Gefördert vom
bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil
Verbundvorhaben

Prosodische Etikettierung des Deutschen mit ToBI

Matthias Reyelt
Martine Grice
Ralf Benzmüller
Jörg Mayer
Anton Batliner

Universität Braunschweig
Universität Saarbrücken
Universität Stuttgart
Universität München

Vm
Report 154
12. August 1996

12. August 1996

Matthias Reyelt
Martine Grice
Ralf Benzmüller
Jörg Mayer
Anton Batliner

Institut für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstr. 22
38092 Braunschweig

Institut für Phonetik
Universität des Saarlandes
66041 Saarbrücken

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Azenbergstr. 12
70174 Stuttgart

Institut für Deutsche Philologie
Ludwig-Maximilian-Universität München
Schellingstr. 3
80799 München

Tel.: (0531) 391 - 2479
Fax: (0531) 391 - 8218
e-mail: `reyelt@ifn.ing.tu-bs.de`

Gehört zum Antragsabschnitt: 14.6 Prosodische Etikettierung

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 N0 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Inhaltsverzeichnis

1 Einleitung

In den letzten Jahren hat die automatische Spracherkennung große Fortschritte gemacht und ist dabei, den Übergang vom Einzelworterkenner zum sprachverstehenden System zu vollziehen. Mittlerweile erreichen statistische Worterkenner auch für Spontansprache Erkennungsraten, die eine weitergehende syntaktisch–semantische Analyse erlauben. Problematisch ist dabei, daß die zu analysierende Wortkette ohne zusätzliche Information nur schwer in syntaktische Einheiten segmentierbar ist. Kompe et. al. [10] geben als Beispiel die folgende Wortkette:

ja | zur | Not | geht's | auch | am Samstag |

Die Balken geben dabei mögliche Satzgrenzen an, die in geschriebener Sprache durch Komma, Punkt oder Fragezeichen gekennzeichnet würden. Insgesamt geben Kompe et. al. 36 (!) syntaktisch korrekte Alternativen an, zwei davon sind:

Ja? Zur Not geht's? Auch am Samstag?
Ja. Zur Not. Geht's auch am Samstag?

Da Satzgrenzen häufig prosodisch – sei es durch Intonation, Längung oder Pausen – markiert sind, kann eine automatische Prosodieanalyse hier wertvolle zusätzliche Information liefern. Batliner et. al. [2] zeigen z.B., daß sich die Dauer des syntaktischen Parsing teilweise stark verkürzt, wenn in das Wortgitter zusätzlich prosodische Information über Phrasengrenzen eingefügt wird. Batliner et. al. weisen auf die Bedeutung der Intonation für die Disambiguierung des Satzmodus hin [3]. Die Integration eines Prosodiemoduls in ein automatisches Übersetzungssystem wird z.B. im Projekt Verbmobil [23] durchgeführt. Der Prosodiemodul reichert dabei den Worthypothesengraphen durch Informationen über prosodische Grenzen und Akzente an, bevor dieser syntaktisch–semantisch analysiert wird (vgl. auch [8]).

Allerdings wird zum Training und Test statistischer Erkenner Datenmaterial in großen Mengen benötigt. Dieses Material muß zuvor von Hand prosodisch etikettiert werden, eine Arbeit, die notwendigerweise von mehreren trainierten Hilfskräften – ev. sogar an unterschiedlichen Orten – durchgeführt wird. Trotzdem müssen die dabei entstehenden Transkriptionen "einheitlich" sein, d.h. die Etiketten müssen von verschiedenen Personen gleich angewandt werden. Eine völlige Übereinstimmung der Transkriptionen, ganz gleich welcher Art, ist allerdings illusorisch. Daher sind Untersuchungen zur Konsistenz wichtig, um den Grad der Übereinstimmung abzuschätzen. Derartige Untersuchungen wurden von Tillmann et. al. [22] für phonetisch enge Transkriptionen und von Reyelt [18] für prosodische Etiketten durchgeführt und sind ein wichtiges Maß für die Qualität der Verfahren.

