PRODUKTION UND PRÄDIKTION. DIE ROLLE INTONATORISCHER UND ANDERER MERKMALE BEI DER BESTIMMUNG DES SATZMODUS.

Anton Batliner (München)

1. EINLEITUNG

Der vorliegende Beitrag ist eine erweiterte und modifizierte Version von Batliner (1987a). Er entstand im Rahmen des DFG-Projekts "Modus-Fokus-Intonation", das eine genauere Charakterisierung der intonatorischen Merkmale von Satzmodus und Fokus zum Ziel hat. Grundlage ist das von Hans Altmann entwickelte Satzmodussystem (vgl. Altmann 1987). Das Testsatzkorpus, die Extraktion der intonatorischen Parameterwerte sowie die daran anschließenden Hörtests werden im nächsten Abschnitt nur kurz charakterisiert, da sie genauer in Oppenrieder (1988) beschrieben sind.

2. SATZMODUSSYSTEM UND TESTSATZKORPUS

Unter einem **Satzmodus** wird die eindeutige Zuordnung eines **Funktionstyps**, z.B. der Ergänzungsfrage, und eines genau festgelegten **Formtyps** verstanden. Die einzelnen Formtypen können als **strukturierte Merkmalbündel** beschrieben werden, die sich in mindestens einem Merkmal voneinander unterscheiden. Nur solche Merkmale werden als satzmodusrelevant angesehen, die zumindest einmal für die Unterscheidung zweier Formtypen unverzichtbar sind. Die Merkmale stammen aus vier Mengen. Sie betreffen

(a) das Vorhandensein von Ausdrücken bestimmter **Kategorien**, beim Formtyp W-Fragesatz z.B. das Vorhandensein irgendeines W-Ausdrucks;
(b) die **Stellungseigenschaften** dieser satzmodusrelevanten Ausdrücke, z.B. die Zweitstellung des finiten Verbs;
(c) die indikativische, konjunktivische oder imperativische **Markierung des finiten Verbs**
(d) die **intonatorische Markierung**.

Wenn intonatorische Merkmale, die anders als die Merkmale aus den Bereichen (a) bis (c) in jeder Äußerung vorhanden sind, satzmodusrelevant sein sollen, müssen sie mindestens an einer Stelle für die Differenzierung zweier ansonsten

identischer Formtypen sorgen können. Genau an den Satzmodi, die sich formal nicht immer durch Merkmale aus den Bereichen (a) bis (c) unterscheiden lassen, können demnach die satzmodusrelevanten intonatorischen Parameter(-werte) aufgefunden und untersucht werden. Es wurden für alle jene Formtypen, die nur intonatorisch unterscheidbar sind, also für intonatorische Minimalpaare, Testsätze gebildet und in Kontexte eingebettet, die jeweils nur mit einem der Formtypen verträglich sind. Die Testsätze wurden zusammen mit ihrem modussteuernden Kontext sechs Versuchspersonen (drei weiblichen, drei männlichen) vorgelegt und von diesen mindestens zweimal realisiert; dabei ergaben sich 956 Äußerungen. Von jeder Äußerung wurde ein Mingogramm mit Zeitsignal (Oszillogramm), Fo-Kurve und Intensitätsverlauf erstellt, an dem die relevanten intonatorischen Parameterwerte wie Fo-Onset, Fo-Offset, Fo-Maximum, Fo-Minimum, Maximum der Intensität, Dauer der am stärksten akzentuierten Silbe etc. abgelesen wurden; diese Werte bildeten die Grundlage für die Berechnung weiterer Werte mit dem Statistikpaket SPSSPC+, wie des Tonumfangs ('Range') jeder Äußerung, der halbtontransformierten Werte etc. Alle Realisationen wurden Hörtests unterzogen, bei denen im Schnitt 12 Versuchspersonen die Testsatzrealisationen u.a. (1) auf ihre Natürlichkeit im Kontext hin beurteilen und sie (2) kontextfrei Moduskategorien zuweisen sollten. Ziel war dabei zuerst eine **deskriptive** Statistik der intonatorischen Parameter, die gegebenenfalls prüfstatistisch überprüft wird. Ergänzend zu diesem Verfahren wurde aber auch ein geeignetes **klassifizierendes** Verfahren, die sog. Diskriminanzanalyse, auf das Korpus angewandt. Es sollte damit untersucht werden, inwiefern automatische Verfahren bei Untersuchungen zur Intonation sinnvoll eingesetzt werden können. In unserem Fall handelt es sich zunächst um teilweise automatische Verfahren, da die Parameterwerte von Hand extrahiert wurden. Die Relevanz solcher Verfahren für die automatische Spracherkennung versteht sich von selbst; im Rahmen der Grundlagenforschung bietet sie eine sinnvolle Korrektur für das notwendigerweise beschränkte Material, mit dem Linguisten und Phonetiker üblicherweise arbeiten. Mit ihrer Hilfe wäre es möglich, große Korpora weitgehend automatisch zu bearbeiten und damit die Beschreibung zunehmend zu verbessern.

