

Annotationsrichtlinien DIMA – Konsenstranskription

Frank Kügler & Stefan Baumann

1. Ziel

Das Ziel dieser Konsenstranskription ist ein weitgehend theorie-neutrales Basistranskript für die Annotation von gesprochenem Deutsch. Dieses System basiert auf den Grundannahmen des autosegmental-metrischen Ansatzes zur Analyse von Intonation (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Gussenhoven, 2004; Ladd, 2008) und stützt sich auf phonetische Kriterien (*“phonetically informed phonological transcription”*). Damit soll eine Vergleichbarkeit von Daten/Korpora unterschiedlicher Arbeitsgruppen im deutschsprachigen Raum gewährleistet werden. Zwei Kriterien soll dieses System erfüllen: (i) Es soll leicht erlernbar bzw. trainierbar und damit transparent sein sowie eine hohe Reliabilität zwischen den Annotatoren, insbesondere zwischen Arbeitsgruppen erzielen. (ii) Das Basistranskript soll grundsätzlich (automatisch) in die phonologischen Systeme der verschiedenen Intonationsmodelle übersetzbar sein (vgl. z.B. Féry, 1993, 2012; Mayer, 1995; Grice, Baumann & Benz Müller, 2005; Peters, 2005, 2014), d.h. die phonologische Interpretation der Daten erfolgt zu einem späteren Zeitpunkt als die Annotation. Somit stellt die Konsenstranskription eine phonetisch orientierte Repräsentation der intonatorischen Oberflächenkontur dar, die jedoch phonologische Kriterien mit einbezieht.

2. Annahmen und Voraussetzungen

Die Durchführung der Annotation wird hier in Praat (Boersma & Weenink, 2013) illustriert. Für die Annotation gesprochener Sprache nehmen wir drei Ebenen der prosodischen Annotation sowie eine zusätzliche Ebene für Kommentare an. Die drei Ebenen der Prosodie kennzeichnen jeweils unabhängig voneinander Phrasengrenzen, Töne und die relative Prominenz von Akzenten. Darüber hinaus soll die Textebene mindestens eine Wortebene und eine Silbenebene beinhalten (vgl. Abbildung 1). Die Annotation der Tonebene und der Prominenzebene erfolgen grundsätzlich mit Bezug auf Silben.

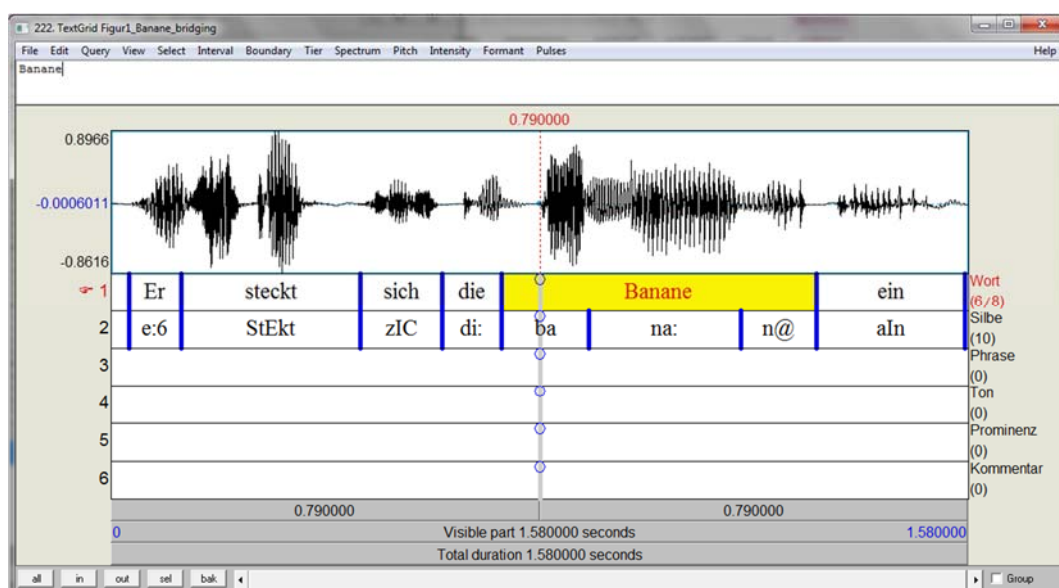


Abbildung 1. Ebenen der prosodischen Annotation nach DIMA. Die einzelnen Ebenen („tiers“) sind von oben nach unten folgendermaßen benannt: Wort, Silbe, Phrase, Ton, Prominenz, Kommentar.

