Zwei Ansätze zur syntaxgesteuerten Prosodiegenerierung

Antje Schweitzer und Martin Haase, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
{antje.schweitzer, martin.haase}@ims.uni-stuttgart.de

Kurzfassung

Diese Arbeit präsentiert zwei Ansätze zur Prosodiegenerierung innerhalb eines deutschen Text-to-Speech-Systems mit unterschiedlichen Analysemethoden. Akzente und Phrasengrenzen werden einmal auf der Grundlage eines syntaktischen Parsers und einmal mit Hilfe eines Part-of-Speech-Taggers und einer darauf aufsetzenden linguistischen Analyse bestimmt. Beide Ansätze erweisen sich als sehr gut geeignet für die Vorhersage intonatorischer Ereignisse und bestätigen den engen Zusammenhang zwischen Syntax und Prosodie.

1 Einleitung

Die prosodische Struktur einer Äußerung wird durch ihre syntaktische Struktur beeinflußt (vgl. dazu u.a. Cinque (1993), Culicover und Rochemont (1983), Selkirk (1984), Hirst (1993)). Aus diesem Grund liegt es nahe, die Prosodiegenerierung in einem Text-to-Speech-System auf einer detaillierten Syntaxanalyse zu basieren.
Wir vergleichen hierzu zwei Ansätze. Einmal wird zur Bestimmung der Prosodie ein Part-of-Speech-Tagger (POS-Tagger), einmal ein syntaktischer Parser herangezogen. Diese Ansätze sollen anhand eines größeren Korpus ausgewertet und miteinander verglichen werden.
Beide Ansätze wurden in der deutschen Version des Text-to-Speech-System *Festival* (Black, Taylor, und Caley (1999), Möhler (1999)) implementiert. Es werden jeweils abstrakte intonatorische Labels vorhergesagt. Bei der Synthese wird dann gemäß diesen Labels der Grundfrequenzverlauf modelliert (Möhler, 1998). Ausführliche Beschreibungen zu beiden Ansätzen finden sich bei Haase (1999) und Schweitzer (1999).

2 Korpus

Für die Evaluation der beiden Ansätze wurde das Nachrichtenkorpus des Instituts für Maschinelle Sprachverarbeitung der Universität Stuttgart verwendet.
Es besteht aus Radionachrichten des Deutschlandfunk von 1995 und wurde manuell prosodisch annotiert (vgl. Mayer (1995), Rapp (1998)). Es umfaßt 105 Meldungen mit einer Gesamtdauer von 67 Minuten. Alle in den folgenden Abschnitten zitierten Beispiele stammen aus diesem Korpus.
Es werden zwei Typen von intonatorischen Ereignissen unterschieden, Grenztöne und Akzente. Grenztöne signalisieren intermediäre Phrasengrenzen (Label -) und Intonationsphrasengrenzen (Labels %, H% und L%). Für unsere Zwecke werden die drei Labels für Intonationsphrasengrenzen auf eine Kategorie % abgebildet.
Analog werden die neun verschiedenen Labels für Akzente (H*L, L*H, L*HL, HH*L, H*M, H*, L*, ..L und ..H) auf zwei Kategorien (H*L und L*H für fallende bzw. steigende Akzente) abgebildet.
Im Folgenden sind mit - und % bzw. mit H*L und L*H immer diese abstrakten Kategorien gemeint.

3 Intonationsbestimmung mit Hilfe eines POS-Taggers

3.1 Linguistische Analyse

Basierend auf den POS-Tags wird eine linguistische Analyse durchgeführt. Der hier verwendete POS-Tagger (Schmid, 1995) unterscheidet 51 verschiedene POS-Tags (Schiller, Teufel, Stöckert, 1995). Differenziert wird z.B. zwischen Eigennamen und sonstigen Substantiven, Finita und Infinita, Auxiliaren, Modalverben und Vollverben.