Ein Inventar prosodischer Etiketten, das für die Bearbeitung großer Korpora geeignet sein soll, muß daher neben der prinzipiellen phonologisch adäquaten

Beschreibung noch weitere Anforderungen erfüllen (vgl. hierzu auch die in [21] geforderten Kriterien):

Das Inventar muß relativ schnell erlernbar sein. Es muß einheitliches Trainingsmaterial geben, das die Anwendung der Etiketten möglichst umfassend an repräsentativen Sprachbeispielen beschreibt, und zwar so, daß das Training der Transkribenten an unterschiedlichen Orten durchgeführt werden kann. Um die Verarbeitung der Daten mit statistischen Methoden zu ermöglichen, müssen die Datenformate maschinell verarbeitbar sein. Weiterhin sollten Evaluierungen der Transkriptionen durchgeführt werden, um Anhaltspunkte über deren Qualität zu erhalten und ggf. Schwachpunkte zu verbessern.

2 Das ToBI–System

Eine Beschreibung der Intonation durch eine Folge hoher und tiefer Töne, wie sie im von Pierrehumbert [15] entwickelten Tonsequenzansatz verwendet wurde, bildet den Ausgangspunkt für eine Reihe prosodischer Beschreibungen. Anfang der neunziger Jahre wurde in den USA unter dem Namen ToBI (Tone and Break Indices) ein einheitlicher Standard für ein Inventar prosodischer Etiketten entwickelt, um die dort existierenden prosodisch etikettierten Datenbasen zusammenzufassen [21] [20]. Neben Tonakzenten, Phrasenakzenten (intermediärer Grenzton) und Grenztönen werden prosodische Grenzen verschiedener Stärke (*intermediate* und *intonational phrase boundaries*) markiert. Das ToBI–System für das Englische wurde in umfangreichen Transkriptionsexperimenten evaluiert, Pitrelli et. al. [16] geben für Tonakzente eine Übereinstimmung von 68% an, für Phrasenakzente 85% und für Grenztöne 91%.

Für das Deutsche wurden an mehreren Instituten unabhängig voneinander Adaptionen des ToBI–Systems durchgeführt, u.a. in Saarbrücken, Stuttgart und Braunschweig/München (Verbmobil). Auf einem Workshop in Stuttgart im Februar 1995 wurden anhand von Sprachdaten, die an den drei Instituten etikettiert wurden, die Beschreibungssysteme verglichen und Unterschiede bzw. Gemeinsamkeiten herausgearbeitet. Auf dem *One Day Workshop on Prosodic Labelling* im August 1995 in Stockholm zeigte sich, daß auch in Hamburg und Nijmegen prosodische Etikettierungen des Deutschen nach dem ToBI–System durchgeführt werden[1]. Auf diesem Workshop wurde auch der Vergleich mit einem konturbasierten Ansatz (PROLAB, vgl. [9]) vorgenommen. Auf einem weiteren Workshop in Stuttgart wurde ein gemeinsames ToBI–System für das Deutsche (GToBI) ent-

[1]In der Vorbereitungsphase dieses Workshops wurde unter den deutschen Teilnehmern eine Umfrage über die verwendeten Etikettensysteme durchgeführt. Die Ergebnisse sind in [13] zusammengefaßt.

worfen und ein Trainingskorpus zusammengestellt. Weiterhin sollte das System in einem ersten Transkriptionsexperiment überprüft werden, bei dem das Sprachmaterial aus einer Mischung unterschiedlicher Korpora bestand.

3 Teilnehmer und Korpora

Die Teilnehmer entwickelten ihre prosodischen Etikettensysteme anhand unterschiedlicher Sprachdaten mit unterschiedlichen Zielsetzungen:

Das Saarbrücker System [6] wurde zur Transkription spontansprachlicher Dialoge entworfen. Es handelt sich um task-orientierte Dialoge (vgl. HCRC Map Task [1]), in denen die Versuchspersonen ohne Sichtkontakt (d.h. nur über den akustischen Kanal) Informationen über die Route auf einer Landkarte austauschen. Dadurch, daß sich die Versuchspersonen kennen und die Aufgabe spielerisch ausgeführt wird, sind relativ viele stilisierte Intonationskonturen (zu Stilisierung (*stylized contours*) vgl. [5]) enthalten. Parallel dazu werden vergleichbare Dialoge in verschiedenen Varietäten des Deutschen, Italienischen und Bulgarischen aufgenommen.