Jede Ebene wird unabhängig von den anderen Ebenen annotiert. Die Unabhängigkeit der Ebenen hat den Vorteil, dass z.B. prosodische Grenzen und Prominenzen nicht in Bezug auf Töne annotiert werden müssen, sondern allein aufgrund ihrer phonetischen Eigenschaften und perceptiven Wahrnehmung als Grenze oder Prominenz. Grenzen müssen mit Tönen einhergehen, die allerdings erst in einem späteren Schritt annotiert werden. Ein Prominenzlabel dagegen, egal welcher Stufe, muss kein korrespondierendes Tonlabel zugewiesen bekommen (vgl. unten „Tonebene“).

Für die prosodische Annotation wird das Symbolinventar gemäß Tabelle 1 festgelegt. Die Symbole sind im Wesentlichen dem klassischen ToBI System (Beckman & Ayers-Elam 1997) entlehnt. Die Bedeutung und Verwendung der einzelnen Symbole wird im Folgenden erläutert und anhand von Beispielen illustriert.

Tabelle 1. Symbole für die prosodische Transkription sowie Kommentar.

Ebene	Symbole
Phrase	% - &
Diakritika	! ^
Ton	H* L* H L
Diakritika	! ^ < >
Prominenz	1 2 3
Kommentar	z.B.: Unterbrechung

2.1 Die Phrasenebene

Auf der Phrasenebene werden zwei Phrasentypen sowie ein Symbol für Disfluenzen unterschieden:

- prosodische Phrase mit starker Grenze: %
- prosodische Phrase mit schwacher Grenze: -
- Disfluenz: &

Aus Sicht der prosodischen Hierarchie (Selkirk, 1984; Nespor & Vogel, 1986) nehmen wir an, dass eine prosodische Phrase mit schwacher Grenze von einer prosodischen Phrase mit starker Grenze dominiert wird. Grundsätzlich postulieren wir damit zwei Ebenen der Phrasierung, die jedoch so theorie-neutral wie möglich über die relative Stärke der Grenze ausgedrückt werden. Phonetisch-perzeptive Kriterien für eine Grenze sind eine Pause, phrasenfinale Längung, eine phrasenfinale tonale Bewegung, Tonhöhen-Reset sowie weitere segmentale Phänomene wie Laryngalisierung oder *creaky voice*. Je nach Ausprägung dieser Kriterien wird eine starke oder schwache Grenze annotiert.

Grundsätzlich wird der Beginn und das Ende einer prosodischen Phrase transkribiert. Dies gilt für beide Phrasentypen. Wenn eine Pause zwischen den Phrasen vorhanden ist, werden zwei Phrasenlabel, jeweils eins für das Ende der ersten und den Beginn der zweiten Phrase annotiert. Wenn keine Pause vorhanden ist, wird nur *ein* Phrasenlabel annotiert, das zugleich das Ende der ersten und den Beginn der zweiten Phrase markiert. Abbildung 2 illustriert das Vorkommen zweier Phrasentypen in einer Äußerung.

Neben der weitgehend linguistisch determinierten Einteilung in kleine und große prosodische Phrasen können in gesprochener, insbesondere spontaner Sprache Disfluenzen vorkommen, die sich in eine Phrase integrieren oder eine Phrase unterbrechen, z.B. Häsitationen wie ‚ähm‘. Für die Markierung von Disfluenzen wird das sogenannte ‚Kaufmanns-Und‘ (&) verwendet. Dieses Label wird an der Grenze einer Disfluenz gesetzt und kann sowohl am Anfang als auch Ende der Konstituente vorkommen. Eine Disfluenz

muss nicht notwendigerweise eine eigene prosodische Phrase konstituieren. Abbildung 3 illustriert die Verwendung des Labels & bei einer Disfluenz.

