Hans G. Tillmann, Gerd Willée (Hg.)

Analyse und Synthese gesprochener Sprache

Vorträge im Rahmen der
Jahrestagung 1987 der Gesellschaft für
Linguistische Datenverarbeitung e. V.,
Bonn, 4.–6. März 1987

1987
Georg Olms Verlag Hildesheim · Zürich · New York
Gesellschaft für Linguistische Datenverarbeitung e. V.

Anton Batliner

Institut für Deutsche Philologie
Ludwig-Maximilians-Universität München
Schellingstr. 3, 8000 München 40

DER EINSATZ DER DISKRIMINANZANALYSE ZUR PRÄDIKTION DES SATZMODUS.

1. Satzmodussystem und Testsatzkorpus

Der vorliegende Beitrag entstand im Rahmen des DFG-Projekts "Modus-Fokus-Intonation", das eine genauere Charakterisierung der intonatorischen Merkmale von Satzmodus und Fokus zum Ziel hat. Das im folgenden beschriebene Testsatzkorpus wurde auf dem Hintergrund eines von Hans Altmann entwickelten Satzmodussystems zusammengestellt, vgl. dazu Altmann (1984) und Altmann (1987). 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-Ausdruck; b) die Stellungseigenschaften dieser satzmodusrelevanten Ausdrücke, z.B. die Zweit-Stellung 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 durch Merkmale aus den Bereichen a) bis c) unterscheiden, lassen sich demnach die satzmodusrelevanten intonatorischen Parameter(-werte) auffinden und untersuchen. 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 6 Versuchspersonen (3 weibl., 3 männl.) vorgelegt und von diesen mindestens zweimal realisiert; dabei ergaben sich 956 Äußerungen. Von jeder Äußerung wurde ein Mingogramm mit Zeitsignal, Fo-Kurve und Intensitäts-

kurve 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 SPSS, 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 inferenzstatistisch überprüft wird. Ergänzend zu diesem Vorgehen wurde aber auch ein geeignetes klassifizierendes Verfahren, die sog. Diskriminanzanalyse, auf das Korpus angewandt. Es soll 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 bieten 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.

2. Beschreibung des Verfahrens

Die Diskriminanzanalyse ist ein Verfahren, bei dem, basierend auf relevanten Variablen (in unserem Fall Onset, Offset etc.), die einzelnen Fälle (in unserem Fall: Äußerungen unseres Korpus') distinkten Kategorien (in unserem Fall: intendierten Modi) zugewiesen werden. Das Verfahren funktioniert ähnlich wie die multiple Regression: Es werden lineare Kombinationen der unabhängigen "Prädiktor"-Variablen gebildet, die möglichst optimal zwischen den Kategorien unterscheiden können. Jede Prädiktorvariable erhält einen Gewichtungskoeffizienten, der anhand der 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 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 klassifi-

ziert werden, es kann aber auch nur ein Ausschnitt aus dem Material (die "Lernstichprobe") analysiert werden und dann Grundlage der Klassifikation eines anderen Ausschnitts (der "Prüfstichprobe") werden.

Es wurden bisher Analysen durchgeführt, bei denen die Äußerungen in unserem Korpus einem der fünf Haupt-Modi Frage, Aussage, Exklamativ, Imperativ oder Wunsch zuzuordnen waren. (In einem späteren Stadium soll das Verfahren auch auf Untergruppen innerhalb eines Modus angewendet werden, z.B. auf W-Versicherungsfrage vs. Ergänzungsfrage, normale Aussage vs. Aussage mit Kontrastakzent o.ä.; vgl. dazu auch Nöth et al. (1987) in diesem Band. Hauptsächlich wurden die kontinuierlichen Grundfrequenz-Variablen Onset, Offset, Maximum und Minimum als Prädiktorvariablen gewählt, da sie 1) intervallskaliert und deswegen gut als Prädiktorvariablen geeignet sind und sich 2) auch relativ einfach mit einem automatischen Verfahren extrahieren lassen dürften. Bevor wir auf die Ergebnisse eingehen, seien einige wichtige Punkte erwähnt:

(i) Maßstab für die Güte der Prädiktion ist der Prozentwert der Fälle, die mit den jeweils angesetzten Prädiktorvariablen richtig klassifiziert, d.h. dem intendierten Modus zugeschlagen 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% richtige Zuordnungen ergeben, bei den fünf Haupt-Modi, wie in unserem Fall, ergäbe sie 20% richtige Zuordnungen.