In Stuttgart wird hauptsächlich gelesene Sprache transkribiert. Das Transkriptionssystem [14] wurde anhand konstruierter, vorgelesener Sätze entwickelt und wird nun vor allem zur Etikettierung von Radionachrichten verwendet. Die Leistungsfähigkeit des Stuttgarter Systems bei der Transkription von Spontansprache wird derzeit getestet.

Im Projekt Verbmobil werden Terminabsprachen zwischen zwei Versuchspersonen aufgenommen. Die Dialoge sind zwar spontan, die Dialogpartner müssen jedoch einen Knopf drücken, bevor sie sprechen können. Die Daten werden in Kiel, Bonn, München und Karlsruhe aufgenommen und in Braunschweig prosodisch etikettiert; das dazu verwendete Transkriptionssystem [19] wurde in Zusammenarbeit mit Projektpartnern in München und Erlangen entwickelt.

Es zeigte sich, daß sich die unterschiedlichen Schwerpunkte der einzelnen Transkriptionssysteme gut ergänzen. Während das Verbmobil–System auf vielen nord– und süddeutschen Varietäten basiert, sind im Saarbrücker Transkriptionssystem stilisierte Konturen sehr genau beschrieben. Das Stuttgarter Transkriptionssystem wiederum ist stärker phonologisch orientiert.

4 Beschreibung des GToBI–Systems

In GToBI werden die Tonhöhenverläufe mit zwei verschiedenen Tönen beschrieben: *H* und *L.* H steht für einen hohen und L für einen tiefen Zielpunkt *(target)* eines Tonhöhenverlaufs. Die Rekonstruktion der beschriebenen Kontur erfolgt durch regelhafte Interpolation zwischen den Zielpunkten. Die Verknüpfung der tonalen mit der textuellen Ebene geschieht über Diakritika. *Akzenttöne* werden durch einen "*" (*stern*) mit der akzentuierten Silbe verknüpft, der *Phrasenton* bei intermediären Phrasen (B3) wird durch "–" (*minus*) an das letzte Wort der Phrase gebunden. Grenzen von Intonationsphrasen (B4) werden immer aus zwei Tönen gebildet, dem *Phrasenton* und einem *Grenzton*, der mit "%" gekennzeichnet wird. Der Grenzton beschreibt dabei den Tonverlauf am Ende der Intonationsphrase, der Phrasenton den Verlauf zwischen letztem Akzentton und Grenzton.

4.1 Tonakzente

Die folgenden Tonakzente können auftreten:

H* : Dies ist der normale *Gipfelakzent*. Er zeichnet sich durch eine Abweichung der Tonhöhe nach oben aus. Folgen zwei oder mehr H*–Akzente aufeinander, so kann die Tonhöhe zwischen diesen leicht abfallen *(sagging)*.

L+H* : Bei diesem Akzent findet auf der akzentuierten Silbe ein steiler Anstieg statt, so daß der Höhepunkt der Bewegung erst recht spät in der akzentuierten Silbe stattfindet. Dieser Akzent tritt häufig als emphatischer oder kontrastiver Akzent auf.

L*+H : Hier gibt es einen tiefen Zielpunkt früh in der akzentuierten Silbe. Der Gipfel ist hinter die akzentuierte Silbe verschoben (vgl. *scooped accent* bei Ladd [12]).

L* : Dieser Akzenttyp zeichnet sich durch eine Abweichung der Tonhöhe auf der akzentuierten Silbe nach unten im unteren Bereich des Stimmumfangs aus, *Talakzent*.

H+L* : Bei diesem Akzent fällt die Tonhöhe von einem hohen Zielpunkt vor der Akzentsilbe in den unteren Bereich des Stimmumfangs ab (*early peak* bei Féry [4]).

H+!H* : Dieser Akzent ähnelt dem H+L*, fällt aber nur bis in den mittleren Bereich des Stimmumfangs ab.

