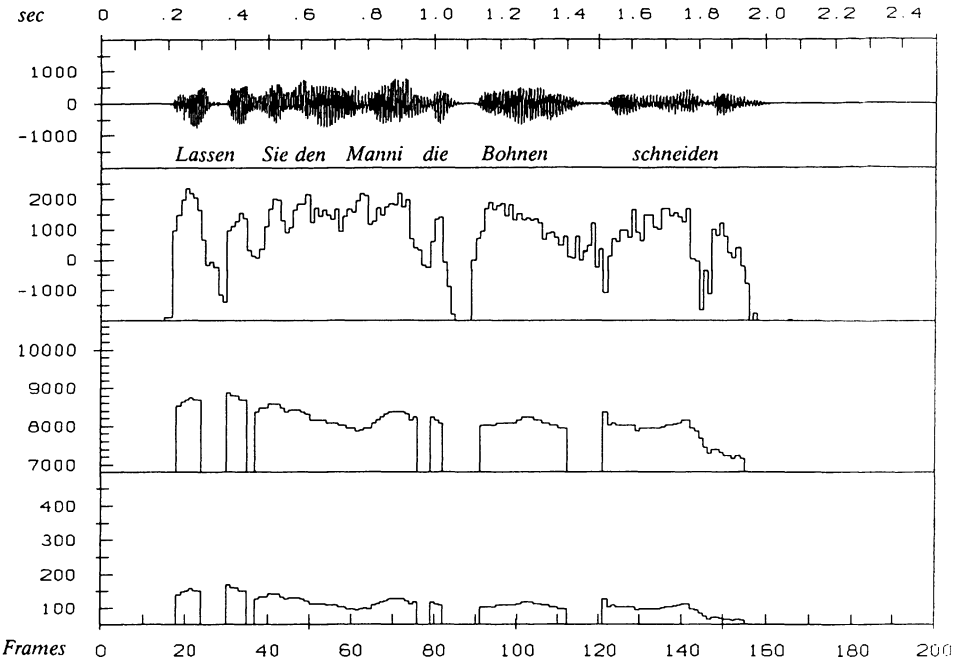
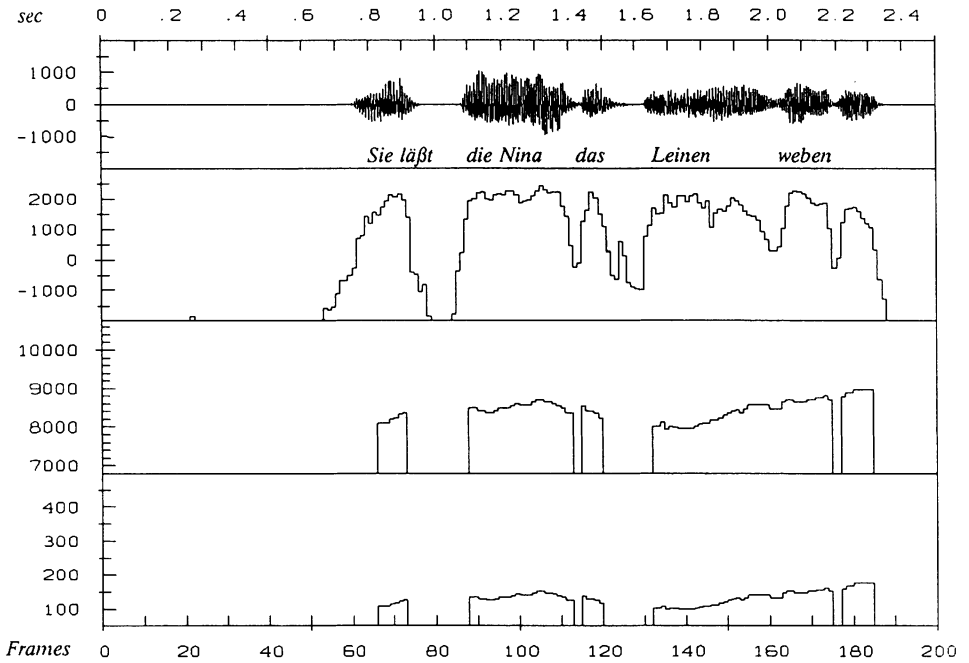
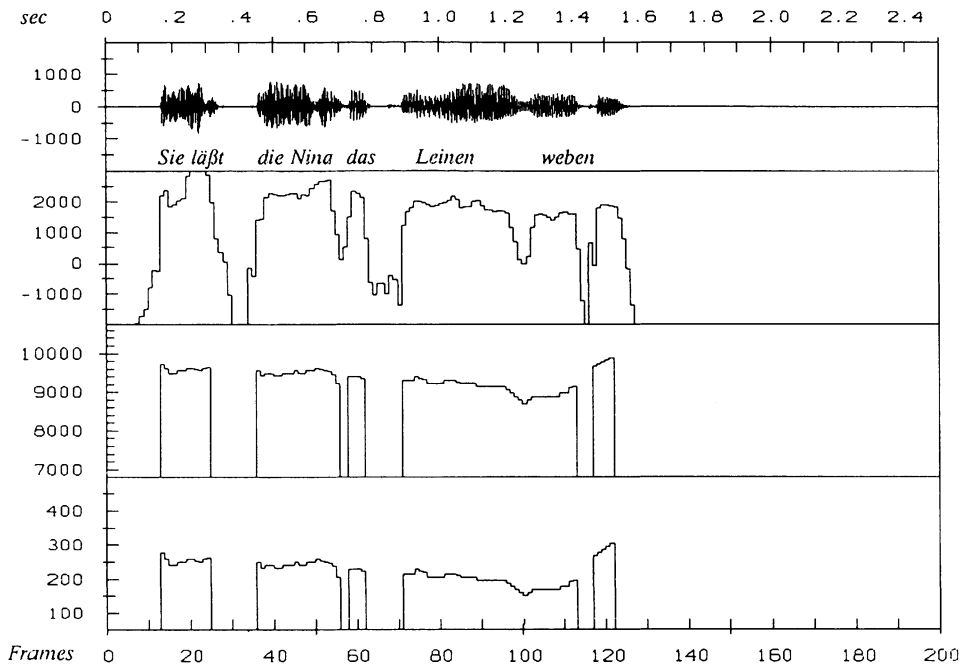