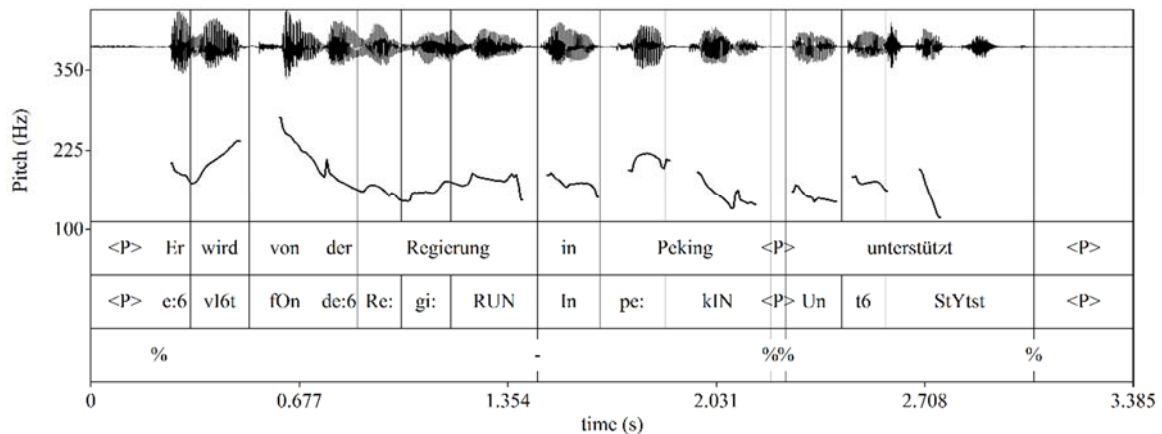


Abbildung 2. Illustration von starken und schwachen Phrasengrenzen, Beispiel „Peking“. Starke Grenzen werden hier jeweils vor und nach Pausen markiert, die schwache Phrasengrenze tritt an einer Stelle ohne Pause auf und wird daher nur einfach markiert.

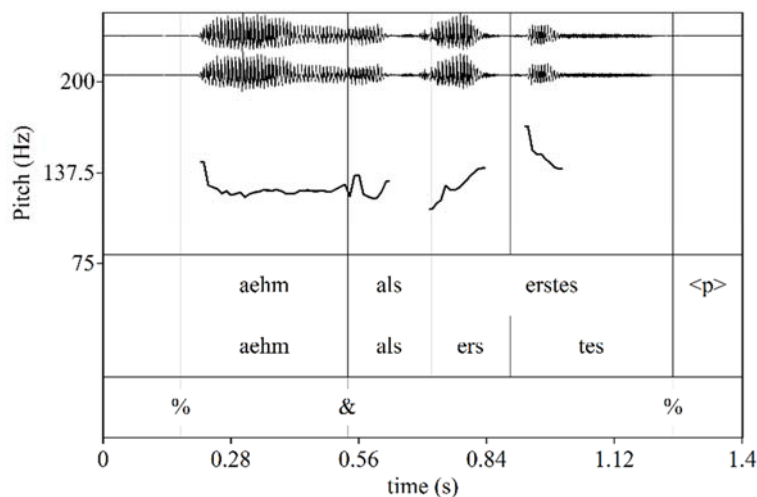


Abbildung 3. Illustration einer Disfluenz, die mit dem Label & auf Phrasenebene markiert wird.

Diakritika: Für Phrasen werden die Diakritika Downstep „!“ und Upstep „^“ verwendet, um phrasalen Downstep oder Upstep zu markieren (vgl. z.B. van den Berg, Gussenhoven & Rietveld, 1992). Damit erlaubt DIMA explizit, unterschiedliche Tonhöhenregister zu erfassen. Hier können die tonalen Relationen innerhalb einer Phrase erfasst werden, so dass die absolute F0 eines H* niedriger sein kann als die eines L* in einer heraufgestuften Phrase.

2.2 Die Tonebene

Die Tonebene differenziert zwischen Akzenttönen und Nicht-Akzenttönen. Grundsätzlich wird, wie im klassischen Ansatz nach Pierrehumbert (1980), nur zwischen zwei phonologischen Tonebenen unterschieden, H und L, die jeweils relativ zueinander als hoch

und tief interpretiert werden; d.h. bezüglich der absoluten F0-Werte kann ein Tiefton höher sein als ein Hochton in einer Äußerung, jedoch nicht in unmittelbarer Nachbarschaft eines Hochtons. Die folgenden Symbole auf Tonebene werden in DIMA verwendet.