Noun Chunks. Mit Hilfe der POS-Tags wird für jede Äußerung eine vereinfachte syntaktische Struktur aufgebaut. Diese beinhaltet im wesentlichen *Noun Chunks* im Sinne von Abney (1995). Bei den Noun Chunks (NC) handelt es sich hier um einfache Fragmente von Nominalphrasen wie in (1). Direkt aufeinander folgende oder koordinierte Nomina einschließlich vorangehender Adjektive und Artikel gehören jeweils zu einem Noun Chunk. Sie werden als flache Struktur repräsentiert.

(1) [$_{NC}$die bosnisch-moslemischen Regierungstruppen] in [$_{NC}$der umkämpften Enklave Bihac]

Verbalkomplex. Folgen von Verben können im Deutschen hierarchisch angeordnet sein (vgl. Bech (1983), Haider (1993)). Jedes infinite Verb hat einen *Status*. Es werden 3 verschiedene Status unterschieden (1. Status ≅ Infinitiv, 2. Status ≅ Infinitiv mit „zu“, 3. Status ≅ Partizip). Das finite Verb *sollen* in (2) bettet z.B. Verben im 1. Status unter sich ein, in diesem Fall das Verb *werden*. Dieses wiederum kann Verben im 3. Status einbetten, hier also *verkauft*. Damit ergibt sich folgende Hierarchie: sollen (*finit*) > werden (*1. Status*) > verkauft (*3. Status*).

(2) Die Dasawerke in Laupheim, Peisenberg und Speyer sollen verkauft werden.

Topologische Analyse. Nach Grewendorf, Hamm, und Sternefeld (1989) läßt sich ein deutscher Satz in fünf aufeinanderfolgende topologische Felder aufteilen: Vorfeld (VF), linke Satzklammer, Mittelfeld, rechte Satzklammer und Nachfeld. Linke Satzklammer und Vorfeld können mit Hilfe der POS-Tags erkannt und bei der Intonationsgenerierung einbezogen werden.

Aufzählungen. Listen von Nomina bzw. Verben, die durch Kommata oder Konjunktionen verbunden sind, werden als Aufzählungen gekennzeichnet.

3.2 Phrasierung

Bei der Bestimmung der Intonation wird zunächst die Phrasierung festgelegt, d.h., die zu synthetisierende Äußerung wird in Intonationsphrasen und diese wiederum in intermediäre Phrasen aufgeteilt. Intermediäre Phrasengrenzen sind schwächer als Intonationsphrasengrenzen.
In den folgenden Abschnitten werden die für die Bestimmung der Phrasierung relevanten aus der linguistischen Analyse stammenden Merkmale kurz beschrieben und anhand von Beispielen aus dem Korpus verdeutlicht.

Noun Chunks. Laut Abney (1995) korrelieren Chunkgrenzen und prosodische Grenzen. Demgemäß werden nach Noun Chunks intermediäre Phrasengrenzen eingefügt, sofern die resultierenden Phrasen nicht eine gewisse Länge unterschreiten (vgl. „phonologische“ Schwächung bei Abney: hier werden prosodische Grenzen zwischen adjazenten Phrasen u.a. dann geschwächt, wenn eine von ihnen nur aus einem einzelnen Funktionswort besteht und beide außerdem in einer gewissen syntaktischen Beziehung stehen, wie z.B. Subjekt - Verb oder Verb - Objekt). Im Korpus findet sich nach Noun Chunks normalerweise eine intermediäre Phrasengrenze, wenn sie wie in (3) aus zwei oder mehr Wörtern bestehen.

(3) [$_{NC}$Radio Sarajevo] - berichtete heute früh, % [$_{NC}$die Fronten] - seien stabilisiert worden. %

Topologische Analyse. Das Vorfeld wird durch eine Phrasengrenze abgeschlossen, sofern es eine Mindestlänge von zwei Wörtern nicht unterschreitet. Z.B. wurde in (4) an allen Vorfeldgrenzen auch eine Phrasengrenze realisiert, außer im letzten Satz, wo das Vorfeld nur aus dem Personalpronomen *Sie* besteht. Es handelt sich meistens um Intonationsphrasengrenzen, jedoch kommen auch intermediäre Phrasengrenzen vor, insbesondere, wenn die resultierende Phrase relativ kurz ist, wie z.B. nach *die Fronten* oder *Radio Sarajevo*.