Fig.13: FOKUS 2, FRAGE (Sp5)**Fig.14: FOKUS 3, FRAGE (Sp1)**

Man sieht, daß sich die beiden Methoden der Prototypengewinnung z.T. ergänzen, z.T. komplementär sind: mit beiden Methoden erhält man die repräsentativen (Kern-) Typen; Randtypen, die nicht repräsentativ, aber ebenso akzeptabel sind, erhält man dagegen nur über die Hörerurteile – wobei wir immer die Einschränkung machen müssen, daß dies alles nur gilt, wenn die Sprecher (= die Stichprobe) in einem gewissen Ausmaß repräsentativ sind für die untersuchte Varietät (= die Population).

Die Abbildungen nach Fig.17 sind Batliner/Nöth (1989) entnommen; in ihnen sind die Mittelwerte der Kerntypen aus Fig.2 (st_{bas} = HT-Basis, t = Centisekunden) auf die konkreten Verläufe der über die Hörerurteile gewonnenen Prototypen von Fig.11 bis 17 projiziert. Dadurch ist ein direkter Vergleich möglich. Neben den Abbildungen sind jeweils ihre Entsprechungen in Fig.2 bzw. in Fig.11 bis 17 angegeben. Die gestrichelte Linie zeigt die Grenze zwischen der 2. und der 3. Phrase für die konkrete Äußerung.

Fig.15: FOKUS 2, FRAGE (Sp6)