4.2 Tonale Grenzmarkierungen

Während Grenzen zwischen intermediären Phrasen nur durch einen *Phrasenton* (also H– bzw. L–) gekennzeichnet werden, sind Grenzen an Intonationsphrasen immer bitonal durch *Phrasenton* und *Grenzton* markiert[2]. Zusammen mit H– und L– an intermediären Phrasengrenzen gibt es sechs mögliche tonale Grenzmarkierungen:

L–L% : terminaler Fall, Grenze im unteren Bereich des Stimmumfangs.

L–H% : Nach Gipfelakzent fallend–steigende Grenze, nach L* leichter Anstieg in der letzten Silbe vor der Grenze.

H–L% : ebene Grenze im mittleren Bereich des Stimmumfangs, "progredient".

H–H% : Anstieg nach dem letzten Akzent bis in die obersten Bereiche des Stimmumfangs, "interrogativ".

4.3 Downstepping

In einer Folge von Gipfelakzenten können diese jeweils "treppenartig" ein Stück nach unten versetzt sein. Dieser Effekt wird als *Downstepping* bezeichnet und diakritisch durch ein dem hohen Ton des Akzents vorgestelltes "!" etikettiert. Weiterhin kann Downstepping auch an hohen Phrasentönen markiert werden.

4.4 Upstepping

Ein treppenartiger Anstieg der Akzente wird durch ein vorgestelltes "∧" etikettiert.

4.5 Grenztypen (break indices)

Die Grenztypen stellen ein (subjektives) Maß für die "Stärke" der Trennung dar.

B1 normale Wortgrenze. Sie trägt *keine* prosodische Markierung und wird auch normalerweise nicht etikettiert. Der Vollständigkeit halber ist sie hier trotzdem aufgeführt.

B2 irreguläre Grenze, Abbruch, Häsitation.

B3 intermediäre Phrasengrenze. Leichte tonale Markierung innerhalb einer Intonationsphrase. B3 werden normalerweise durch einen Phrasenton (H– bzw. L–) markiert.

[2] Im Englischen *phrase accent* und *boundary tone*. Ladd [11] führt allgemein für tonale Markierung prosodischer Grenzen den Begriff *edge tone* ein.

B4 Grenze einer Intonationsphrase. Starke tonale Markierung, häufig verbunden mit Dehnung oder Pause. B4 werden normalerweise bitonal durch einen Phrasenton (H–/L–) und einen Grenzton (H%/L%) markiert.

Grenztypen und Grenztöne sind nur teilweise voneinander unabhängig. Dadurch, daß B4 immer bitonal und B3 immer durch einen Phrasenton markiert sind, brauchen sie nicht unbedingt explizit etikettiert werden. Sie lassen aus der tonalen Etikettierung rekonstruieren. Es wird dann nur eine explizite Etikettierung irregulärer Grenzen (B2) benötigt. Das heißt jedoch nicht, daß die in GToBI etikettierten prosodischen Grenzen rein auf der intonatorischen Markierung beruhen. Z. B. liegt der Unterschied zwischen H- (B3) und H-L% (B4) hauptsächlich in der unterschiedlich wahrgenommenen "Tiefe" der Grenze (die auch durch Dehnung bzw. Pause verursacht sein kann).[3]

5 Transkriptionsexperiment

Um die Verwendbarkeit des Inventars zu untersuchen, wurde ein Transkriptionsexperiment durchgeführt. Dazu wurde zunächst ein Trainingskorpus festgelegt. Dieses enthielt zu den einzelnen Etiketten passende Sprachbeispiele. Weiterhin wurde eine kurze Beschreibung ausgearbeitet, in der die Anwendung der Etiketten in den Sprachbeispielen erläutert wurde.