(4) [$_{VF}$Die bosnisch-moslemischen Regierungstruppen % in der umkämpften Enklave Bihac] % haben den serbischen Vormarsch - nach eigenen Angaben - gestoppt. % [$_{VF}$Radio Sarajevo] - berichtete heute früh, % [$_{VF}$die Fronten] - seien stabilisert worden. % [$_{VF}$An einzelnen Abschnitten] % habe die Armee - sogar kleinere Gebiete - zurückerobern können. % [$_{VF}$Im Südwesten - Bosniens] % geht die Offensive bosnisch-kroatischer Einheiten % gegen die Städte Glamoc - und Bosanko-Grahovo weiter. % [$_{VF}$Sie] werden von der regulären Armee Kroatiens - unterstützt. %

Aufzählungen. Zwischen Koordinationsgliedern befinden sich intermediäre Phrasengrenzen, wie z.B. bei den koordinierten Adjektiven in (5).

(5) Handel - bringe Wandel, % oft auch menschenrechtliche, - demokratische - und rechtsstaatliche Fortschritte. %

3.3 Akzentuierung

Die Akzentuierung wird erst nach Bestimmung der Phrasierung behandelt. Innerhalb jeder Phrase muß mindestens einem Wort ein Akzent zugewiesen werden. Dabei werden die unten beschriebenen linguistischen Merkmale berücksichtigt. Es wird zunächst nur bestimmt, ob ein Akzent realisiert werden soll oder nicht. Über die Kategorie des Akzents entscheidet eine Heuristik: Am Satzende werden fallende Akzente zugewiesen. Dabei wird von der Tatsache profitiert, daß neue Information, die typischerweise mit fallendem Akzent versehen wird, in Texten eher am Satzende steht.

Noun Chunks. Normalerweise wird innerhalb jedes Chunks dem am weitesten rechts stehenden Element ein Akzent zugewiesen. Da die deutsche Satzstruktur stark rechtsverzweigend ist, handelt es sich bei diesem Element meist um den Kopf des Chunks, der am tiefsten eingebettet ist und deshalb nach Cinque (1993) den Satzakzent tragen muß. Die Akzente auf *Bihac*, *Kämpfe*, *Regierungstruppen* und *Verbänden* in (6) können so erklärt werden.

(6) In $[_{NC}$der L*H bosnischen Moslem-Enklave L*H Bihac] - gingen $[_{NC}$die L*H Kämpfe] - zwischen $[_{NC}$den L*H Regierungstruppen] - und $[_{NC}$ L*H serbischen L*H Verbänden] - auch heute früh H*L weiter. % [1]

Verbalkomplex. Innerhalb einer Folge von Verben soll demjenigen Verb ein Akzent zugewiesen werden, das hierarchisch und damit auch syntaktisch am tiefsten eingebettet ist, wie z.B. das Verb *verkauft* in (7). Mit Hilfe der POS-Tags läßt sich der Status der infiniten Verben bestimmen. Aufgrund der jeweiligen Status aufeinanderfolgender Verben und deren Anzahl lassen sich Rückschlüsse auf ihre hierarchische Anordnung ziehen. So sind z.B. in einer Struktur wie in (8) die Verben in ihrer hierarchischen Struktur angeordnet, mit dem hierarchisch am tiefsten eingebetteten Verb *verkauft* an erster Stelle. Eine solche hierarchische Struktur läßt sich für die Folge *sollen verkauft werden* in (7) mit großer Sicherheit ausschließen, weil das erste Verb *sollen* finit ist.

(7) Die L*H Dasawerke in L*H Laupheim - H*L Peisenberg - und L*H Speyer - sollen H*L verkauft werden. %

(8) ..., weil die Dasawerke in Laupheim, Peisenberg und Speyer verkauft werden sollen.

[1] Akzente der angegebenen Akzentkategorien wurden jeweils auf dem folgenden Wort realisiert.

