

## "Neue Information" im Sprachsignal: Die prosodische Markierung der Fokusstruktur

A. BATLINER, W. OPPENRIEDER, E. NÖTH\*, G. STALLWITZ\*

(Institut für Deutsche Philologie, Universität München;

\* Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg)

### 1. Fragestellung

Der linguistische Begriff der Fokussierung bezieht sich auf einen Aspekt der semantischen Gliederung von Sätzen. Sätze werden gegliedert in einen Fokus- und einen dazu komplementären Hintergrundteil. Die im Fokus stehende Information wird aus einer Menge von Alternativen ausgewählt, die im jeweiligen Verwendungskontext zur Debatte stehen; beim Hintergrund dagegen fehlt ein Bezug auf Alternativen [5]. Im allgemeinen enthält der Fokus relativ zur Verwendungssituation 'neue', der Hintergrund dagegen 'alte', kontextuell 'gegebene' Information. Die angemessene Fokus-Hintergrund-Gliederung eines Satzes ist also eine Funktion des sprachlichen und nicht-sprachlichen Kontextes, in dem dieser Satz verwendet wird. Eine klar definierte Abgrenzung von Fokus und Hintergrund ergibt sich z.B. in Aussagesätzen, die als Antworten auf *w*-Fragen verwendet werden: Nur derjenige Ausdruck, der die erfragte *w*-Konstituente ersetzt, ist fokussiert, d.h. aus einer Anzahl von Alternativen ausgewählt, während der Hintergrund durch den Rest der *w*-Frage kontextuell eindeutig bestimmt ist. Ein Beispiel zeigt (1):

- (1) **Situationsbeschreibung:** In einem Textilbetrieb; eine Mutter erkundigt sich bei einer Angestellten nach den handwerklichen Fortschritten ihrer Tochter.

**Mutter:** *Was läßt die Meisterin meine Nina denn gerade weben?* (Kontextsatz)

**Angestellte:** *Sie läßt die Nina das Leinen weben.*

In der Antwort ist die Konstituente *das Leinen* fokussiert (**enger Fokus**). Fokussierung wird auf der phonetischen Oberfläche durch die akzentuelle Hervorhebung mindestens einer Silbe in der Fokuskonstituente angezeigt; die Auswahl dieser Silbe erfolgt nach spezifischen Regeln, die wesentlich von der syntaktischen Struktur des Satzes abhängen [7, S.212ff.]. In (1) ist *Lei-* akzentuell hervorgehoben. Die Akzentuierung allein bestimmt die Ausdehnung des Fokus jedoch nicht eindeutig. So ist die Hervorhebung von *Lei-* in der Antwort aus (1) auch mit einem umfangreicheren Fokus verträglich, wie er in der Antwort auf (2) auftritt:

- (2) **Mutter:** *Was läßt die Meisterin die Nina denn gerade machen?* (Kontextsatz)

**Angestellte:** *Sie läßt die Nina das Leinen weben.*

Hier ist *das Leinen weben* fokussiert. Es handelt sich um Fokusprojektion (**weiter Fokus**): Der Fokus umfaßt mehr als die betonte Objekt-NP *das Leinen*, die den Fokusakzent trägt.

In diesem Beitrag soll untersucht werden, ob sich die intonatorische Markierung des engen Fokus von der des weiten unterscheiden läßt (zur intonatorischen Markierung des Fokus sowie zur wechsel-

seitigen Beeinflussung von intonatorischer Fokus- und Modusmarkierung siehe [3], [6]). Als Verfahren zur Trennung der beiden Klassen wurde die Diskriminanzanalyse eingesetzt.

## **2. Untersuchungsmaterial**

Unserer Untersuchung lagen die folgenden drei Strukturen zugrunde:

- (3) *Sie läßt die Nina das Leinen weben./?*
- (4) *Lassen Sie den Mann die Bohnen schneiden!/?*
- (5) *Lassen wir den Leo die Blumen düngen!/?*

Satz (3) kann als Aussagesatz oder als assertive Frage realisiert werden, (4) und (5) als Imperativsatz oder als Entscheidungsfragesatz. Eine ausführliche Beschreibung des Materials findet sich in [1, S.305ff.]. Die Sätze waren in Situationsbeschreibungen und Kontexte wie in Beispiel (1) eingebettet; damit konnten Satzmodus, Position des Fokusakzents und Fokusausdehnung eindeutig festgelegt werden. Sechs untrainierte Sprecher (drei weibliche, drei männliche) produzierten die Kontext-Testsatz-Paare insgesamt 360 mal. Hier interessieren die Fälle, in denen nur das eingebettete Objekt (*das Leinen, die Bohnen, die Blumen*) oder aber die eingebettete VP (*das Leinen weben, die Bohnen schneiden, die Blumen düngen*) fokussiert sind (72 Nicht-Fragen bzw. 81 Fragen). In beiden Fällen ist das Nomen der Objekt-NP akzentuell hervorgehoben.