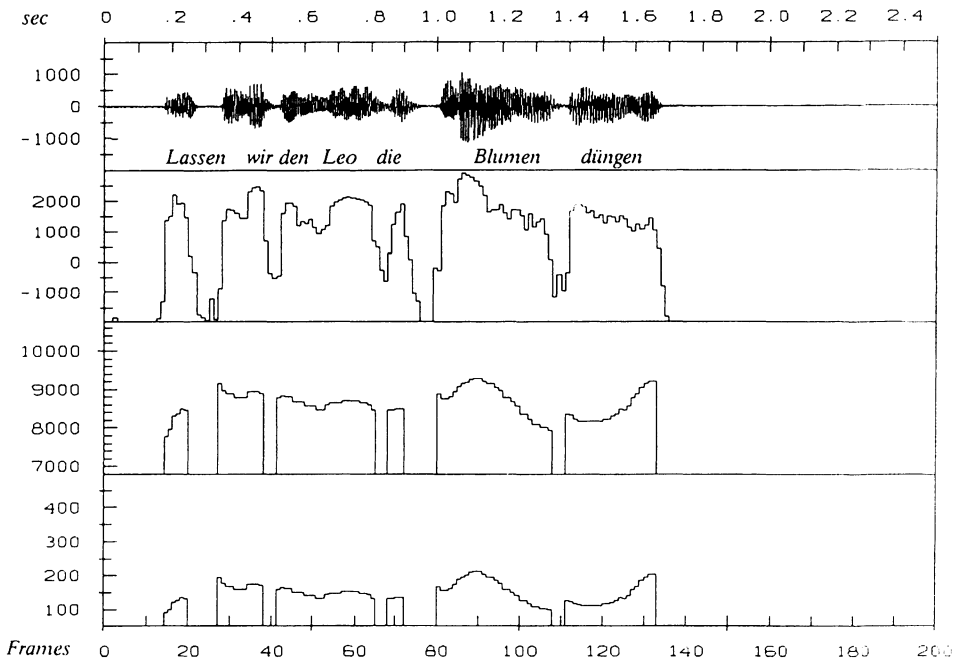
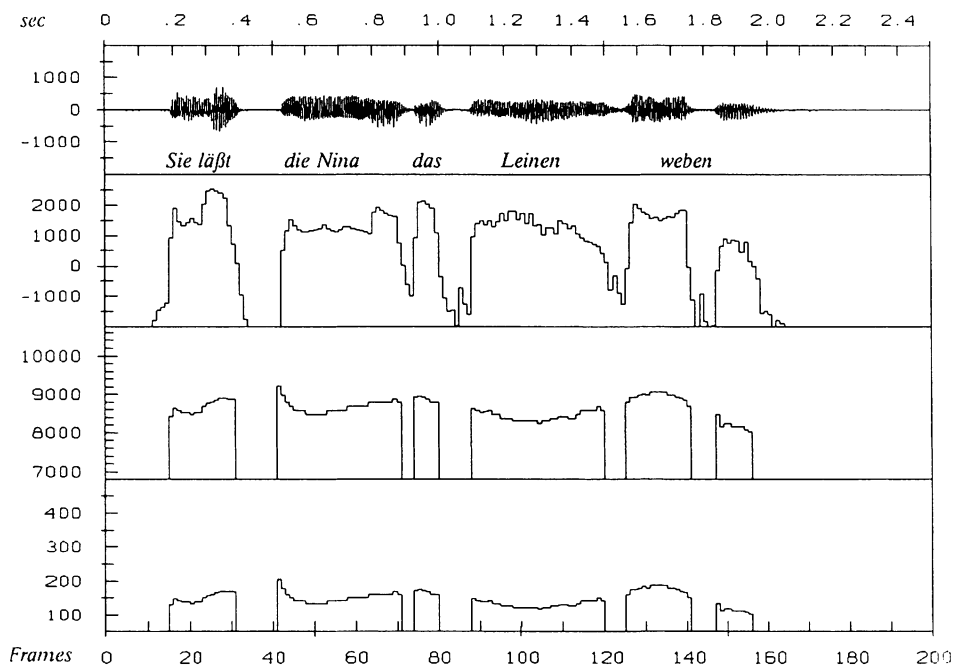
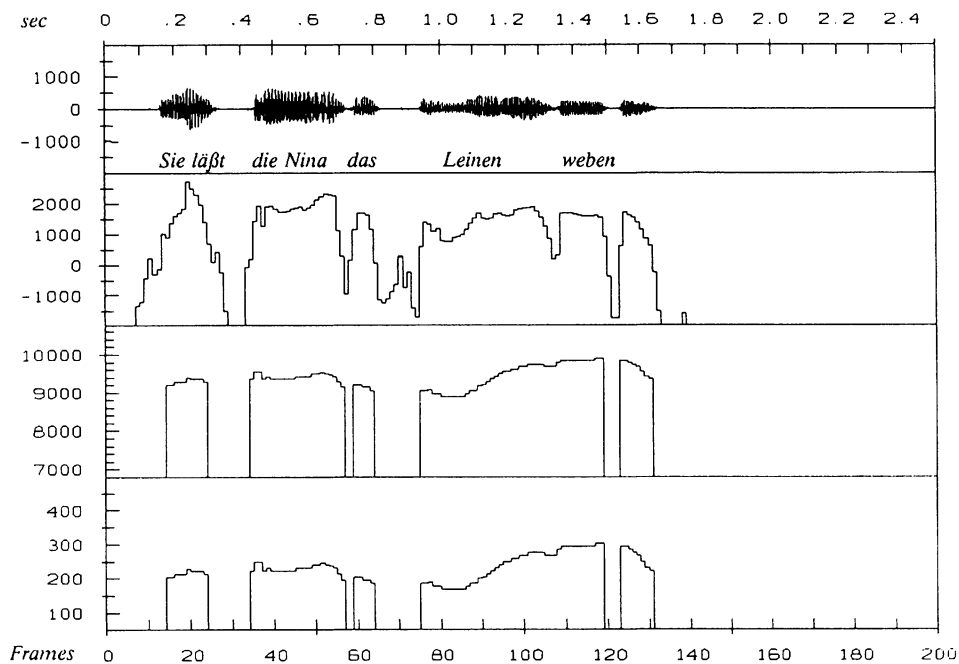
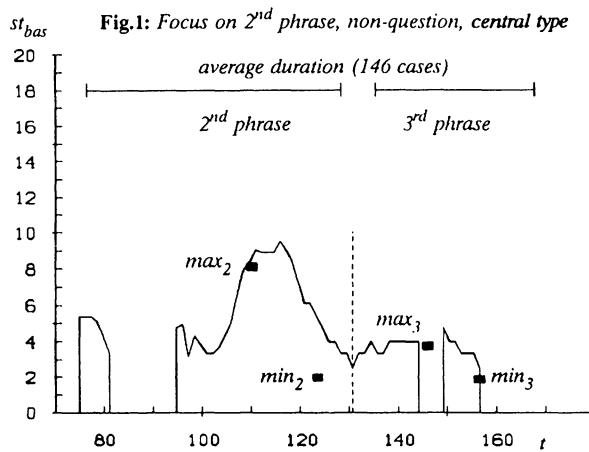