Das Testmaterial enthielt zum Teil gelesene Sprache:
1) Nachrichten, gesprochen von einem trainierten Sprecher (DLF)
2) Buchausschnitt, gelesen von einem trainierten Schauspieler
3) Abschnitte aus einem Reiseführer, gelesen von einem untrainierten Sprecher
Der andere Teil bestand aus Spontandialogen:
1) Eine Terminabsprache aus dem Verbmobil Korpus
2) Ein Dialog aus dem Map Task Korpus
Insgesamt bestand das Material aus 35 Äußerungen, 733 Wörtern und 304 Sekunden. Das Material wurde parallel von 13 Transkribenten etikettiert. Zu den Transkribenten gehörten drei der AutorInnen, die restlichen waren Studenten mit nur wenig Transkriptionserfahrung, die sich anhand des Trainingsmaterials eingearbeitet hatten.

Wegen des beschränkten Materialumfangs und um die Studenten nicht zu überfordern, wurde ein eingeschränktes Inventar verwendet: Upstepping wurde

[3] Im Verbmobil GToBI werden die Grenztypen etwas anders etikettiert: normale Wortgrenze als B1, intermediäre PG als B2, Intonationsphrasengrenze als B3. Die irreguläre Phrasengrenze wird mit B9 etikettiert. Der Unterschied ist aber rein formal, die Übersetzung nach GToBI ist eindeutig.

nicht etikettiert, ein diakritisches Fragezeichen konnte verwendet werden, um Unsicherheit der Transkription anzuzeigen. Auch wurden für das Sprachmaterial nur norddeutsche Sprecher verwendet.

6 Arbeitsumgebung

Die Transkriptionen wurden in Stuttgart und Saarbrücken mit *ESPS xwaves(tm)* durchgeführt. In Braunschweig wurde das Programm *fish* verwendet, ein freies Softwarepaket zur Etikettierung von Sprachsignalen. Die Transkriptionen wurden akustisch mit visueller Unterstützung (Sprachsignal und Sprachgrundfrequenz) erstellt.

7 Resultate

Die resultierenden 13 Transkriptionen wurden miteinander verglichen, und die Korrespondenzen zwischen den einzelnen Transkribenten ermittelt. In Tabelle 1 sind zunächst für die 13 Transkribenten die Anzahlen vergebener Akzente und Grenzen angegeben. Von den 733 Wörtern markierten die Transkribenten im Mittel 301 als akzentuiert und etikettierten 60 B3 und 90 B4. Die bei Tab. 1 angegebenen Standardabweichungen (σ) zeigen, daß die Anzahl vergebener Etiketten unterschiedlich streut. Während Tonakzente und B4 relativ zum Mittelwert wenig streuen, liegt die Standardabweichung bei B3 mit 25 bei im Mittel 60 etikettierten B3 relativ hoch. Dies zeigt, daß die Etikettierung der intermediären Phrasengrenze weniger verläßlich ist als die der anderen Kategorien.

Die prozentualen Korrespondenzen zwischen den Transkribenten wurden analog dem in [21] für das englische ToBI–System beschriebenen Verfahren ermittelt. Bei diesem Verfahren werden für jedes der 733 Wörter alle Transkribentenpaare (78 bei 13 Transkribenten) miteinander verglichen. Insgesamt wurden also 57174 Transkriptionspaare ausgewertet; die Korrespondenz ergibt sich dabei aus der Anzahl gleich transkribierter Wörter bezogen auf die Gesamtheit. Die Korrespondenzen wurden für die unterschiedlichen Etiketten einzeln ermittelt. Sie sind in Tab. 2 angegeben. Dabei wird zunächst verglichen, ob jeweils zwei Transkribenten überhaupt einen Akzent auf einem Wort etikettiert hatten. Der zweite Vergleich bezog sich auf die Art des Tonakzentes (dabei wurden alle Wörter, also auch die unakzentuierten mit einbezogen). Der dritte Vergleich berücksichtigte zusätzlich, ob beide übereinstimmend Downstepping etikettierten. Für die Grenzmarkierungen wurde einmal die tonale Markierung verglichen und dann der Grenztyp, also die “Tiefe” der Grenze. Zunächst wurde dieser Vergleich für alle 13 Transkribenten

Tabelle 1: Anzahl vergebener Etiketten für die 13 Transkribenten. Der Mittelwert für die Anzahl der Akzente liegt bei 301 ($\sigma = 17$), für B3 liegt der Mittelwert bei 60 ($\sigma = 25$) und für B4 bei 90 ($\sigma = 18$).