(ii) 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 die Wahrscheinlichkeit von 20%. Wenn die Verteilung der Gruppen in der Stichprobe der Verteilung in der Population (der Grundgesamtheit) entspricht und von der Gleichverteilung abweicht, so kann die Auftretenswahrscheinlichkeit der einzelnen Gruppen vorgegeben und damit meist eine bessere Prädiktion erzielt werden, da die Auftretenswahrscheinlichkeit in die Berechnung der Prädiktion mit eingeht. Was die Verteilung der Modi in der Grundgesamtheit betrifft, so sind Aussagen darüber problematisch: Über die Grundgesamtheit aller in einem bestimmten Zeitraum im deutschen Sprachraum realisierten Modi kann man nur spekulieren; es kann aber als sicher angenommen werden, daß diskursspezifische Stichproben Ungleichverteilungen aufweisen. So werden beim Militär relativ gesehen mehr Imperative realisiert werden, im Reisebüro mehr Fragen usw. So gesehen, ist die Ungleichverteilung in unserem Korpus, die sich aus

der speziellen Form der Minimalpaarauswahl ergibt, gar nicht so ungewöhnlich. Eine Berücksichtigung der Ungleichverteilung führt auch für unser Korpus immer zu einem besseren Ergebnis; darauf werden wir allerdings in diesem Beitrag kaum eingehen. Die Zugrundelegung aller möglichen Minimalpaare hat nämlich in unserem Korpus zur Folge, daß Fragen stark überwiegen - und eine Berücksichtigung dieses Umstandes führt dazu, daß dieser Modus auf Kosten der anderen Modi besser klassifiziert wird.

(iii) Bei jeweils nur einer Variablen als Prädiktor ist eine Interpretation noch relativ einfach. Sie wird schwieriger bei mehreren Variablen, da die Berücksichtigung einer zusätzlichen Variablen, von der man schon weiß, daß sie für bestimmte Distinktionen relevant ist, nicht automatisch einen besseren Klassifikationswert ergibt, wenn alle Fälle und alle Gruppen klassifiziert werden müssen. Wir beschränken uns der Einfachheit halber in der folgenden Darstellung auf eine Kombination der vier intervallskalierten und leicht zu extrahierenden Fo-Variablen Offset, Onset, Maximum und Minimum als Prädiktorvariablen.

3. Ergebnisse

Tabelle 1 zeigt die richtig klassifizierten Fälle in Prozent, wobei zum einen alle 956 Fälle, zum anderen nur die 353 "Prototypen" zugrundelagen. "Prototypen" nennen wir die Realisationen, die von unseren 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 Sicherheit von mehr als 80% zuweisen konnten. Die Prädiktorvariablen sind 1) transformiert in Halbtöne zur Basis 1 (Ht); 2) sind die Halbtonwerte transformiert zum sprecherspezifischen Basiswert (Ht_s), d.h. von jedem Wert wird der vom jeweiligen Sprecher tiefste erreichte Wert abgezogen ; 3) sind die Halbtonwerte transformiert zu einem angenäherten Mittelwert der jeweiligen Äußerung (Ht_a), der sich aus dem Mittel von Onset, Offset, Maximum und Minimum ergibt. Die Annahme, daß eine Transformation zum sprecherspezifischen Basiswert die Prädiktion verbessert, da damit Frauen- und Männerstimmen eher vergleichbar werden, bestätigt sich. Allerdings ergibt auch die Transformation zum Mittelwert der Äußerung kaum schlechtere Klassifikationen. (Dieses Ergebnis bestätigt sich im folgenden; vgl. auch Nöth et al. (1987) in diesem Band.)