Fig.16: FOKUS 3, NICHT-FRAGE (Sp6)**Fig.17: FOKUS 2, FRAGE (Sp1)**

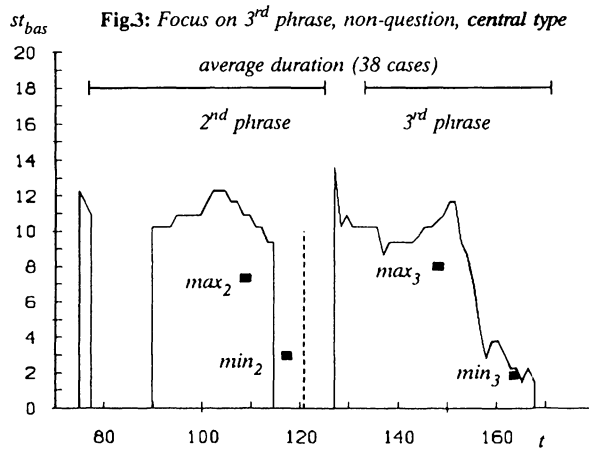
Aus Batliner/Nöth (1989):
Projektion von

Fig.2 (oben links)
auf Fig.11



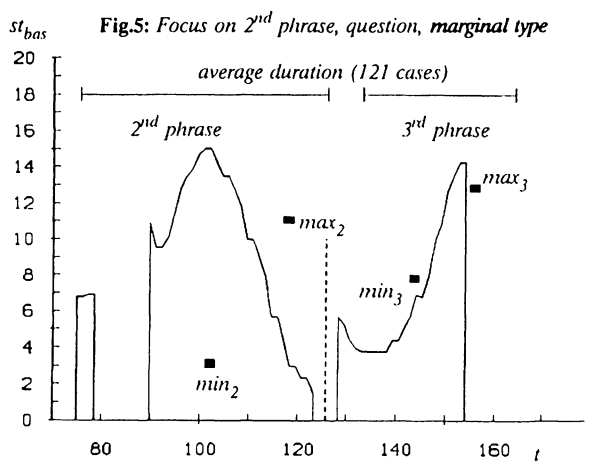
Projektion von

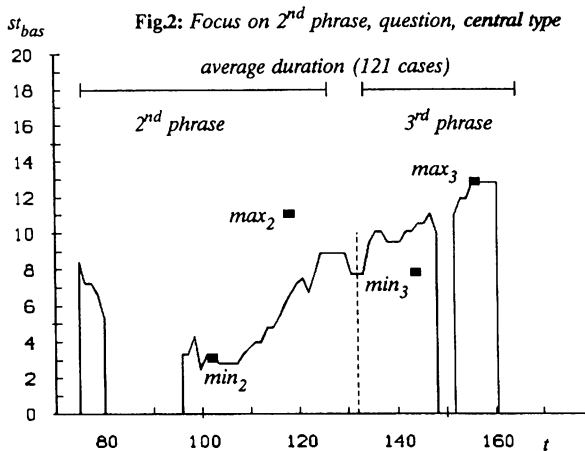
Fig.2 (unten links)
auf Fig.12



Projektion von

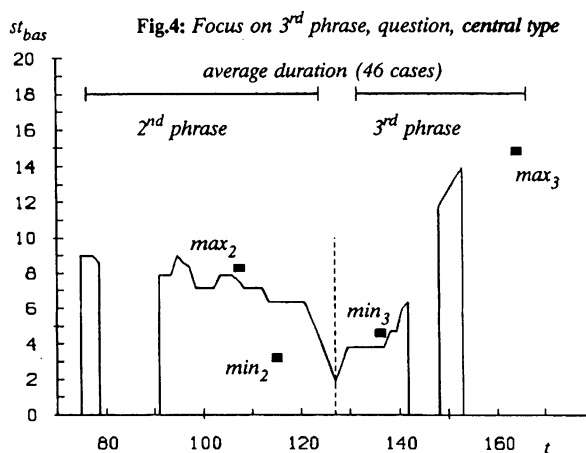
Fig.2 (oben rechts)
auf Fig.15





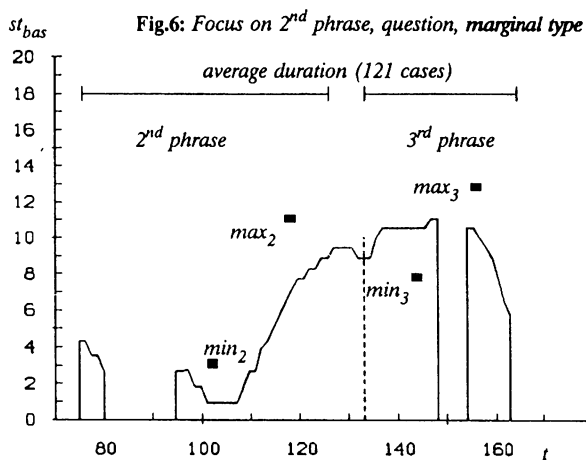
Projektion von

Fig.2 (oben rechts)
auf Fig.13



Projektion von

Fig.2 (unten rechts)
auf Fig.14

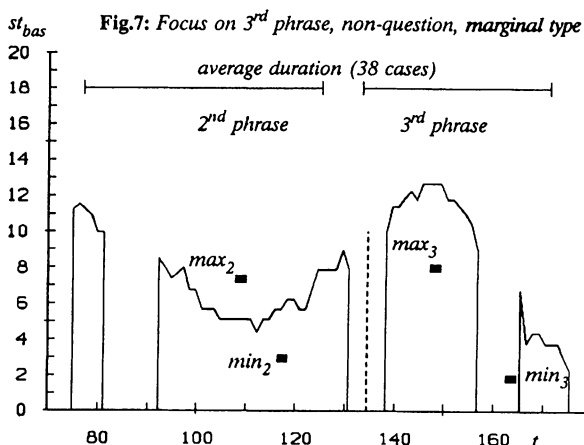


Projektion von

Fig.2 (oben rechts)
auf Fig.17

Projektion von

Fig.2 (unten links)
auf Fig.16



9. Fehleranalyse der Klassifikation durch die Diskriminanzanalyse

Die Fehleranalyse wird anhand der Konstellation (3) aus Tab.9, also für Lernstichprobe = Prüfstichprobe, durchgeführt. Der Unterschied zwischen dieser Konstellation und Konstellation (4) ('leave-one-out') liegt zwischen 2% und 8%. Wir legen also eine Klassifikation zugrunde, die in diesem Ausmaß auf Überadaptation beruhen dürfte (vgl. Ruske 1988:25). Auf der anderen Seite ist zu bedenken, daß die Stichprobe immer noch relativ klein ist; man kann also damit rechnen, daß sich bei einer genügend großen Stichprobe die Prädiktion auch bei 'leave-one-out' verbessern würde (für $n \rightarrow \infty$ gilt: Die Klassifikationsgüte bei 'leave-one-out' geht gegen die Klassifikationsgüte bei Lern- = Prüfstichprobe). Wir entschieden uns für dieses Vorgehen, da bei dieser