3. BESCHREIBUNG DES VERFAHRENS

Die Diskriminanzanalyse ist ein Verfahren, bei dem, basierend auf relevanten Variablen (in unserem Fall Fo-Onset, Fo-Offset etc.), die einzelnen Fälle (die Äußerungen unseres Korpus') distinkten Kategorien (den intendierten Satzmodi) zugewiesen werden. Das Verfahren funktioniert ähnlich wie die multiple Regression: Es werden lineare Kombinationen der unabhängigen 'Prädiktor'-Varia-

blen gebildet, die möglichst optimal zwischen den Kategorien unterscheiden können. Jede Prädiktorvariable erhält einen Gewichtungskoeffizienten, der anhand der vorgegebenen Daten so geschätzt wird, daß die resultierende Diskriminanzfunktion zwischen den Gruppen so stark wie möglich differiert; anders gesagt, die Variabilität **zwischen** den Gruppen soll im Verhältnis zur Variabilität **innerhalb** der Gruppen möglichst groß sein. Prädiktorvariablen sollten normalverteilte kontinuierliche Werte aufweisen, d.h. Fo-Werte und davon abgeleitete Werte sind auf alle Fälle gute Kandidaten. (Das Verfahren funktioniert aber auch oft recht gut bei dichotomen Variablen.) Es kann das gesamte Material Grundlage der Analyse sein und dann klassifiziert werden, es kann aber auch nur ein Ausschnitt aus dem Material (die 'Lernstichprobe') analysiert und Grundlage der Klassifikation eines anderen Ausschnitts (der 'Prüfstichprobe') werden. Bevor wir im nächsten Abschnitt unser spezielles Vorgehen beschreiben, wollen wir auf einige allgemeine Punkte eingehen:

Erwartungswert: Maßstab für die Güte der Prädiktion ist der Prozentwert der Fälle, die mit den jeweils angenommenen Prädiktorvariablen richtig klassifiziert werden, d.h. in unserem Fall, einem der im Korpus vorhandenen fünf Haupt-Modi **Frage (-satz)**, **Aussage (-satz)**, **Exklamativ (-satz)**, **Imperativ (-satz)** oder **Wunsch (-satz)** richtig zugeordnet werden. Dieser Wert muß immer in Relation gesetzt werden zu einer zufällig richtigen Zuordnung (dem 'Erwartungswert'), deren Wahrscheinlichkeit sich ergibt, wenn man 100 durch die Anzahl der Gruppen dividiert. Bei zwei Gruppen würde eine zufällige Verteilung im Schnitt 50%, bei fünf Gruppen 20% richtige Zuordnungen ergeben.

Auftretenswahrscheinlichkeit: Die Diskriminanzanalyse setzt normalerweise die Wahrscheinlichkeit, mit der Fälle, über die keine Information vorliegt, einer bestimmten Gruppe zugeordnet werden, für alle Gruppen gleich an; d.h. bei fünf möglichen Gruppen erhält jede eine Wahrscheinlichkeit von 20%. Wenn die Verteilung der Gruppen in der Stichprobe der Verteilung in der Grundgesamtheit (der 'Population') entspricht und von der Gleichverteilung abweicht, so kann die Auftretenswahrscheinlichkeit der einzelnen Gruppen vorgegeben und damit eine bessere Prädiktion erzielt werden, da die Auftretenswahrscheinlichkeit in die Berechnung der Prädiktion mit eingeht. Was die Verteilung der Satzmodi in der Population aller in einem bestimmten Zeitraum im deutschen Sprachraum realisierten Äußerungen betrifft, so läßt sich darüber nur spekulieren. Es kann aber als sicher angenommen werden, daß diskursspezifische Stichproben Ungleichverteilungen aufweisen. So werden beim Militär relativ gesehen mehr Imperativsätze realisiert werden, im Reisebüro oder bei einer Mensch-Maschine-Kommunikation mehr Fragesätze usw. So gesehen, ist die Ungleichverteilung in unserem Korpus, die sich aus der speziellen Form der Minimalpaarauswahl ergibt, gar nicht so ungewöhnlich. Eine Berücksichtigung dieses Umstands führt zwar zu besseren Ergebnissen; wir werden aber eine Gleichverteilung annehmen, da die Zugrundelegung aller möglichen Minimalpaare in unserem Korpus zur Folge hat, daß die stark überwiegenden Fragesätze immer auf Kosten der anderen Satzmodi besser klassifiziert werden.