	tr1	tr2	tr3	tr4	tr5	tr6	tr7
Anz. Akzente	295	304	295	301	319	312	288
Anz. B3	74	50	31	62	38	73	34
Anz. B4	104	88	56	92	100	110	87
	tr8	tr9	tr10	tr11	tr12	tr13	
Anz. Akzente	279	290	328	271	319	312	
Anz. B3	50	65	70	131	53	54	
Anz. B4	104	93	91	62	113	63	

Tabelle 2: Korrespondenzen insgesamt und getrennt für Experten und Studenten. Die Spalten geben an: (1) Kategorie, (2) Anzahl der unterschiedlichen Etiketten (jeweils +1 für keine Markierung) , (3) Gesamtkorrespondenz, (4) Expertengruppe, (5) Studentengruppe

	Auswahl	Korr. ges.	Exp.	Stud.
akz./unakz.	2	87%	91%	86%
Tonakzente	7	74%	81%	73%
Tonakz. + Downst.	10	71%	78%	70%
tonale Grenzmarkierung	7	86%	90%	85%
Grenztypen	4	87%	89%	86%

gemeinsam durchgeführt, dann wurden zwei Gruppen gebildet, einmal die 3 Experten (Autoren) und die 10 Studenten.

Mit Ausnahme der Tonakzente liegen die Korrespondenzen durchgehend deutlich über 80%. Die Werte sind denen in [16] angegebenen für englisches ToBI vergleichbar, ein detaillierter Vergleich beider Experimente wird von Grice [7] durchgeführt. Allein die Tonakzente liegen nur bei etwa 70%. Dies liegt zum Teil an dem reichhaltigen Inventar, zum Teil sind die Kategoriengrenzen zwischen den Tonakzenten aber auch unscharf, sie haben Randbereiche, in denen die Unterscheidung schwerfällt. Z.B. verfügen H* und L+H* beide über einen hohen Zielpunkt in der akzentuierten Silbe und sind, gerade am Beginn einer Intonationsphrase, beide tendenziell steigende Akzente. Der durch den tiefen Führungston hervorgerufene stärkere Anstieg bei L+H* ist nicht immer klar auszumachen.

Es zeigt sich weiterhin, daß es zwar, wie zu erwarten war, einen Unterschied zwischen der Studenten– und der Expertengruppe gibt, daß dieser Unterschied aber nicht sehr groß ist. Die Studenten scheinen trotz des kurzen Trainings keine prinzipiellen Probleme mit der Anwendung des GToBI–Etikettensystems zu haben. Durch ein erweitertes und verbessertes Training könnte der Unterschied zur Expertengruppe noch verringert werden. Eine Korrespondenz von 90% wäre hier ein erreichbares Ziel (außer wohl für die Etikettierung der Tonakzente).

In Tab. 7 wurden die Korrespondenzen für die Tonakzente einzeln ermittelt. Für die Berechnung konnte das oben verwendete Verfahren nicht benutzt werden. Es wurde daher ein Verfahren angewandt, das an die Erkennungsrate bei der automatischen Spracherkennung angelehnt ist. Die Berechnung erfolgt gemäß Gleichung 1[4].

$$corr_{1,2,label} = \frac{n_{corr(1,2),label}}{(n_{1,label} + n_{2,label})/2} \qquad (1)$$

Werden die übereinstimmungen für die Tonakzente wie in Tab. 7 einzeln durchgeführt, so zeigen sich ganz erhebliche Unterschiede zur Gesamtkorrespondenz. In letztere gehen die einzelnen Akzente mit ihrer Häufigkeit gewichtet ein, so daß selten vorkommende Akzente wie z.B. H+L* trotz sehr geringer Übereinstimmung den Wert kaum verschlechtern. Gerade die selten vorkommenden Akzente werden aber offensichtlich leicht verwechselt.

Interessant ist auch die Frage, ob sich Unterschiede zwischen den einzelnen Transkribenten zeigen. Frühere Untersuchungen [17] zeigten, daß die Streuung sehr uneinheitlich verteilt sein kann. Dazu wurde jeweils die Korrespondenz eines Transkribenten zu allen anderen berechnet und gemittelt. Die Ergebnisse zeigt Tab. 7.