Konstellation wirklich nur die Fälle fehlerklassifiziert werden, bei denen selbst eine sehr gute, überadaptierte Klassifikation versagt (das heißt nicht unbedingt, daß das statistische Modell für diese Fälle nicht adäquat ist; vgl. Teil 9.2 und 10).

Obwohl die Güte der Klassifikation sehr hoch ist, bleiben noch 27 Fehlerklassifikationen, 10 bei MODUS und 19 bei FOKUS; d.h. daß zwei Fälle sowohl bei MODUS als auch bei FOKUS fehlerklassifiziert wurden. Bei einer automatischen Zuweisung zu Frage/Nicht-Frage durch die Diskriminanzanalyse und darauffol-

gender Fokusprädiktion ist das Ergebnis sogar besser, da zwei Fälle ('Fehlproduktionen', vgl. Teil 9.1) 'dank' der falschen Moduszuweisung richtig klassifiziert werden.

9.1 Modus

In allen 10 Fällen wurden Fragen fehlkategorisiert; In 8 der 10 Fälle ist MOD gleich Null, d.h. die Hörer konnten im kontextfreien Kategorisierungstest die Äußerungen nicht als Frage erkennen. Die Natürlichkeitsbewertung war schlecht bis sehr schlecht (>3.5). Es handelt sich also bei diesen 'Fragen' mit einem tiefen Offset eindeutig um Fehlproduktionen. Eine Äußerung wurde zwar als Frage erkannt, erhielt aber eine sehr schlechte Natürlichkeitsbewertung. Es bleibt ein letzter und interessanter Fall: Diese Äußerung wurde zwar nicht als Frage erkannt, erhielt aber eine relativ gute Natürlichkeitsbewertung. Möglicherweise handelt es sich dabei um den in unserem Korpus singulären Fall einer assertiven Frage mit tiefem Offset (vgl. dazu Geluykens 1987 und Batliner/Oppenrieder 1988).