## **3. Merkmale**

Wie in [2] erläutert, wurden zur Beschreibung des Verlaufs der intonatorischen Parameter 'Tonhöhe', 'Lautstärke' und 'zeitliche Strukturierung' für die Objekt-NP (zweite Phrase) und das Verb (dritte Phrase) die folgenden Merkmale extrahiert:

- Maximaler und minimaler  $F_0$ -Wert, transformiert in Halbtöne minus sprecherspezifischem Basiswert;
- Der Abstand von maximalem und minimalem  $F_0$ -Wert auf der Zeitachse;
- Relative Phrasendauer, berechnet nach der Formel  
(aktuelle Phrasendauer / mittlere Phrasendauer pro Sprecher) \* (aktuelle Dauer / mittlere Silbendauer);
- Maximale logarithmische Energie im Bereich 0 - 5000 Hz.

In [8, S.16ff.] wird für Nicht-Fragen berichtet, daß bei weitem Fokus der  $F_0$ -Abfall auf das nachfolgende Verb ausgedehnt werden kann. Daher berechnet wird als zusätzliches Merkmal die Steigung zwischen dem maximalen  $F_0$ -Wert auf der zweiten Phrase und dem  $F_0$ -Minimum vor dem Anstieg zum  $F_0$ -Maximum der dritten Phrase.

## **4. Ergebnisse**

Diskriminanzanalysen wurden für Frage- und Nicht-Fragesätze getrennt durchgeführt, da diese beiden Satztypen intonatorisch unterschiedlich markiert werden [3]. Auf die Fragesätze soll im folgenden nicht weiter eingegangen werden: Die starke Überlagerung der intonatorischen Markierung des Fokus

mit der des Satzmodus sowie die unterschiedlichen, sprecherabhängigen Realisationstypen [3] erschweren eine Trennung der Klassen. Bei der relativ kleinen Stichprobe können somit die Ergebnisse nicht sinnvoll interpretiert werden.

Wir berechneten die folgenden Lern-Test-Konstellationen:

- l=t: Die Äußerungen aller Sprecher dienen sowohl als Lern- als auch als Teststichprobe. Damit läßt sich die Obergrenze der Klassifikationsgüte feststellen. Allerdings besteht die Gefahr der Überadaption, d.h. das Verfahren 'lernt' u. U. die Stichprobe auswendig. Mit einem Auswahlverfahren ('stepwise variable selection') wurde die relevante Merkmalmenge bestimmt. Die Erkennungsrate betrug 73 Prozent;
- l5t1: Reihum bilden die Äußerungen von fünf Sprechern die Lernstichprobe und die des jeweils sechsten Sprechers die Teststichprobe ('leave-one-out' zur Simulation der Sprecherunabhängigkeit). Die Erkennungsrate bei einer 'stepwise variable selection' betrug 63 Prozent.

Die relevantesten Merkmale bei l=t waren die relative Phrasendauer auf der zweiten Phrase sowie für die zweite und die dritte Phrase der Abstand zwischen maximalem und minimalem  $F_0$ -Wert. Mit diesen drei Merkmalen erhielten wir bei l=t 65 Prozent und bei l5t1 63 Prozent Erkennungsrate.

### **5. Diskussion**

Beim Gebrauch aller Merkmale überadaptiert das Verfahren bei l=t, beim Gebrauch der drei relevantesten nicht. Man kann also sagen, daß diese drei Merkmale - im Rahmen der nicht besonders hohen Erkennungsrate von 63 Prozent - sprecherübergreifend verwendet werden.

Die Erkennungsraten sind eher niedrig, besonders im Vergleich zu den Erkennungsraten von über 90 Prozent für den Fokusakzent in [3]. Die Stichprobe ist allerdings zu klein, um daraus den Schluß ziehen zu können, daß zwischen der intonatorischen Markierung des engen und des weiten Fokus kein Unterschied besteht. Zum einen gibt es die Möglichkeit, daß die Sprecher zwar grundsätzlich den Unterschied markieren, daß sie aber einige Male fehlproduzierten, und daß diese Fehlproduktionen bei der relativ geringen Zahl der Fälle das Ergebnis stark beeinflussen. Zum anderen deutet eine Inspektion der für die einzelnen Sprecher unterschiedlich relevanten Merkmale darauf hin, daß diese sprecherspezifisch eingesetzt werden.