- Akzenttöne sind durch das Diakritikum „*“ gekennzeichnet: (H* / L*)
- Die Nicht-Akzenttöne tragen zunächst kein Diakritikum: (H / L)
- Diakritika:
Zusätzlich wird Downstep oder Upstep an Akzenttönen oder Nicht-Akzenttönen mit den Diakritika „!“ und „^“ ausgedrückt. Downstep und Upstep sind relational zum vorangehenden Ton gleicher Qualität zu interpretieren (!H* / !H / ^H* / ^H).
- Für den Fall, dass der tonale Zielpunkt eines Akzenttons nicht in der prominenten Silbe liegt, werden Diakritika an dem Zielpunkt in der vorangehenden oder nachfolgenden Silbe gelabelt:
 „<“ Zielpunkt erscheint in der nachfolgenden Silbe und signalisiert Zugehörigkeit zum *vorhergehenden* Ton: (H* <)
 „>“ Zielpunkt erscheint in der vorhergehenden Silbe und signalisiert Zugehörigkeit zum *nachfolgenden* Ton: (> H*)

Abbildung 4 illustriert das Zusammenspiel von Akzenttönen und Nicht-Akzenttönen. Tiefe Töne erscheinen in diesem Beispiel zu Beginn und am Ende von Phrasen mit starker prosodischer Grenze. Der erste Akzentton (auf *wird*) ist perzeptiv ein hoher Ton (H*), dessen F0-Maximum erst in der nachfolgenden Silbe (*von*) erreicht wird, ausgedrückt durch „<“. Außerdem stehen zwei spätere Hochtöne in einer Downstep-Relation zu einem vorangehenden H. Dies betrifft sowohl Akzenttöne (auf *-stützt*) als auch Nicht-Akzenttöne (an der Grenze nach *Regierung*).

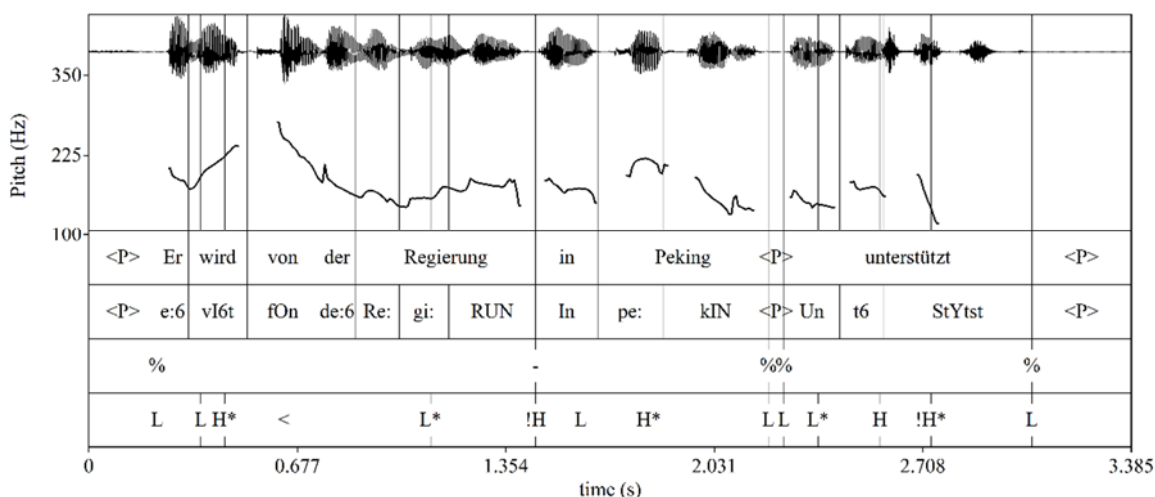


Abbildung 4. Illustration der Tonebene, Beispiel „Peking“. Akzenttöne werden mit „*“ gekennzeichnet. Nicht-Akzenttöne werden an phonologischen Grenzen gesetzt und erscheinen an F0-Wendepunkten zwischen Grenzen und Akzenttönen sowie zwischen Akzenttönen.