Es wurde mit drei verschiedenen Konstellationen von Lern- und Prüfstichprobe gerechnet: 1) wird reihum ein Sprecher analysiert ($n-5$) und dient als Grundlage der Klassifikation der übrigen fünf Sprecher. Damit kann eine Prototypizität von einzelnen Sprechern abgeschätzt werden. 2) werden reihum fünf Sprecher analysiert und

dienen als Grundlage für die Klassifikation des restlichen Sprechers (*n-1*): damit wird eine Sprecherunabhängigkeit der Klassifikation simuliert. 3) werden alle sechs Sprecher analysiert und klassifiziert (*n*); damit dürfte eine obere Grenze der Klassifikationsgüte beschrieben werden können. Es bestätigt sich die Annahme, daß von 1) zu 3) eine schrittweise Verbesserung eintritt. Weiter bestätigt sich, daß die Prototypen wirklich eine in sich konsistentere Gruppe darstellen als alle Fälle, die ja auch Sprecheridiosynkrasien, echte Fehlrealisationen und vermehrt untypische Realisationen enthalten.

Tabelle 1: fünf Haupt-Modi, Erwartungswert 20%

	Alle 956 Fälle			353 Prototypen		
Präd. var.	n-5	n-1	n	n-5	n-1	n
Ht	41.70	50.16	55.69	52.12	64.17	68.82
Ht_b	49.71	53.15	58.59	61.18	66.07	68.53
Ht_e	51.53	53.02	56.58	61.84	65.68	68.82

Die bisher dargestellten Analysen haben nur rein intonatorische Variablen berücksichtigt. Natürlich wäre es möglich, auch syntaktische Merkmale wie Verb-Erst- vs. Verb-Zweit-Stellung, Vorhandensein eines W-Worts etc. als Prädiktorvariablen anzusetzen. Ihr Charakter 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 Minimalpaarkonstellationen a priori distinktiven Charakter hat. Für die intonatorischen Parameter hingegen wird die Diskriminanzanalyse gerade als Instrument eingesetzt, mit dem ein eventueller distinktiver Charakter erst entdeckt werden soll. Auch wird es sich bei diesen Parametern grundsätzlich nicht um binäre, sondern um graduelle Distinktionen handeln, bei denen z.T. nur die Extremwerte eindeutige Indikatoren darstellen können: eine Verb-Erst-Stellung ist entweder vorhanden oder nicht, ein hoher Offset kann mehr oder weniger ausgeprägt sein.

Wir wollen deshalb nun eine andere Vorgehensweise beschreiben, die auch eher der Wahrnehmung durch den natürlichen Sprecher/Hörer entspricht, da auch nicht-intonatorische Merkmale mit berücksichtigt werden. Der Hörer 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 als Nicht-Aussage und Nicht-Imperativ kennzeichnet. Natürlich soll eine klassifizierende statistische Analyse kein Hörermodell darstellen, es ist

aber doch sinnvoll, der Diskriminanzanalyse jeweils pro Fall nur die Modi zur Auswahl vorzugeben, denen die Äußerung auf Grund der nicht-intonatorischen Merkmale auch wirklich zugerechnet werden könnte. Auch ein automatisches Verfahren der Spracherkennung kann prinzipiell z.B. ein einleitendes W-Element erkennen und damit Aussagesatz und Imperativsatz als Formtyp ausschließen. Als Möglichkeit ausgeschlossen sind durch Verb-Erst-Stellung Aussagesatz, durch Verb-Zweit-Stellung Wunschsatz und Imperativsatz, durch Konjunktiv II Imperativsatz, durch ein einleitendes W-Wort 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 **Minimalpaar- oder Minimaltripel-Konstellationen**; in Klammern ist jeweils die Zahl der Fälle im ganzen Korpus und bei den Prototypen angegeben:
1) **Fragesatz vs. Exklamativsatz,** z.B. *Wie der läuft* (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 sein kann; 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 die weitere Bearbeitung ausgeschieden. Für jede Konstellation wurde eine Diskriminanzanalyse mit den gleichen Vorgaben wie beim oben beschriebenen ersten Vorgehen durchgeführt.

Tabelle 2: je eine Moduskonst., Erwartungswert 50% bzw. 33.33%

		Alle 956 Fälle		353 Prototypen	
Präd. var.	Konst.	n-1	n	n-1	n
Ht.	1	86.00	87.95	82.90	87.90
Ht.	2	64.46	78.05	75.33	93.02
Ht.	3	69.18	76.55	85.88	88.89
Ht.	4	70.94	80.60	91.66	100.00
Ht.	1-4	72.64	80.79	83.94	92.45
Ht.	1	86.56	87.44	83.43	85.99
Ht.	2	66.69	76.83	72.06	93.02
Ht.	3	67.29	77.24	87.27	88.89
Ht.	4	73.71	79.10	91.66	100.00
Ht.	1-4	73.56	80.15	83.60	91.97

Die Legende zu Tabelle 2 ist analog der zu Tabelle 1. Die fünfte und die zehnte Zeile zeigen jeweils die Mittelwerte für alle vier Konstellationen. Für die beste Prädiktion (92.45% vgl. fünfte Zeile bei den Prototypen) sind in Tabelle 3 die Fehlklassifikationen aufgeführt; die erste Spalte zeigt die jeweilige Konstellation.