Offensichtlich gibt es unter den Transkribenten keine “Ausreißer”, die deutlich anders etikettieren als der Rest. Auch die Experten zeigen weder bessere noch schlechtere Werte als die Studenten, was ein Indiz dafür wäre, daß beide Gruppen unterschiedlich etikettieren. Das Training scheint also zumindest so beschaffen zu sein, daß die Transkribenten die Etiketten zwar etwas unsicherer, aber nicht falsch anwandten.

[4]Das Prinzip ist folgendermaßen: zunächst wird die Etikettierung des einen Transkribenten als Referenz angenommen und die “Erkennungsrate” des anderen ermittelt. Danach wird getauscht und die “Erkennungsrate” des ersten berechnet. Zuletzt werden beide Werte gemittelt.

Tabelle 3: Korrespondenzen für die einzelnen Tonakzente, Häufigkeit der Tonakzente (Gesamtsumme ist 733 Wörter x 13 Transkriptionen = 9529).

	ohne Downstepping	Anzahl	mit Downstepping	Anzahl
unakz.	90%	5757	90%	5757
H*	62%	2272	56%	1914
!H*			28%	358
L+H*	38%	610	38%	576
(L+!H*)			6%	34
L*+H	35%	408	34%	382
L*+!H			10%	26
H+!H*	14%	171	14%	171
H+L*	14%	72	14%	72
L*	18%	239	18%	239

Tabelle 4: Korrespondenz jedes der 13 Transkribenten zu den restlichen. Alle Werte in Prozent. Die Experten sind mit "*" markiert

tr1	tr2	tr3	tr4	tr5	tr6	tr7	tr8	tr9	tr10	tr11	tr12	tr13
86	87*	82	87	84	86*	86	87	85	86*	84	82	83

8 Zusammenfassung

Mit GToBI ist ein System zur prosodischen Transkription des Deutschen entwickelt worden, das bereits an mehreren Instituten in Deutschland angewandt wurde. Das hier beschriebene Experiment zeigte, daß Transkribenten schon nach einer kurzen Einarbeitungszeit mit dem GToBI–System Etikettierungen von guter Konsistenz durchführen können. Damit ist eine Vorbedingung für die Erstellung umfangreicher prosodisch etikettierter Datenbasen erfüllt. Um die Konsistenz weiter zu verbessern, muß an Hand der Ergebnisse des hier beschriebenen Transkriptionsexperiments das Trainingsprogramm überarbeitet und ergänzt werden. Das System soll auch noch in einigen weiteren Aspekten erweitert werden (Markierung von Kern- vs. Randtypen sowie der Sicherheit der Entscheidung).

Literatur

[1] Anne Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. The HCRC map task corpus. *Language and Speech*, 34(4):351–366, 1991.

[2] A. Batliner, A. Feldhaus, S. Geißler, T. Kiss, R. Kompe, and E. Nöth. Prosody, Empty Categories and Parsing — A Success Story. In *Proc. ICSLP*, Philadelphia, 1996. In Druck.

[3] A. Batliner, A. Kießling, R. Kompe, E. Nöth, and B. Raithel. Wann geht der Sonderzug nach Pankow? (Uhrzeitangaben und ihre prosodische Markierung in der Mensch-Mensch– und in der Mensch-Maschine–kommunikation). In *Fortschritte der Akustik – DAGA92*, pages 541–544, Bad Honnef, 1992. DPG–GmbH.

[4] Caroline Féry. *German Intonational Patterns*. Linguistische Arbeiten 285. Niemeyer, Tübingen, 1993.

[5] Dafydd Gibbon. *Perspectives of Intonation Analysis*. Lang, Hamburg, 1976.

[4] Ein Teil dieser Arbeiten wurde mit Mitteln des Bundesministers für Bildung und Forschung unter den Förderkennzeichen 01 IV 101 N/0 und 01 IV 102 F/4 gefördert. Die Verantwortung für den Inhalt liegt bei den Autoren. Die Sprachdaten für dies Experiment sowie GToBI–Trainingsmaterial ist via ftp verfügbar (Kontakt: benzm@coli.uni-sb.de oder mgrice@coli.uni-sb.de). Zur Visualisierung der Daten kann das Programm *fish* ebenfalls via ftp bezogen werden (Kontakt: m.reyelt@tu-bs.de).