2.3 Die Prominenzebene

Auf der Prominenzebene werden drei Stufen der perzeptiven Prominenz unterschieden (vgl. Kieler Intonationsmodell KIM; Kohler, 1991), die auf Silbenebene annotiert werden. Nicht-

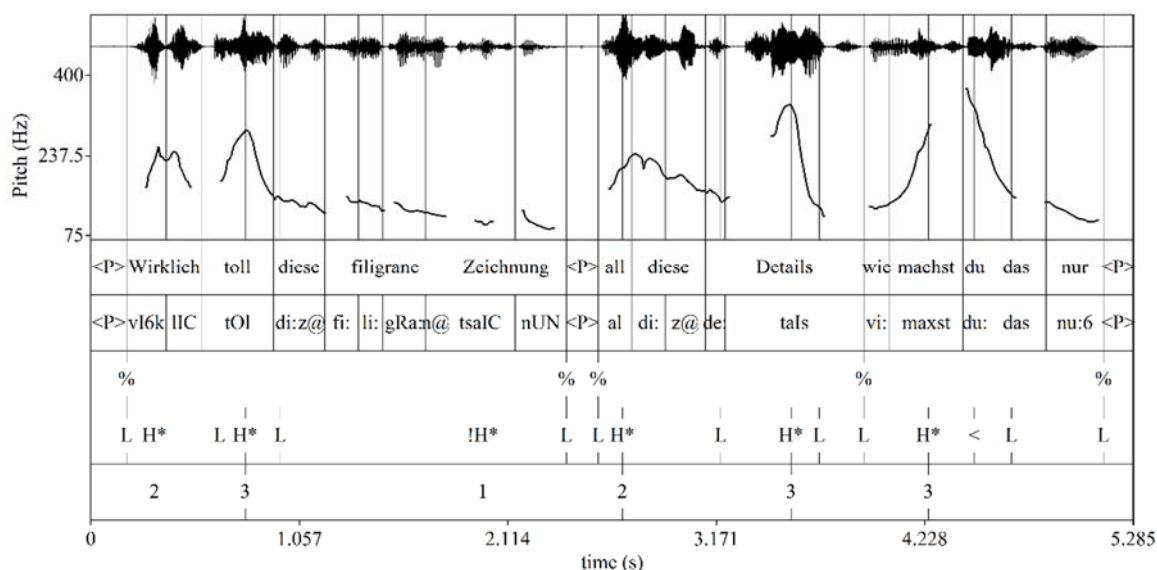


Abbildung 6. Illustration der Prominenzstufe 3, die mehrfach in dieser Äußerung vorkommt. Beispiel „Details“.

2.4 Die Kommentarebene

Diese Ebene ist für Kommentare vorgesehen. Hier können z.B. unterschiedliche Disfluenzen als Phänomene der gesprochenen Sprache wie etwa Redeflussstörungen, Häitationen, Unterbrechungen, Abbrüche o.ä. gekennzeichnet werden; auf Phrasenebene sollte dies mit der Markierung „&“ einhergehen. Außerdem können hier Unklarheiten oder Unsicherheiten mit Hilfe des Fragezeichens „?“ an den entsprechenden Grenz-, Ton- oder Prominenzlabeln sowie Kommentare aller Art markiert werden.

3. Grundsätzliches Vorgehen der Annotation (Annotationsprozess)

Die prosodische Annotation erfolgt in mehreren Schritten und jeweils eigenständig auf den drei unterschiedlichen Ebenen. Auf allen Ebenen wird von links nach rechts gelabelt. Grundsätzlich empfohlen wird, zuerst die Phrasenebene, als zweites die Prominenzebene und als letztes die Tonebene zu annotieren. Dieses Vorgehen lenkt die Aufmerksamkeit auf einzelne Aspekte des komplexen prosodischen Ereignisses und erleichtert somit z.B. die Entscheidung über die An- bzw. Abwesenheit einer Grenze oder der Prominenzstufe unabhängig vom tonalen Wert einer Silbe. Außerdem soll auf Phrasenebene begonnen werden, weil die Einteilung in prosodische Phrasen mit starker Grenze die Domäne für die weitere Annotation definiert. Dies hat zur Folge, dass weitere Grenzmarkierungen, Prominenzen und Töne immer phrasenweise annotiert werden.