9.2 Fokus

Drei Beobachtungen können gemacht werden:

1. Bei den 19 Fehlklassifikationen von FOKUS ist bei 11 Fällen eine auffällige Diskrepanz (geringe Übereinstimmung oder gar Gegenläufigkeit) zwischen den beiden Akzenttests zu beobachten. Zum einen mag ein systematischer Einfluß des unterschiedlichen Versuchsdesigns die Ursache sein: Es ist nicht ganz klar, ob die Versuchspersonen das gleiche machen, wenn die Fälle unterschiedlich randomisiert sind (wie das beim 1. und beim 2. Akzenttest der Fall ist) und einmal nur die am stärksten betonte Silbe, das andere Mal zwei Akzentstufen vergeben werden müssen. Zum anderen sind Akzenturteile grundsätzlich nicht sehr konsistent, vgl. die Diskussion in Teil 10.
2. In 12 Fällen (60% der Fehlklassifikationen) ist die Akzentstruktur offensichtlich nicht sehr ausgeprägt, da FOK hier zwischen .4 und -.4 liegt. In diesem Bereich liegen nur 15% aller Fälle, vgl. Fig.1.
3. In 7 Fällen zeigt sich eine 'Unsicherheit' der Diskriminanzanalyse dadurch, daß die Wahrscheinlichkeit der Gruppenzuordnung nicht ausgeprägt ist, sondern nahe der 50%-Grenze liegt. 'Unsicherheiten' dieser Art sind sehr selten, normalerweise entscheidet sich die Diskriminanzanalyse mit hoher Wahrscheinlichkeit für eine der Gruppen. Bei solchen 'Grenzfällen' kann auch eine Verletzung der für die Diskriminanzanalyse erforderlichen Voraussetzungen eine Fehlklassifikation bewirken.

Wie man sieht, summieren sich die Zweifelsfälle auf mehr als 19, da in 11 Fällen mehr als eine der Beobachtungen gemacht werden konnte, also etwa eine

gleichzeitige 'Unsicherheit' der Diskriminanzanalyse und eine geringe Ausprägung der Akzentstruktur. Zwei Fälle bleiben, die auch durch Inspektion der Daten nicht erklärt werden können.

Bei den Fehlklassifizierungen von FOKUS muß man bedenken, daß die Zuweisung zu FOKUS 2 oder FOKUS 3 auf Hörerurteilen beruht, die – wie der Unterschied zwischen den beiden Akzenttests und die relativ hohe Anzahl von nicht sehr entschiedenen Urteilen zeigt – selbst mit einer gewissen Unschärfe behaftet sind. Diese Unschärfe war auch beabsichtigt, da der Kontext nur in einem Drittel der Fälle einen einfachen Fokus vorgab, ansonsten aber entweder Fokusprojektion oder zweifachen Fokus. Die Annahme erscheint also plausibel, daß das Verfahren unter ansonsten vergleichbaren Randbedingungen dann, wenn der Sprecher eindeutig fokussiert – und nur dann können die Hörer auch zu eindeutigen Urteilen kommen – praktisch fehlerfrei arbeitet.

10. Kerntypen, Randtypen und Trennschärfe

Ob ein Vokal kurz oder lang ist bzw. sein soll, kann man auch dem Aussprachelexikon entnehmen; ein vergleichbares FA-Lexikon gibt es nicht – und das ist kein Zufall: Ohne daß uns einschlägige Arbeiten bekannt wären, vermuten wir, daß Vokallängen und -kürzen von nativen Sprechern konsistenter produziert und auch eindeutiger perzipiert werden, als das bei den für den FA relevanten Parametern der Fall ist (diese Vermutung gilt für im Labor elizitierte Sprache; bei spontaner Sprache mag es anders aussehen). Wie auch schon Lieberman (1965) festgestellt hat, ändert sich die Transskription der Tonverläufe durch kompetente Linguisten in starkem Maße, wenn statt der natürlich produzierten Äußerungen synthetisierte Äquivalente der Fo-Verläufe vorgegeben werden. Brown et al. (1980:139ff) berichten von ähnlichen Schwierigkeiten bei der Zuweisung des *tonic* (also des FA); vgl. auch Lickey/Waibel (1985). Bei uns ergaben sich auffällige Diskrepanzen zwischen Akzenttest 1 und Akzenttest 2; sicher hätten sich weitere Diskrepanzen ergeben, wenn wir in einem zusätzlichen Test die Äußerungen in ihren modus- und fokussteuernden Kontexten zur Akzentzuweisung vorgegeben hätten.¹⁶ Erst bei einer Kontexteinbettung hätten aber die intonatorischen Merkmale den ihnen zukommenden Platz als eine – wenn auch wichtige – Gruppe fokusindizierender Merkmale unter anderen grammatischen (phonetischen, syntaktischen, semantischen).