Prädiktorvariablen: Bei jeweils nur einer Variablen als Prädiktor ist eine Interpretation der Ergebnisse noch relativ einfach. Sie wird schwieriger bei mehreren Variablen, da die Berücksichtigung einer zusätzlichen Variablen, von der man schon weiß, daß sie für eine bestimmte Distinktion relevant ist, nicht automatisch eine bessere Klassifikation ergibt, wenn alle Fälle und alle Gruppen klassifiziert werden. Eine Optimierung der Variablenauswahl ist natürlich erstrebenswert, bei unserem Testsatzkorpus allerdings noch nicht zu verwirklichen. Das sei an einem Beispiel erläutert: Eine Dehnung der Hauptakzentsilbe oder der ganzen Äußerung ist sicher eines der Merkmale, mit denen man einen Exklamativsatz kennzeichnen kann. Wenn man nun die Länge der Hauptakzentsilbe oder der Äußerung als Prädiktorvariable einsetzen möchte, so muß man natürlich wissen, um welche Segmente es sich handelt (Langvokal oder Kurzvokal etc.). Wenn darüberhinaus genug Exemplare mit der gleichen segmentalen Struktur im Korpus vorhanden sind, so kann man entweder nur die gleichen Strukturen miteinander vergleichen oder zumindest die relative Länge der Hauptakzentsilbe sinnvoll bestimmen (vgl. zu diesem Vorgehen Batliner 1987b). Da unser Korpus aber nicht auf diese Fragestellung hin konzipiert ist, gibt es in ihm zuwenig Fälle, bei denen die (relative) Länge als Prädiktorvariable sinnvoll angenommen werden könnte. Wir werden also nicht mit Variablen aus dem Zeitbereich, sondern nur mit Variablen aus dem Frequenzbereich arbeiten. Wir sehen auch davon ab, die einzelnen Diskriminanzfunktionen zu interpretieren (Welche Satzmodi werden in welchem Ausmaß durch welche Prädiktorvariable bestimmt?), da eine solche Interpretation in diesem Stadium noch recht fehleranfällig wäre.

4. STICHPROBENAUSWAHL UND TRANSFORMATION DER PRÄDIKTORVARIABLEN

Wir legen im folgenden immer die vier Fo-Variablen Onset (Fo-Wert am Beginn der Äußerung), Offset (Fo-Wert am Ende der Äußerung), Maximum (absolut höchster Fo-Wert der Äußerung) und Minimum (absolut tiefster Wert der Äußerung) den Berechnungen zugrunde. (D.h. daß die Variablen auch zufällig gleiche Werte annehmen und manchmal - z.B. bei Fragen mit einem sehr hohen Offset, der gleichzeitig das Maximum bildet - zusammenfallen können.) Diese Variablen haben den Vorteil, daß sie eindeutig bestimmbar sind, nicht auf der Zeitachse positioniert werden müssen und keine zusätzliche Information voraussetzen. (Eine solche zusätzliche Information ist z.B. beim Konturverlauf auf der Hauptakzentsilbe nötig, da diese Silbe ja erst bestimmt werden muß - z.B. durch Hörtests, so wie es in Oppenrieder 1988 beschrieben wird.) Die Variablen können sowohl mit analogen Geräten und nachfolgender Messung per Hand, wie in unserem Fall, als auch mit automatischen Algorithmen ermittelt werden. Den meßtechnischen Nachteil, daß sie ab und an nicht bestimmt werden können - etwa wenn der Sprecher am Ende der Äußerung laryngalisiert-, teilen sie mit anderen Variablen. (Bei unserer Stichprobe konnten wir in ca 5% der Fälle

einen der Fo-Werte nicht bestimmen. Solche Fälle werden von der Diskriminanzanalyse zwar nicht analysiert, wohl aber klassifiziert.) Vorteilhaft für eine Bearbeitung mit statistischen Verfahren ist, daß es sich um eindeutig intervallskalierte Daten handelt. Man beachte, daß die Variablen zwar gesondert in die Berechnungen eingehen, daß in ihnen aber andere, bekannte Zusammenhänge bzw. Beschreibungsgrößen 'versteckt' sind, von denen man annimmt, daß sie satzmodusrelevant sind: der globale **Deklinationsverlauf** (Differenz von Onset und Offset), der Tonumfang (**Range**, die Differenz zwischen Maximum und Minimum) und der finale hohe Tonverlauf (Offsethöhe). Durch die weiter unten beschriebenen Transformationen erhält man auch Angaben über die **Auslenkung** (monotoner vs. bewegter Tonverlauf) der Äußerung bzw. ihre **Positionierung** im Frequenzbereich bezüglich der sprecherspezifischen Basislinie.

Wir wollen mit unseren Analysen untersuchen, welche Verbesserung der Prädiktion sich durch die im folgenden beschriebenen Maßnahmen erreichen läßt:

Maßnahme 1 (Teilstichproben):

Zum einen sollen alle 956 Realisationen, zum anderen nur die 353 'Prototypen' zugrundegelegt werden. Prototypen nennen wir die Realisationen, die von den Versuchspersonen beim Hörtest im Kontext als natürlich, d.h. auf einer Skala von 1 bis 5 besser als 2.5, eingestuft wurden und die sie kontextfrei den intendierten Modi mit einer Trefferquote von mehr als 80% zuweisen konnten. (Damit sollte ein Filter gebildet werden, mit dem wir weniger akzeptable Realisationen aussondern; vgl. im einzelnen Oppenrieder 1988). Wir nehmen an, daß die Prototypen nicht nur von den Versuchspersonen richtig klassifiziert, sondern auch von der Diskriminanzanalyse besser prädiziert werden können. (Dieser Gesichtspunkt ist für die automatische Spracherkennung relevant, da man ja von einem automatischen Verfahren billigerweise nicht verlangen kann, wozu der Mensch nicht fähig ist: also etwa eine kontextfreie Kategorisierung von Nicht-Prototypen, die im Hörtest nicht dem richtigen Modus zugewiesen wurden.)