Auch wenn ein sprecherspezifischer Einsatz der intonatorischen Parameter vorliegt, so zeigen die Mittelwerte der drei relevantesten Merkmale über alle Sprecher doch deutliche und gut interpretierbare Unterschiede:

In Abb. 1 sind die Mittelwerte über alle Sprecher für den Abstand von maximalem und minimalem  $F_0$ -Wert der beiden Phrasen auf der Zeitachse dargestellt. In X-Richtung ist die Zeit vom Beginn der Äußerung in Centisekunden angegeben, in Y-Richtung die Grundfrequenz in Halbtönen minus sprecherspezifischem Basiswert. Bei engem Fokus ist der Abstand bei der zweiten Phrase größer und bei der dritten Phrase kleiner. Damit ändert sich auch die Steigung, die bei engem Fokus auf der zweiten Phrase flacher und auf der dritten Phrase steiler ist als bei weitem Fokus. Da bei gleichen

Ausgangs- und Zielwerten ein flacher und damit längerer  $F_0$ -Verlauf prominenter als ein weniger flacher und damit kürzerer ist [4], kann man sagen, daß bei engem Fokus die zweite Phrase als Trägerin des Fokusakzents eindeutiger fokussiert ist.

Das gleiche gilt für das relevanteste Merkmal 'relative Dauer der zweiten Phrase', wo der Mittelwert 3.47 bei engem und 3.22 bei weitem Fokus beträgt.

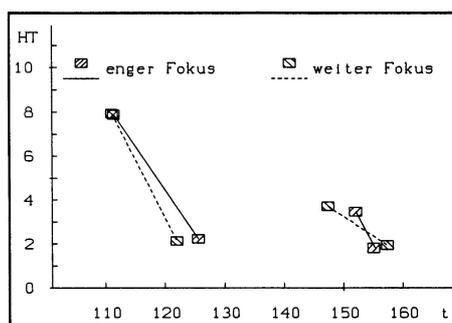


Abb. 1

Interessant ist, daß in einem Hörtest der Fokusakzent bei engem und weitem Fokus gleichermaßen zu jeweils 80 Prozent auf der zweiten Phrase gehört wurde [2, S.30]. Dieses Ergebnis war zu erwarten, da der Träger des Fokusakzents in beiden Fällen die Objekt-NP sein muß. Zwar werden die drei oben genannten relevantesten Merkmale auch für die Markierung des Fokusakzents eingesetzt, sie sind allerdings bei weitem nicht so wichtig wie z.B. die  $F_0$ -Maxima [2, S.41ff.]. Man kann also vorsichtig daraus folgern, daß dann, wenn die Sprecher **engen** vs. **weiten Fokus** überhaupt intonatorisch markieren, sie dafür Merkmale einsetzen, die für die Markierung des **Fokusakzents** eher zweitrangig sind.

### 6. Zusammenfassung

Während die Position des Fokusakzents relativ sicher mit Hilfe von intonatorischen Merkmalen bestimmt werden kann [3], ist dies für die Ausdehnung des Fokus nicht so einfach möglich. Daher sollte an einer größeren Stichprobe untersucht werden, ob sich zuverlässig sprecherspezifische intonatorische Markierungsstrategien ermitteln lassen.

Die Arbeiten an der Universität München entstanden im Rahmen des DFG-Projekts 'Modus-Fokus-Intonation', die an der Universität Erlangen im Rahmen des BMFT-Verbundvorhabens 'Sprachverstehende Systeme'.

- [1] Altmann, H., Batliner, A., Oppenrieder, W. (Hgg.): "Zur Intonation von Modus und Fokus im Deutschen", Max Niemeyer Verlag, Tübingen, 1989.
- [2] Batliner, A.: "Fokus, Modus und die große Zahl. Zur intonatorischen Indizierung des Fokus im Deutschen", in [1], S. 21-70.
- [3] Batliner, A., Nöth, E.: "The Prediction of Focus", Proceedings ECST, Vol. 1, Paris, 1989, S. 210-213.
- [4] Black, J.: "The Magnitude of Pitch Inflections", Proceedings of the Sixth International Congress of Phonetic Sciences, Prag, 1967, Hueber Verlag, München, 1970, S. 177-181.
- [5] Jacobs, J.: "Fokus-Hintergrund-Gliederung und Grammatik", in: Altmann, H. (Hg.): "Intonationsforschungen", Tübingen, 1988, S. 89-134.
- [6] Oppenrieder, W.: "Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung", in [1], S. 267-280.
- [7] Uhmman, S.: "Fokussierung und Intonation. Eine Untersuchung zum Deutschen an Frage/Antwort-Sequenzen in experimentellen Dialogen.", Masch. Phil. Diss., Konstanz, 1987.
- [8] Uhmman, S.: "On the Tonal Disambiguation of Focus Structures in German", unveröffentlichtes Manuskript, Wuppertal, 1989.