Tabelle 3: Fehlklassifikationen

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

Bei allen Fehlklassifikationen ist der Exklamativ mit beteiligt, der sich auch in den anderen von uns durchgeführten Untersuchungen als wenig trennscharf herausgestellt hat. Für die mit einem Stern gekennzeichneten Sätze lassen sich zusätzliche Merkmale finden, mit deren Hilfe sich der richtige Modus klassifizieren läßt: Die Verbsemantik schließt bei *Gehört das Ihnen hier* eine Exklamativinterpretation aus, ebenso wie 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 bleiben also 8 Fehlklassifikationen übrig.

Wie in Tabelle 1 zeigt sich auch in Tabelle 2 kein entscheidender Unterschied zwischen der Transformation zum Basiswert und der zum Mittelwert. Die gemittelte Trefferquote ist bei den Prototypen für Ht_b bei n-1 83.94 und bei n 92.45. Wenn man die in Tabelle 3 mit einem Stern gekennzeichneten Fälle, die sich prima vista mit den zusätzlichen Merkmalen erklären lassen, hinzunimmt, so verbessert sich diese letzte Trefferquote von 92.45% auf ca 97%.

4. Schlußbemerkungen

Bedenkt man, daß unsere Stichprobe aus sechs verschiedenen Sprechern bestand und von ihrer Konstruktion her für Verwechslungen prädestiniert ist, so ist eine Trefferquote von 92.45% bzw. 97% ein sehr gutes Ergebnis. Es ist auch insoweit realistisch, als

die übrig gebliebenen Fehlklassifikationen alle die weniger ausgeprägte Kategorie 'Exklamativ' betreffen, mit der auch unsere Versuchspersonen im Kategorisierungstest, bei dem sie die kontextfreien Äußerungen den intendierten Modi zuordnen mußten, die größten Schwierigkeiten hatten. Auf der anderen Seite müssen zumindest die folgenden Gesichtspunkte zusätzlich erwogen werden:

1. Eine Trennung in Lern- und Prüfstichprobe (Simulierung von Sprecherunabhängigkeit) ergibt z.B. für n-1 vs. n eine um 5-10% schlechtere Trefferquote.
2. Im "Ernstfall" können Fehlrealisationen und untypische Verläufe nicht wie bei unseren Prototypen ausgefiltert werden. Es zeigte sich denn auch, daß das gesamte Korpus um ca. 10% schlechter klassifiziert wurde.
3. Die Sätze in unserer Stichprobe waren alle relativ kurz. Bei längeren Sätzen, die aus mehreren Phrasen mit jeweils einem Maximum und einem Minimum bestehen, dürften die Variablen Onset, Offset, Maximum und Minimum allein keine so gute Prädiktion mehr gewährleisten.
4. Allerdings steht eine Berücksichtigung anderer relevanter Faktoren, die die Prädiktion verbessern könnten, noch aus, so etwa die von Intensitäts- oder Dauerparametern.

Zu diesen Punkten sind in unserem Projekt noch weitere Untersuchungen geplant.

LITERATUR:

Altmann, Hans (1984): Linguistische Aspekte der Intonation am Beispiel Satzmodus. In: Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 19, 130-152.

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

Nöth, Elmar / Batliner, Anton / Lang, Roswitha / Oppenrieder, Wilhelm (1987): Automatische Grundfrequenzanalysen und Satzmodusdifferenzierung. In: H.G. Tillmann, G. Willée (Hrsg.): Analyse und Synthese gesprochener Sprache. Vorträge im Rahmen der Jahrestagung 1987 der Gesellschaft für Linguistische Datenverarbeitung e.V., Bonn, 4.-6. März 1987. (Dieser Band)