Maßnahme 2 (Analysierte vs. klassifizierte Stichprobe):

Wie oben erwähnt, kann man der Diskriminanzanalyse eine unterschiedlich große Auswahl aus der Gesamtstichprobe zur Analyse bzw. Klassifikation vorgeben. Erst dann, wenn man nicht die gesamte Stichprobe analysiert und anschließend klassifiziert, wird 'echt' prädiziert, also die Projektivität des Verfahrens ge-

testet. Es ist z.B. üblich, nur jeden zweiten Fall zu analysieren und dann die andere Hälfte zu klassifizieren. Da es sich bei unserer Stichprobe um sechs verschiedene Sprecher handelt und die Frage der Sprecherabhängigkeit bzw. -unabhängigkeit der Merkmale interessiert, bietet sich ein anderes Vorgehen an:

(1) Es wird reihum ein Sprecher analysiert und dann der Rest der Stichprobe, also fünf Sprecher, klassifiziert (*n-5*). Damit kann abgeschätzt werden, wie gut eine Prädiktion ist, die auf einer sprecherabhängigen Analyse beruht.
(2) Es werden reihum fünf Sprecher analysiert und als Grundlage für die Klassifikation des restlichen Sprechers genommen (*n-1*). Damit wird eine Sprecherunabhängigkeit der Klassifikation simuliert.
(3) Es werden alle sechs Sprecher analysiert und klassifiziert (*n*); damit läßt sich eine obere Grenze der Klassifikationsgüte beschreiben (Vgl. zum Stellenwert dieser Auswahl die Bemerkungen unten in der Diskussion).

Die Annahme, daß von *n-5* über *n-1* nach *n* eine Verbesserung eintritt, ist an sich trivial - es wäre verwunderlich, wenn sie nicht bestätigt würde. Interessant ist aber das Ausmaß der Verbesserung.

Maßnahme 3 (Transformation der Hz-Werte):

Es gibt verschiedene Ansichten darüber, ob Fo-Daten grundsätzlich auf einer Hz- oder einer Halbtonskala bzw. eher als absolute Werte oder bezogen auf einen Vergleichswert (gehörs-)adäquat repräsentiert sind. Bei einem Vergleich von Männer- und Frauenstimmen dürfte man sich aber immer für eine Transformation der Hz-Werte entscheiden. Um herauszufinden, wie sich die beste Prädiktion erzielen läßt, werden wir die Fo-Daten auf sechs unterschiedliche Weisen den Analysen zugrundelegen:

(1) Untransformierte Hz-Werte (Hz)

(2) Hz-Werte transformiert zum sprecherspezifischen Basiswert(Hz_{basis})
Der sprecherspezifische Basiswert ergibt sich aus dem tiefsten, vom jeweiligen Sprecher erreichten Offsetwert. Dieser Wert wird von jedem der vier Fo-Werte abgezogen.

(3) Hz-Werte, transformiert zum äußerungsspezifischen Mittelwert (Hz_{mittel})
Bei Hz_{basis} muß natürlich auf ein Wissen rekurriert werden, das man nicht anhand der jeweils betrachteten Äußerung erhält. Wir nehmen deshalb als äußerungsspezifischen Vergleichswert den Mittelwert der Äußerung an, der sich einfach aus dem Mittelwert der vier zur Verfügung stehenden Werte Onset, Offset, Maximum und Minimum ergibt. (Eine solche Berechnung simuliert sowohl die Situation, in der sich ein Hörer befindet, der eine einzige Äußerung eines ihm

fremden Sprechers beurteilen soll, als auch die Bedingungen eines sprecherunabhängigen automatischen Spracherkennungssystems; vgl. Nöth et al. 1987.)

(4) Halbtonwerte zur Basis 1 (Ht)
Die 'gehörsadäquate' Transformation in Halbtonwerte ergibt sich aus der 12. Wurzel aus zwei multipliziert mit dem natürlichen Logarithmus des Hz-Wertes - anders ausgedrückt, aus der Formel *17.31xLN(Hz)*.

(5) Ht-Werte transformiert zum sprecherspezifischen Basiswert (Ht_{basis})
Die Berechnung ist analog der zu Hz_{basis}.

(6) Ht-Werte, transformiert zum äußerungsspezifischen Mittelwert (Ht_{mittel})
Die Berechnung ist analog der zu Hz_{mittel}.

Maßnahme 4 (Moduskonstellationen):

Für Maßnahme 1-3 haben wir nur intonatorische Merkmale, eben die vier gewählten Fo-Werte, berücksichtigt. Natürlich wäre es möglich, auch syntaktische Merkmale wie Verb-Erst- vs. Verb-Zweit-Stellung, Vorhandensein eines *W*-Wortes etc. als Prädiktorvariablen anzusetzen. Der Status dieser Merkmale ist aber grundsätzlich anders: Es steht z.B. von vornherein fest, daß eine Aussage nicht mit einem *W*-Wort beginnen kann, d.h. wir haben es hier mit einem Merkmal zu tun, das bei unseren Merkmalkombinationen a priori distinktiven Charakter hat. Bei den intonatorischen Parametern wird dagegen die Diskriminanzanalyse als Instrument eingesetzt, mit dem ihre Relevanz für die Modusunterscheidungen erst entdeckt werden soll. Auch wird es sich bei intonatorischen Parametern nicht unbedingt um binäre, sondern oft um graduelle Merkmale handeln, bei denen z.T. nur die Extremwerte eindeutige Indikatoren darstellen: Eine Verb-Erst-Stellung ist entweder vorhanden oder nicht, ein hoher oder tiefer Offset kann mehr oder weniger ausgeprägt sein.