[6] Martine Grice and Ralf Benzmüller. Transcription of German using ToBI–tones: The Saarbrücken System. Technical report, Institut für Phonetik, Universtität Saarbrücken, Saarbrücken, 1995.

[7] Martine Grice, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer, and Anton Batliner. Consistency in transcription and labelling of German intonation with GToBI. In *Proc. ICSLP*, Philadelphia, 1996. In Druck.

[8] W. Hess, A. Batliner, A. Kießling, R. Kompe, E. Nöth, A. Petzold, M. Reyelt, and V. Strom. Prosodic modules for speech recognition and understanding in Verbmobil. In Yoshinori Sagisaka, Nick Campbell, and Norio Higuchi, editors, *Computing Prosody*, pages 363–383. Springer, New York, 1996.

[9] K.J. Kohler. PROLAB – the Kiel system of prosodic labelling. In *Proc. International Congress of Phonetic Sciences*, volume 3, pages 162–165, Stockholm, August 1995.

[10] R. Kompe, A. Kießling, H. Niemann, E. Nöth, E.G. Schukat-Talamazzini, A. Zottmann, and A. Batliner. Prosodic scoring of word hypotheses graphs. In *Proc. EUROSPEECH*, volume 2, pages 1333–1336, September 1995.

[11] D. Robert Ladd. "linear" and "overlay" descriptions: An autosegmental–metrical middle way. In *Proc. ICPhS*, volume 2, pages 116–123, Stockholm, August 1995.

[12] D. Robert Ladd, Kim E.A. Silverman, and Klaus R. Scherer. Parametrische und kategoriale Ansätze bei der Erforschung intonatorischer Funktion. *Zeitschrift für Literaturwissenschaft und Linguistik*, 49:124–133, 1983.

[13] Jörg Mayer. Towards the workshop on prosodic labelling. Results of a questionnaire concerning current works on the prosodic labelling of German speech data bases. Technical report, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Stuttgart, 1995. http://www.ims.uni-stuttgart.de/phonetik/joerg/stockholm/questionnaire.html.

[14] Jörg Mayer. Transcription of German intonation: The Stuttgart System. Technical report, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Stuttgart, 1995.

[15] Janet Pierrehumbert. *The Phonology and Phonetics of English Intonation.* PhD thesis, M.I.T., 1980.

[16] John Pitrelli, Mary Beckman, and Julia Hirschberg. Evaluation of prosodic transcription labeling. In *Proc. ICSLP*, pages 123–126, Yokohama, 1994.

[17] Matthias Reyelt. Experimental investigation on the perceptual consistency and the automatic recognition of prosodic units in spoken German. In *Working papers*, volume 41, pages 238–241, Lund University, 1993. Dept. of Linguistics.

[18] Matthias Reyelt. Ein flexibles Programmpaket zur Visualisierung von Sprachdaten. In K. Fellbaum, editor, *Tagungsband Elektronische Sprachsignalverarbeitung*, pages 358–365, Berlin, 1994.

[19] Matthias Reyelt and Anton Batliner. Ein Inventar prosodischer Etiketten für Verbmobil. Verbmobil–Memo 34, Techn. Univ. Braunschweig, LM-Univ. München, 1994.

[20] K. Silverman, E. Blaauw, J. Spitz, and J. Pitrelli. Prosodic comparison of spontaneous speech and read speech. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, pages 1299–1302, 1992.

[21] Kim Silverman, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. Tobi: A standard for labeling english prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, pages 867–870, 1992.

[22] H.G. Tillmann, B. Eisen, and Ch. Draxler. Consistency of judgements in manual labelling of phonetic segments: The distinction between clear and unclear cases. In *Proc. ICSLP 92*, pages 871–874, 1992.

[23] W. Wahlster. Verbmobil: Translation of face-to-face dialogs. In *Proc. EUROSPEECH*, 1993.