¹⁶ Der normale 'Default'-FA liegt bei unseren Äußerungen auf der 2. Phrase. Ohne Kontext erwarten die Hörer diese Position, d.h. im Zweifelsfall dürften sie den FA dieser Position zuweisen.

Die von uns vorgenommene Zuweisung des FA auf die 2. oder die 3. Phrase anhand des Akzenttests ist also zwar intersubjektiv abgestützt; da ihr aber die Beurteilung der isolierten Äußerungen zugrundelag, kann nicht ausgeschlossen werden, daß die intonatorischen Merkmale im einzelnen oder der Komplex der intonatorischen Merkmale im ganzen im Zusammenhang mit all den anderen grammatischen Merkmalen anders gewichtet wären. Wenn wir also **innerhalb** der intonatorischen Merkmale multivariate (und damit 'realistischere' als univariate) Analysen durchführen, so arbeiten wir auf einer höheren Stufe trotzdem nur univariat, da wir von den relevanten grammatischen Merkmalskomplexen nur den intonatorischen berücksichtigen.

Eine gleichzeitige und systematische Berücksichtigung **aller** grammatischen Merkmale in einem experimentellen Design ist auf lange Sicht wegen des damit verbundenen Aufwands nicht vorstellbar; die Lage ist aber nicht hoffnungslos, wenn man sich von den unscharfen (*fuzzy*) Grenzen zwischen Kategorien ab- und den Kernbereichen zuwendet: Eindeutig fokussierte Phrasen werden vom Menschen auch eindeutig als fokussiert wahrgenommen und vom automatischen Verfahren richtig klassifiziert. Es dürfte also sinnvoll sein, sich stärker als bisher nicht mit Kategoriengrenzen (wie im Paradigma der Kategorialen Wahrnehmung, vgl. Schlefer/Batliner 1988) zu beschäftigen, sondern mit den Kernbereichen der Kategorien, eben den Prototypen.

11. Schlußbemerkungen

Wir haben versucht, nicht nur einen Faktor zu variieren und alle anderen konstant zu halten, sondern mehrere Faktoren zu variieren, um dadurch eine größere Nähe zur natürlichsprachlichen Kommunikation zu erreichen: es wurden Äußerungen **mehrerer Sprecher** untersucht, die **ohne explizite Instruktionen** und bei **gleichzeitiger Variorung von Modus und Fokus** produziert wurden. Dagegen stehen zwei Beschränkungen, zum einen die auf eine **bestimmte Varietät** des Deutschen und zum anderen die auf drei Testsätze mit **syntaktisch und segmental ähnlicher Struktur**. Die Ergebnisse sollten auch an **spontansprachlichen** Korpora überprüft werden.

LITERATUR

- Adriaens, L.M.H. (1984): A preliminary description of german intonation. In: IPO Annual Progress Report 19. S.36-41.
- Altmann, H. (1987): Zur Problematik der Konstitution von Satzmodi als Formtypen. In: Melbauer, J. (Hg.) (1987): Satzmodus zwischen Grammatik und Pragmatik. Tübingen. S.22-56.
- Altmann, H. (Hg.) (1988): Intonationsforschungen. Tübingen.
- Bannert, R. (1985): Fokus, Kontrast und Phrasenintonation im Deutschen. In: Zeitschrift für Dialektologie und Linguistik 52. S.289-305.
- Batliner, A. (1988): Produktion und Prädiktion. Die Rolle intonatorischer und anderer Merkmale bei der Bestimmung des Satzmodus. In: Altmann, H. (Hg.) (1988). S.207-221.
- Batliner, A. (1989a): Fokus, Deklination und Wendepunkt. (In diesem Band)
- Batliner, A. (1989b): Wieviel Halbtöne braucht die Frage? Merkmale, Dimensionen, Kategorien. (In diesem Band)
- Batliner, A./Nöth, E./Lang, R./Stallwitz, G. (1989): Zur Klassifikation von Fragen und Nicht-Fragen anhand intonatorischer Merkmale. (Erscheint in: Tagungsband der 15. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik (DAGA). 13.-16. März 1989)
- Batliner, A./Oppenrieder, W. (1988): Rising intonation: Not passed away but still alive. A reply to R. Geluykens. In: Journal of Pragmatics 12. S.227-233.
- Batliner, A./Oppenrieder, W. (1989): Korpora und Auswertung. (In diesem Band)
- Batliner, A./Nöth, E. (1989). The prediction of focus. Proceedings of the European Conference on Speech Communication and Technology, Paris, 26-28 September 1989. (Im Druck)
- Beckman, M.E. (1986): Stress and Non-Stress Accent. Dordrecht.
- Brown, G./Currie, K.L./Kenworthy, J. (1980): Questions of Intonation. London.
- Cohen, A./'t Hart, J. (1967): On the anatomy of intonation. In: Lingua 19. S.177-192
- Geluykens, R. (1987): Intonation and speech act type: An experimental approach to rising intonation in queclaratives. In: Journal of Pragmatics 11. S.483-494.
- 't Hart, J. (1986): Declination has not been defeated - A reply to Lieberman *et al.* In: J. Acoust. Soc. Am. 80. S.1838-1840.
- Hildebrand, D.K./Laing, J.D./Rosenthal, H. (1977): Analysis of Ordinal Data. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-008. Beverly Hills and London.
- Hörmann, H. (1978): Meinen und Verstehen. Frankfurt am Main.