Wir wollen deshalb nicht nur ausschließlich intonatorische Merkmale zugrundelegen, sondern auch nicht-intonatorische. Das entspricht ja auch eher der Situation, in der sich der natürliche Sprecher/Hörer befindet: Man führt ja wohl keine komplette intonatorische Analyse einer Äußerung durch, bei der alle Modi als gleichermaßen wahrscheinlich angesetzt werden, wenn z.B. ein einleitendes *W*-Element die Äußerung von vornherein formal als Nicht-Aussagesatz und Nicht-Imperativsatz kennzeichnet. Natürlich soll und kann eine klassifizierende statistische Analyse kein Hörermodell darstellen, es ist aber doch sinnvoll, der Diskriminanzanalyse jeweils pro Fall nur die Modi zur Auswahl vorzulegen, denen

die Äußerung aufgrund von nicht-intonatorischen Merkmalen auch wirklich zugerechnet werden kann. (Auch ein automatisches Verfahren der Spracherkennung ist grundsätzlich in der Lage, z.B. ein einleitendes *W*-Element zu erkennen und damit Aussagesatz und Imperativsatz als Formtyp auszuschließen.) Als Möglichkeit in unserem Korpus ausgeschlossen sind durch **Verb-Erst-Stellung** Aussagesatz, durch **Verb-Zweit-Stellung** Wunschsatz und Imperativsatz, durch **Konjunktiv** II Imperativsatz, durch ein einleitendes **W-Element** Aussagesatz, Imperativsatz und Wunschsatz. Als letztes Merkmal wurden die **Modalpartikeln** berücksichtigt, die auch nur jeweils bestimmte Modi indizieren; *wohl* schließt z.B. einen Exklamativsatz aus, *etwa* einen Wunschsatz etc. Jeder Satz unseres Korpus' wurde aufgrund dieser Merkmale danach klassifiziert, welche Modi mit ihm **nicht** ausgedrückt werden können.

In unserem Korpus gibt es vier mögliche **Modus- (Minimalpaar- oder Tripel-) Konstellationen**; in Klammern ist jeweils die Zahl der Fälle im ganzen Korpus und bei den Prototypen angegeben:

(1) **Fragesatz vs. Exklamativsatz**, z.B. *Hat der geflucht* (390/157)
(2) **Fragesatz vs. Exklamativsatz vs. Imperativsatz**, z.B. *Stellt ihr euch an* (82/43)
(3) **Fragesatz vs. Aussagesatz vs. Exklamativsatz**, z.B. *Die ist naiv* (145/45)
(4) **Fragesatz vs. Exklamativsatz vs. Wunschsatz**, z.B. *Wäre ich glücklich* (67/12).

Hinzu kommen Fälle, bei denen jeweils nur ein einziger Modus indiziert wird; so kann der Satz *Stellt ihr euch etwa an* wegen der Modalpartikel *etwa* nur als Frage aufgefaßt werden.

Jede der Äußerungen wurde nun einer der vier Konstellationen zugeteilt oder für diesen Bearbeitungsschritt ausgeschieden. Für jede Konstellation wurden analog zur alleinigen Berücksichtigung der intonatorischen Parameter Diskriminanzanalysen durchgeführt.

5. ERGEBNISSE

Die Ergebnisse aller durchgeführten Diskriminanzanalysen sind in Tab.1 und 2 verzeichnet.

Tab.1 : Richtig klassifizierte Fälle in Prozent
(Alle Realisationen)

Werte	n-5		n-1		n	
Hz	40.4	*51.1*	53.4 +	*67.7*	55.6	*73.2*
Hz_{basis}	45.0	*61.8*	54.8 +	*68.3*	57.3 +	*74.4*
Hz_{mittel}	44.8	*60.5*	51.1	*70.2*	54.3	*75.0*
Ht	41.7	*51.1*	50.2	*70.2*	55.7	*79.6* +
Ht_{basis}	49.7 +	*66.5* +	53.1 +	*72.6* +	58.6 +	*80.8* +
Ht_{mittel}	51.5 +	*65.9* +	53.0 +	*73.6* +	56.6 +	*80.1* +

Tab.2 : Richtig klassifizierte Fälle in Prozent
(Prototypen)

Werte	n-5		n-1		n	
Hz	48.6	*58.0*	59.2	*75.1*	66.3	*89.5*
Hz_{basis}	58.4	*66.5*	61.3	*79.4*	69.1 +	*87.9*
Hz_{mittel}	58.4	*70.9*	59.1	*77.6*	68.3 +	*87.1*
Ht	52.1	*57.3*	64.2 +	*85.2* +	68.8 +	*91.4* +
Ht_{basis}	61.2 +	*72.0*	66.1 +	*83.9* +	68.5 +	*92.4* +
Ht_{mittel}	61.8 +	*79.5* +	65.7 +	*83.6* +	68.8 +	*92.0* +

Legende zu Tab.1 und 2:
kursiv: zusätzliche Berücksichtigung der möglichen Satzmodi.
analysierte vs. klassifizierte Stichproben ('Lern'- vs. 'Prüf'-Stichprobe):
n-5: reihum ein Sprecher analysiert, fünf klassifiziert
n-1: reihum fünf Sprecher analysiert, einer klassifiziert
n: alle Sprecher analysiert und klassifiziert.
'+': beste Prädiktionen pro Spalte (Bereich: maximaler Prozentwert-2%).