4 Intonationsbestimmung mit Hilfe eines Parsers

4.1 Der Parser

Wir verwenden einen robusten stochastischen Parser, dessen Grammatik aus einem kontextfreien Regelkern mit lexikalisierten Kopfwahrscheinlichkeiten besteht (*Head-Lexicalized Probabilistic Context-Free Grammar*, vgl. Carroll und Rooth (1998)).
Die Grammatik wurde ursprünglich entwickelt, um deutsche Verb-letzt-Sätze auf Subkategorisierungsrahmen zu untersuchen (Beil, Carroll, Prescher, Riezler, Rooth, 1999). Um unbeschränkten Text parsen zu können, wurden neben den handgeschriebenen Regeln weitere Robustheitsmaßnahmen zur Realisierung der Abfolge phrasaler Kategorien eingeführt. Dadurch werden zusätzlich zu Verb-letzt-Sätzen auch alle Präpositonal- und Nominalphrasen mit ihrer internen Struktur erkannt.
Aus dem mehrere Analysen umfassenden Parse-Wald wird die wahrscheinlichste ausgewählt (*Viterbi*-Parse). Als Beispiel ist in (9) die geklammerte Baumstruktur einer Analyse angegeben.[2] Trainiert wurde auf einem Zeitungskorpus mit ca. 40 Millionen Wörtern.

(9) $[_{X}[_{Y}[_{NN1}[_{NN}$Radio] $[_{NE}$Sarajevo]] $[_{X}[_{Y}[_{VPA}[_{VVFIN}$berichtete]] $[_{X}[_{Y}[_{ADV1}[_{ADV}$heute] $[_{ADJ}$früh]] $[_{X}[_{Y}[_{KOMMA}$,] $[_{X}[_{Y}[_{NP}[_{NC}[_{ART}$die]$[_{NN}$Fronten]]] $[_{X}[_{Y}[_{VSFIN}$seien] $[_{X}[_{Y}[_{VPPP}[_{VVPP}$stabilisiert] $[_{VWPP}$worden]] $[_{X}[_{Y}[_{PER}$.]]]]]]]]]]]]]]

4.2 Phrasierung

Konstituentengrenzen. Aus dem Syntaxbaum wird für jedes Wort die Anzahl der danach endenden Teilbäume berechnet, relativ zu der Anzahl der vor dem Wort beginnenden Teilbäume. In der geklammerten Struktur entspricht das der Differenz zwischen schließenden Klammern nach einem Wort und öffnenden Klammern davor. Die sich für (9) ergebenden Indizes sind in (10) nach den Wörtern eingefügt.

(10) Radio -3 Sarajevo 1 berichtete -2 heute -3 früh 1 , die -4 Fronten 2 seien -2 stabilisiert -3 worden 1 . [3]

[2] Die Kategorien orientieren sich am STTS-Tagset, s. Schiller et al. (1995). Darüberhinaus bezeichnet NC hier Noun Chunks, X und Y sind zusätzliche Elemente, die für die robuste Realisierung der Abfolge benötigt werden.

[3] Die Indizes für Interpunktion sind nicht angegeben, da sie hier nicht relevant sind.

Je nach Position innerhalb des Teibaums ergibt sich ein verschieden starker Grenzindex. Je mehr Konstituenten nach einem Wort enden, desto höher ist er.
Tabelle 1 zeigt, wie sich für alle im Nachrichtenkorpus vorkommenden Wörter der Grenzindex zu den danach gefundenen Phrasengrenzen verhält. Hierbei wird abstrahiert von der Kategorie der Grenze, selten vorkommende Indizes sind weggelassen.

Grenzindex	keine Grenze	Grenze	Verhältnis
-7	41	1	0.023
-6	151	6	0.038
-5	271	6	0.021
-4	770	12	0.015
-3	418	32	0.071
-2	2497	617	0.198
-1	837	45	0.051
0	97	1	0.01
1	396	188	0.321
2	339	349	0.507
3	264	328	0.554
4	124	168	0.575
5	68	138	0.669
6	22	54	0.71

Tabelle 1: Die Grenzindizes, die für jedes Wort im Korpus berechnet wurden (1. Spalte), aufgeteilt nach den tatsächlich vorgefundenen Phrasengrenzen (Spalten 2 und 3). Spalte 4 gibt das Verhältnis von Grenze/(Nichtgrenze + Grenze) an.