- Klecka, W.R. (1980): Discriminant Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-019. Beverly Hills and London.
- Klein, W. (1980): Der Stand der Forschung zur deutschen Satzintonation. In: Linguistische Berichte 68. S.3-33.
- Klein, W. (1982): Einige Bemerkungen zur Frageintonation. In: Deutsche Sprache (1982). S.289-310.
- Lakoff, G. (1987): Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. Chicago und London.
- Lickey, S.A./Waibel, A. (1985): Perceptual Stress Assignments. (Anhang zu: Waibel, A. (1986): Prosody and Speech Recognition. Carnegie Mellon University, Computer Science Departement. S.192-198.
- Lieberman, P. (1965): On the Acoustic Basis of the Perception of Intonation by Linguists. In: Word 21. S.40-54.
- Lieberman, P. (1986): Alice in declinationland - A reply to Johan 't Hart. In: J. Acoust. Soc. Am. 80. S.1840-42.
- Lieberman, P./Katz, W./Jongman, A./Zimmerman, R./Miller, M. (1985): Measures of the sentence intonation of read and spontaneous speech in American English. In: J. Acoust. Soc. Am. 77. S.649-657.
- Medin, D.L./Barsalou, L.W. (1987): Categorization processes and categorical perception. In: Harnad, S. (Hg.) (1987): Categorical perception. The groundwork of cognition. Cambridge etc. S.455-490.
- Niemann, H. (1974): Methoden der Mustererkennung. Frankfurt am Main.
- Norusis, M.J. (1986): SPSSPC+ Advanced Statistics. Chicago.
- Nöth, E. (1989): Prosodische Information in der automatischen Spracherkennung - Berechnung und Anwendung. Diss., Lehrstuhl für Informatik 5, Universität Erlangen. (in Vorbereitung)
- Nöth, E./ Batliner, A./ Lang, R./ Oppenrieder, W. (1987): Automatische Grundfrequenzanalyse und Satzmodusdifferenzierung. In: Tillmann, H.G./Willée G. (Hgg.) (1987): Analyse und Synthese gesprochener Sprache. Hildesheim etc. S.59-66.
- Oppenrieder, W. (1988a): Intonation und Identifikation. Kategorisierungstests zur kontextfreien Identifikation von Satzmodi. In: Altmann, H. (Hg.) (1988). S.153-167.
- Oppenrieder, W. (1988b): Intonatorische Kennzeichnung von Satzmodi. In: Altmann, H. (Hg.) (1988). S.169-205.
- Oppenrieder, W. (1989): Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung. (In diesem Band)
- Pierrehumbert, J.B. (1980): The Phonology and Phonetics of English Intonation. PhD Dissertation, MIT.

- Ruske, G. (1988): Automatische Spracherkennung. Methoden der Klassifikation und Merkmalsextraktion. München/Wien.
- Schiefer, L./Batliner, A. (1988): Intonation, Ordnungseffekt und das Paradigma der Kategorialen Wahrnehmung. In: Altmann, H. (Hg.) (1988). S.273-291.
- Taylor, S./Wales, R. (1987): Primitive mechanisms of accent perception. In: Journal of Phonetics 15. S.235-246.
- Thorsen, N.G. (1987): Suprasegmental Transcription. In: Almeida, A./Braun, A. (Hgg.): Probleme der phonetischen Transkription. Stuttgart. S.79-109.
- Wunderlich, D. (1988): Der Ton macht die Melodie – Zur Phonologie der Intonation des Deutschen. In: Altmann, H. (Hg.) (1988). S.1-40.