In Tab.1 stehen die Werte für alle Realisationen, in Tab.2 nur die der Prototypen (vgl. oben Maßnahme 1); ansonsten ist der Aufbau der beiden Tabellen identisch. In der ersten Spalte ist angegeben, welche Art der Transformation den Zeilenwerten zugrundeliegt (vgl. oben Maßnahme 3). Jeweils zwei Spalten stehen für analysierte vs. klassifizierte Stichprobe (vgl. oben Maßnahme 2). Bei den nicht-kursiven Werten wurden nur die intonatorischen Variablen zugrundegelegt, bei den kursiven wurden nur die jeweils einer Moduskonstellation angehörigen Fälle gemeinsam berechnet (vgl. oben Maßnahme 4). Die Werte stehen in der linken (nicht-kursiv gekennzeichneten) Spalte von *n* für eine einzige Diskrimi-

nanzanalyse; ansonsten stellen sie Mittelwerte dar. So repräsentiert z.B. bei *n-1*, kursiv gekennzeichnete Spalte, jeder Wert einen Mittelwert aus 24 unterschiedlichen Diskriminanzanalysen (sechs Sprecher reihum bei der Analyse ausgelassen, pro Sprecher vier Moduskonstellationen). Die Mittelwerte sind **ungewichtet**, d.h. es wurde nicht berücksichtigt, daß z.B. die erste Moduskonstellation mehr Fälle repräsentiert als die anderen. Da aber nie gewichtet wurde, sind alle Werte untereinander vergleichbar. Um einen schnellen Überblick zu ermöglichen, sind die besten Werte pro Spalte (im Bereich 'maximaler Prozentwert-2%') mit '+' gekennzeichnet.

Wir fassen nun zuerst zusammen, in welchem Ausmaß die von uns vorgenommenen Maßnahmen die Prädiktion verbessern konnten. (Die dabei angegebenen Verbesserungen sind nicht genau aus den Werten von Tab.1 und 2 berechnet, sondern geben einen abgerundeten Betrag der Verbesserung wieder.)

Maßnahme 1 (Teilstichproben):

Die Prototypen können um gut 10% besser klassifiziert werden als alle anderen Realisationen zusammen. (Eine Klassifikation der Nicht-Prototypen, also der in unserem Sinne ungenügenden Realisationen, ergab eine etwa 20% schlechtere Prädiktion als die der Prototypen. Der Unterschied ist also systematisch und nicht etwa durch die geringere Anzahl der Prototypen im Verhältnis zu allen Realisationen bedingt.)

Maßnahme 2 (Analysierte vs. klassifizierte Stichprobe):

Von *n-5* über *n-1* nach *n* verbessert sich die Prädiktion um ca. 30%. (Bei *n-5* und *n-1* sind die Werte ein genaues Maß der Effektivität der Diskriminanzfunktion; d.h. wenn man die Funktion auf eine neue Stichprobe anwendet, so wird die Prädiktion die gleiche Güte haben. Bei *n* dagegen sind die Werte mit einem Bias behaftet und gelten deshalb nur für die vorliegende Stichprobe.)

Maßnahme 3 (Transformation der Hz-Werte):

Halbtonwerte resultieren in einer um ca. 5% besseren Prädiktion als Hz-Werte. Am besten ist die Prädiktion bei den transformierten Halbtonwerten, wobei es unerheblich ist, ob die Transformation sprecherspezifisch (zum Basiswert) oder äußerungsspezifisch (zum Mittelwert) erfolgte: Die Differenzen sind mit einer

Ausnahme geringer als 2%. Beide Transformationen sind also ein geeignetes Mittel, die unterschiedlichen Sprechlagen von Frauen und Männern vergleichbar zu machen. Am wichtigsten ist eine solche Transformation offensichtlich dann, wenn nur auf der Grundlage eines Sprechers oder einer Sprecherin prädiziert wird: bei *n-5* ist der Unterschied zwischen transformierten und nicht-transformierten Werten am höchsten.

Maßnahme 4 (Moduskonstellationen):

Bei Berücksichtigung der möglichen Moduskonstellationen verbessert sich die Prädiktion um gut 20%. Man beachte, daß sich dabei auch der Erwartungswert geändert hat; er ist nun - gewichtet aus den vier verschiedenen Moduskonstellationen und der jeweiligen Anzahl der Fälle - ca. 43%, während er sonst 20% beträgt. Ohne eine vergleichende Studie mit einem Modell, das unsere verschiedenen Konstellationen mit Zufallszahlen simuliert, kann prima facie nicht entschieden werden, in welchem Ausmaß die Verbesserung allein auf die Erhöhung des Erwartungswertes oder darauf zurückzuführen ist, daß die stärkere funktionale Belastung der Intonation in solchen Fällen eindeutigere und leichter zu klassifizierende Fo-Werte zur Folge hat.