Es zeigt sich eine deutliche Korrelation zwischen diesem Grenzindex und dem Verhältnis der tatsächlich im Korpus gefundenen prosodischen Phrasengrenzen (Korrelationskoeffizient $r = 0.919$. Bei Berücksichtigung der unterschiedlichen Zahl der Fälle pro Index ergibt sich eine gewichteter Koeffizient von 0.911).

Nichtrealisierte Grenzen. Bislang sind also alle Wörter mit einem Grenzindex > 0 mögliche Kandidaten für eine folgende Phrasengrenze. Nicht alle Grenzen dürfen aber realisiert werden: Da sich selbst für Wörter mit hohem Grenzindex (> 2) nur bei 59% der Daten Phrasengrenzen fanden, werden für die Phrasierung noch weitere Merkmale herangezogen: Bei direkt aufeinanderfolgenden Phrasengrenzen wird nur die letzte realisiert, ebenso wird vor von Interpunktion gefolgten Wörtern keine Grenze zugewiesen (vgl. „phonologische Schwächung“ in Abschnitt 3.2).

Unbekannte Wörter. Zusätzlich wurde eine Grammatik-spezifische Heuristik gebildet, die auf den beim Training nicht aufgetretenen Wörtern und Wortformen basiert. Diese als UNTAGGED markierten Elemente werden in der Grammatik wie phrasale Komponenten nebeneinander realisiert:

(11) ...[$_{\text{NC}}$der [$_{\text{NN1}}$ehemalige SS-Offizier]] [$_{\text{UNTAGGED}}$Priebke] in Italien eingetroffen.

In der Mehrzahl der Fälle sind diese Wörter Eigennamen, die dem vorangehenden Noun Chunk zuzuordnen sind. Deshalb werden zwischen Noun Chunk und UNTAGGED-markierten Wörtern auftretende Grenzen nicht realisiert.
Auf die konkrete Kategorie der vorhergesagten Phrasengrenze (% oder -) wird hier kein besonderer Wert gelegt. Man könnte den Unterschied u. a. von der Phrasen-*Länge* abhängig machen. In der Implementierung des Algorithmus wurde hier nur die Interpunktion berücksichtigt: Bei Interpunktion wurde % zugewiesen, sonst -.

4.3 Akzentuierung

Durch den Einsatz eines stochastischen Parsers erschließen sich weitere Möglichkeiten für die Akzentvorhersage.

Wortklassen. Die Analyse des Korpus ergab, daß es einen Zusammenhang zwischen Wortklasse und Akzenthäufigkeit gibt: Die meisten Wortklassen sind entweder sehr häufig oder sehr selten akzentuiert. Zu den akzentuierbaren Einheiten gehören u. a. sämtliche nominalen Wortklassen (hier sind 70.3% akzentuiert) und vom Training unbekannte Wörter (meist Eigennamen, 86.5% akzentuiert).

Wortwahrscheinlichkeiten. Die lexikalischen Wahrscheinlichkeiten gegeben die Wortklasse P(Wort|POS) können auch zur Akzentuierung herangezogen werden. Sie sind in Tabelle 2 für die im Korpus vorkommenden Wörter angegeben.
Die relative Häufigkeit, mit der „unwahrscheinlichere“ Wörter akzentuiert sind, nimmt nahezu monoton zu: Korrelationskoeffizient $r = 0.966$ für $[-log(\text{lex. Wahrscheinlichkeit}), \text{Akzentverhältnis}]$; gewichtet nach der Zahl der Wörter pro Wahrscheinlichkeitsklasse ist $r = 0.973$.