6. DAS FÄHNLEIN DER SIEBEN AUFRECHTEN

Vor der abschließenden Diskussion wollen wir nun die Fälle betrachten, die auch bei der besten Prädiktion von 92.4% (vgl. Tab.2, 6. Spalte, vorletzter Wert) noch fehlklassifiziert wurden. Da es sich dabei um Prototypen handelt - also um Fälle, die von den Hörern richtig klassifiziert und als natürlich bewertet wurden -, ist anzunehmen, daß dabei von uns nicht berücksichtigte Merkmale eine Rolle spielen. Tab.3 zeigt diese Fehlklassifikationen; in der ersten Spalte ist dabei die betreffende Moduskonstellation vezeichnet (vgl. Maßnahme 4).

Tab.3: Fehlklassifikationen:

Konst.	Satz	Anzahl Fälle	tatsächl. Modus	zugewiesener Modus
1	*Gehört das Ihnen hier* *	4	Frage	Exkl.
1	*Wie ist der reich geworden* *	10	Frage	Exkl.
1	*Wie laut ist es hier* °	2	Frage	Exkl.
1	*Stellt ihr euch vielleicht an* °	1	Exkl.	Frage
1	*Hat der geflucht* °	1	Exkl.	Frage
2	*Stellt ihr euch an* °	3	Imp.	Exkl.
3	*Er sieht was* *	1	Aussage	Exkl.
3	*Du kommst* *	2	Aussage	Exkl.
3	*Die ist naiv*	2	Exkl.	Aussage

Für die mit einem Stern gekennzeichneten Sätze lassen sich leicht zusätzliche Merkmale finden, mit deren Hilfe der richtige Modus klassifiziert werden kann: Die Verbsemantik schließt bei *Gehört das ihnen hier* eine Exklamativinterpretation aus, ebenso bei *Er sieht was* und *Du kommst*. Die Realisationen von *Wie ist der reich geworden* haben alle, durch die Kontextvorgabe bedingt, den Satzakzent auf dem *W*-Wort - eine Akzentuierung, die ebenfalls eine Exklamativinterpretation ausschließt. Es ist sicher kein Zufall, daß bei allen diesen Fehlklassifikationen der Exklamativ mit beteiligt ist - ein Modus, dessen Status grundsätzlich nicht ganz eindeutig ist (vgl. Batliner 1988). Als von uns nicht berücksichtigte exklamativtypische Merkmale kommen zumindest Dehnung und die Position des Fo-Gipfels in Betracht. Nimmt man an, daß die Verwechslung von Exklamativ und Aussage bei *Die ist naiv* keine kommunikativen Schwierigkeiten nach sich ziehen würde, so bleiben sieben, in Tab.3 mit '°' gekennzeichnete 'kritische' Fehlklassifikationen übrig - kritisch insofern, als solche Verwechslungen der Intention des Sprechers beim Adressaten unadäquate Reaktionen hervorrufen. Das sind weniger als 3% der Stichprobe.

7. DISKUSSION

Linguistische oder phonetische Analysen validieren ihre Regeln meistens am aktuellen Korpus - die Überprüfung, ob diese Regeln auch auf andere Korpora, also andere Stichproben aus der Population, zutreffen, bleibt dann nachfolgenden Arbeiten überlassen. Insofern ist die zuletzt genannte Fehlerquote von weniger als 3% (die ja auch nur auf unsere Stichprobe zutrifft und bei anderen Stichproben höher ausfallen würde) vergleichbar und ohne 'doppelten Boden' erzielt.

Die Behauptung, damit sei ein Verfahren entdeckt, mit dem man in 'real life'-Situationen 97% der vom Sprecher intendierten Satzmodi richtig klassifizieren kann, wäre natürlich unsinnig. Wir möchten deshalb nun anhand einiger relevanter Punkte den Stellenwert unserer Ergebnisse diskutieren:

Sprecherunabhängigkeit: Eine Sprecherunabhängigkeit, so wie sie billigerweise bei automatischen Verfahren verlangt oder zumindest angestrebt werden sollte, ist bei unseren Berechnungen im Falle von *n-1* simuliert. Das beste Ergebnis ist hier 73.6% für alle Realisationen. Man muß dabei allerdings bedenken, daß bei weitem nicht alle relevanten intonatorischen oder nicht-intonatorischen Merkmale in Betracht gezogen oder als Prädiktorvariablen optimiert wurden.