Zur endgültigen Akzentbestimmung bekommen also a) zunächst diejenigen Wörter einen Akzent, die obigen Wortklassen zugeordnet werden können. Da die Akzentbestimmung nach der Phrasierung erfolgt, muß noch beachtet werden, daß in jeder Phrase min-

P(W|POS)	kein Akzent	Akzent	Verhältnis
< 1	2640	44	0.016
< 0.1	382	24	0.059
< 0.01	401	29	0.067
< 0.001	250	107	0.299
< 0.0001	264	162	0.38
$< 10^{-5}$	210	234	0.527
$< 10^{-6}$	246	318	0.563
$< 10^{-7}$	227	259	0.532
$< 10^{-8}$	100	215	0.682
$< 10^{-9}$	54	114	0.678
$< 10^{-10}$	34	69	0.669
$< 10^{-11}$	15	53	0.779
$< 10^{-12}$	35	121	0.775
$< 10^{-13}$	7	48	0.872
$< 10^{-14}$	3	29	0.906

Tabelle 2: Lexikalische Wahrscheinlichkeiten der Wörter im Korpus (1. Spalte), bezogen auf das Vorhandensein eines Akzents (Spalten 2 und 3). Das Verhältnis Akzent/(Nichtakzent + Akzent) zeigt Spalte 4.

destens ein Akzent realisiert wird. Wenn also in einer Phrase kein Wort mit einer solchen Wortklasse vorkommt, erhält b) dasjenige Wort den Akzent, das innerhalb der Phrase die niedrigste Wortwahrscheinlichkeit hat.

Die Kategorie des Akzents wurde hier nicht berücksichtigt. Dazu werden u. a. semantische Methoden benötigt, da mit der Akzentform oft eine bestimmte Bedeutung verbunden wird. Als Heuristik wurde in der vorliegenden Implementierung H*L dem letzten Wort im Satz zugeordnet, und Phrasen-internen Wörtern. Bei Phrasen-finalen Wörtern wurde L*H vergeben.

5 Evaluierung und Diskussion

Integriert in ein TTS-System liefern beide Ansätze eine sehr natürlich klingende Intonation, ohne sich perzeptiv wesentlich zu unterscheiden.

Zur Evaluierung wurde jeweils das komplette Korpus synthetisiert. Dabei wurden die Merkmale $\pm$*akzentuiert* bzw. $\pm$*Phrasengrenze* wortweise zugewiesen. Für beide Ansätze wurden *Precision* und *Recall* berechnet. Diese Metriken beziehen sich jeweils auf korrekte Zuweisungen von Grenzen bzw. Akzenten. Precision (korrekte Zuweisungen/(korrekte Zuweisungen + falsche Zuweisungen)) ist ein Maß für die Korrektheit des Modells. Recall (korrekte Zuweisungen/(korrekte Zuweisungen + Nicht gefundene Grenzen bzw. Akzente)) gibt ein Maß für die Abdeckung des Modells an.

	parserbasiert	taggerbasiert
Akzentuierung		
Precision	74.0%	80.5%
Recall	67.3%	67.2%
Phrasierung		
Precision	65.5%	82.2%
Recall	86.2%	70.5%

Tabelle 3: Parser- und taggerbasierte Modellierung im Vergleich. Bei der Akzentuierung ist die taggerbasierte Modellierung deutlich genauer, während sich bei der Phrasierung die hohe Precision durch niedrigeren Recall ausgleicht.

Beim Vergleich der Ansätze zeigt sich, daß die Precision beim taggerbasierten Ansatz insgesamt höher ist. Während bei der Phrasierung die hohe Precision des taggerbasierten Ansatzes zu Lasten der Abdeckung geht, liegt bei der Akzentuierung eine echte Verbesserung vor.

Man kann sagen, daß sich mit Hilfe eines POS-Taggers und einer darauf aufsetzenden linguistischen Analyse Ergebnisse erzielen lassen, die mit einer parserbasierten Intonationsbestimmung vergleichbar sind bzw. diese unter Umständen übertreffen können.

Der Grund könnte darin liegen, daß, wie Abney (1995) postuliert, nicht die syntaktische Struktur, sondern die einfachere Chunkstruktur mit der prosodischen Struktur korreliert. Die Chunkstruktur besteht aus kleineren Einheiten, hauptsächlich aus Nominalphrasen und Präpositionalphrasen, die zum großen Teil auch mit Hilfe der POS-Tags erkannt werden können. Das spricht zusätzlich für den Einsatz einer speziell für die Bedürfnisse von Text-to-Speech-Synthese zugeschnittenen linguistischen Analyse, die auf einem Tagger basiert.

Aus demselben Grund erwarten wir, daß eine komplexere Grammatik als die für den Parser verwendete robuste Grammatik keine signifikant besseren Ergebnisse bei der Phrasierung liefern würde. Eine Verbesserung erwarten wir jedoch durch die veränderte Behandlung unbekannter Wörter.[4] Außerdem sind für die Akzentuierung noch nicht alle Möglichkeiten ausgeschöpft, hier könnte z. B. die Berücksichtigung der Köpfe phrasaler Konstituenten weitere Informationen liefern.

[4]Die Grammatik und der Parser werden weiter entwickelt. Für den aktuellen Stand sei auf die Homepage der Theoretischen Computerlinguistk am IMS verwiesen: `http://www.ims.uni-stuttgart.de/tcl/`

Abschließend ist anzumerken, daß das Gesamtergebnis sowohl beim taggerbasierten als auch beim parserbasierten Ansatz umso höher zu bewerten ist, als aufgrund der natürlichen Variabilität von Intonation auch bei der Reproduktion einer Nachricht durch denselben Sprecher keine vollständige Übereinstimmung erreicht wird. So kann im Korpus z.B. beobachtet werden, daß Sprecher dieselbe Nachricht zu zwei verschiedenen Uhrzeiten mit unterschiedlicher Phrasierung oder Akzentuierung vorlesen. Bei der Berechnung von Precision und Recall wird eine natürlich wahrgenommene Intonation jedoch nur dann als korrekt gewertet, wenn sie mit der konkret geäußerten Intonation genau übereinstimmt. Eine Unterscheidung dieser Fälle von tatsächlichen Fehlern kann bei keiner Metrik erfaßt werden.

Literatur

Abney, S. (1995). Chunks and dependencies: Bringing processing evidence to bear on syntax. In *Computational Linguistics and the Foundations of Linguistic Theory, CSLI*. Stanford, Kalifornien.

Bech, G. (1983). *Studien über das deutsche verbum infinitum* (2. Auflage). Tübingen, Niemeyer.

Beil, F., Carroll, G., Prescher, D., Riezler, S., Rooth, M. (1999). Inside–outside estimation of a lexicalized PCFG for German. In *37th Annual Meeting of the ACL*. College Park, Maryland.

Black, A. W., Taylor, P., Caley, R. (1999). *The Festival Speech Synthesis System* (1.4 Auflage). University of Edinburgh.

Carroll, G. Rooth, M. (1998). Valence induction with a head-lexicalized PCFG. In *3rd Conference on Empirical Methods in Natural Language Processing*. Granada.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry, 24*(2), 239–297.

Culicover, P. W. Rochemont, M. (1983). Stress and focus in English. *Language, 59*(1), 123–165.

Grewendorf, G., Hamm, F., Sternefeld, W. (1989). *Sprachliches Wissen* (3. Auflage). Suhrkamp.

Haase, M. (1999). Aspekte einer Syntax-Prosodie-Schnittstelle. Studienarbeit, Universität Stuttgart.

Haider, H. (1993). *Deutsche Syntax - generativ*. Gunter Narr Verlag, Tübingen.

Hirst, D. (1993). Detaching intonational phrases from syntactic structure. *Linguistic Inquiry, 24*(4), 781–787.

Mayer, J. (1995). Transcription of German intonation - the Stuttgart system. Technical Report, Universität Stuttgart.

Möhler, G. (1998). *Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese*. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung. Universität Stuttgart.

Möhler, G. (1999). The German Festival system. http://www.ims.uni-stuttgart.de/phonetik/synthesis/.

Rapp, S. (1998). *Automatisierte Erstellung von Korpora für die Prosodieforschung*. Dissertation, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.

Schiller, A., Teufel, S., Stöckert, C. (1995). *Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS*. Universität Stuttgart.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to german. Revidierte Version eines Vortrags beim EACL SIGDAT Workshop, Dublin.

Schweitzer, A. (1999). Bestimmung der Intonation mit Hilfe von Wortklassen. Diplomarbeit, Universität Stuttgart.

Selkirk, E. O. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. MIT Press, Cambridge, Massachusetts.