Produktionsbedingungen: Die Bedingungen, unter denen Korpora zustandekommen, sind ein wesentlicher Faktor für die Konsistenz dieser Korpora und damit für die Möglichkeit, Regeln aufzustellen und die Fälle erfolgreich zu klassifizieren. Auf der einen Seite stehen Korpora, bei denen erfahrene und kooperative Sprecher (z.B. die Autoren selbst) unter genau und explizit festgelegten Bedingungen die Äußerungen produzieren. Auf der anderen Seite stehen Aufnahmesituationen, die dem Ideal einer unbeobachteten und zwanglosen Kommunikation möglichst nahe kommen. Unser Korpus ist hier irgendwo dazwischen einzuordnen (vgl. zu den Aufnahmebedingungen im einzelnen Oppenrieder 1988): Einerseits werden 'naive' Sprecher nicht explizit (z.B. durch Vorgabe der Akzentposition, der Emphasestufe etc.) instruiert, wie sie produzieren sollen, sondern sie sollen nur durch eine Kontextvorgabe zur intendierten Produktion veranlaßt werden. Andererseits stehen die spezielle Minimalpaarkonstruktion sowie die Aufnahmebedingungen im schallarmen Raum 'ganz natürlich' produzierten Äußerungen entgegen. U.W. gibt es kein Maß, mit dem man den Grad der Natürlichkeit bestimmen könnte. Es gibt aber für unser Korpus doch einen gewissen Anhaltspunkt durch die Hörtests und dadurch, daß die mit deren Hilfe ausgewählten Prototypen um etwa 10% besser klassifiziert werden als alle Realisationen.

Diskursspezifität: Unser Korpus weist zwei Eigenheiten auf, die wahrscheinlich eine richtige Klassifikation anhand der von uns gewählten Prädiktorvariablen begünstigen: Zum einen sind die Äußerungen relativ kurz. Bei längeren Sätzen, die aus mehreren intonatorischen Phrasen mit jeweils einem Maximum und Minimum bestehen, dürften unsere vier Fo-Variablen allein keine so gute Prädiktion mehr gewährleisten. Zum anderen sind Fragen überrepräsentiert (gut die Hälfte der Fälle). Diese zwei Eigenheiten weisen aber auch restringierte Mensch-Maschine-Kommunikationen auf, z.B. Bahnauskünfte auf Fragen wie *Wann geht*

der Sonderzug nach Pankow? Insofern ist unser Korpus zwar nicht für die Gesamtpopulation, aber doch für bestimmte Anwendungen repräsentativ.

Prädiktorvariablen: Das stabilste intonatorische Merkmal dürfte die Offsethöhe sein, die sehr oft Fragen von Nicht-Fragen unterscheidet (vgl. Oppenrieder 1988). Da die Fragen in unserem Korpus so häufig vorkommen, ergibt der Offset allein als Prädiktorvariable eine nur um ca. 5% geringere Prädiktionsgüte (bei den besten Werten in Tab.1 und 2 z.B. 73.8 statt 80.8 und 88.8 statt 92.4). Das heißt aber nicht, daß der Offset ausreicht. Zum einen ist im Bereich von über 80% eine Verbesserung um 5% nicht zu verachten; sie zeigt, daß auch die anderen Variablen relevant sind. Zum anderen differenziert der Offset zwar gut zwischen Fragen und Nicht-Fragen, aber nicht innerhalb der Nicht-Fragen. Es ist deshalb nötig, an anderen Korpora zu untersuchen, welche weiteren Merkmale als Prädiktorvariablen in Frage kommen (vgl. dazu Batliner 1987b), um die Auswahl der Prädiktoren optimieren zu können.

8. SCHLUSSBEMERKUNGEN

Wie wir schon erwähnt haben, stand eine Optimierung der Prädiktorvariablen nicht im Mittelpunkt unserer Analysen, sondern die Frage, welche Verbesserung sich mit den von uns vorgenommenen vier Maßnahmen erzielen läßt. Wir wollen zwei Punkte zum Schluß noch einmal kommentieren:

(1) Die (plausible) Annahme, daß zu einem Vergleichswert transformierte Halbton-Werte die besten Ergebnisse bringen, hat sich voll bestätigt. Dies ist ein Beitrag zur Methodik.

(2) Bedenkt man, daß unser Korpus von seiner Konstruktion her für Verwechslungen prädestiniert ist, und daß unsere Prädiktorvariablen alles andere als optimiert waren, so läßt die erzielte Prädiktionsgüte hoffen. Dies ist ein Beitrag zur Frage, welche Aussicht eine automatische Bestimmung des Satzmodus in Zukunft haben dürfte.

LITERATUR:

Altmann, Hans (1987): Zur Problematik der Konstitution von Satzmodi als Formtypen. In: Meibauer, Jörg (Hg.): Satzmodus zwischen Grammatik und Pragmatik. Tübingen, Niemeyer, 22-56. (=Linguistische Arbeiten 180).

Batliner (1987a): Der Einsatz der Diskriminanzanalyse zur Prädiktion des Satzmodus. In: Tillmann, Hans G. / Willée, Gerd (Hgg.): Analyse und Synthese gesprochener Sprache. Hildesheim et al., Olms, 125-132.

Batliner, Anton (1987b): Die intonatorische Indizierung des Fokus. Erste Ergebnisse zur Perzeption und Produktion. Ms.

Batliner, Anton (1988): Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? Experimente zur Rolle von Höhe und Position des Fo-Gipfels. (In diesem Band).

Oppenrieder, Wilhelm (1988): Intonatorische Kennzeichnung von Satzmodi. (In diesem Band).

Nöth, Elmar / Batliner, Anton / Lang, Roswitha / Oppenrieder, Wilhelm (1987): Automatische Grundfrequenzanalyse und Satzmodusdifferenzierung. In: Tillmann, Hans G. / Willée, Gerd (Hgg.): Analyse und Synthese gesprochener Sprache. Hildesheim et al., Olms, 59-66.