- **Phrasenebene** – Identifizieren von Phrasengrenzen (vgl. Abbildung 2)
 - Beginn und Ende einer prosodischen Phrase mit starker Grenze → %
 - Beginn und Ende einer prosodischen Phrase mit schwacher Grenze → -
 - Markierung von Disfluenzen → &

Hinweise zur Phrasenebene:

1. Grundsätzlich werden der Beginn und das Ende einer prosodischen Phrase annotiert. Sollte keine Pause zwischen beiden Phrasen auftreten, ist nur ein Grenzlabel zu verwenden, das gleichzeitig das Ende und den Beginn einer Phrase markiert.
2. Die Annahme zur hierarchischen Struktur der Phrasen impliziert, dass eine schwache prosodische Phrase immer innerhalb einer starken prosodischen Phrase vorkommt.
3. Disfluenzen können – müssen aber nicht – die Phrasierung unterbrechen. Grundsätzlich werden Disfluenzen mit „&“ annotiert und können zusätzlich auf der Kommentarebene näher klassifiziert werden.

- **Prominenzebene** – Identifizieren von Prominenzen und deren Stufe
 - Punkt-Label (*point tier* in Praat) im Vokal der prominenten Silbe einfügen und Prominenzstufe schwach (=1), stark (=2), extra-stark (=3) festlegen → z.B. |2
 - [[Hinweis: späteres, gleichzeitiges Verschieben zweier / mehrerer Punkt-Labels in Praat durch Klicken eines Punktes, ⇧Shift-Taste gedrückt halten und Punkte verschieben]]

Hinweise zur Prominenzebene:

1. Prominenzlabel können ohne Tonlabel vorkommen.
 2. In der Regel korrespondiert Prominenzstufe 2 mit einem Akzentton auf der Tonebene.
- **Tonebene** – Labeln der Akzenttöne und der Nicht-Akzenttöne von links nach rechts
 - Begonnen wird mit den Grenztönen. Darauf folgt jeweils der nächste Akzentton, in dessen linker und rechter Umgebung Nicht-Akzenttöne gelabelt werden können. Dieses iterative Vorgehen gewährleistet, dass Töne relativ zueinander bewertet werden, was z.B. Downstep- und Upstep-Relationen einschließt.

○ **Grenztöne**

- Als Orientierung dienen hier die Grenzen der Phrasenebene; Tonlabel unterhalb des Grenzlabels platzieren →

Phrase %
Ton L

Hinweise zu Grenztönen:

1. Grundsätzlich erhalten Grenzlabel einen korrespondierenden Ton auf der Tonebene. Repräsentiert ein Grenzlabel sowohl den Anfang als auch das Ende einer Phrase, können zwei Tonlabel vergeben werden, aber nur wenn diese sich in ihrem Tonwert unterscheiden; andernfalls wird nur ein Tonwert gelabelt.
2. In der Regel beginnt eine große prosodische Phrase mit einem tiefen Grenzton „L“. Dieser Wert wird als Standardwert verstanden und repräsentiert tiefe bis mittlere Phrasenanfänge. Wenn eine Phrase mit einer außergewöhnlich hohen F0 beginnt, wird ein hoher linker Grenzton „H“ festgesetzt – es sei denn, die Kontur kann durch einen hohen Ton (z.B. Akzentton) in der ersten Silbe der Phrase beschrieben werden.
3. Disfluenzlabel müssen nicht notwendigerweise mit einem Tonlabel korrespondieren.

Die nächsten beiden Schritte wiederholen sich bis zum Ende der Phrase:

○ **Akzenttöne**

- Als Orientierung dienen hier die Punkt-Label der Prominenzebene – allerdings muss nicht jede Prominenz mit einem Akzentton einhergehen!

- Ein Akzentton wird immer in der prominenten (wortbetonten) Silbe annotiert
 → H* / L*
 Das Tonlabel wird auf den Wendepunkt in der F0-Kurve (Maximum, Minimum) platziert, wenn dieser innerhalb der prominenten Silbe auftritt. Akzenttöne als Teil einer Plateaukontur haben nicht immer einen Wendepunkt; die Annotation erfolgt dann in der Mitte des Vokals der wortbetonten Silbe. Das Tonlabel wird in die Mitte des Vokals platziert, wenn das Maximum/Minimum außerhalb der prominenten (wortbetonten) Silbe auftritt.
 - Diakritika I: Die Diakritika werden an dem phonetischen Minimum/Maximum der F0 außerhalb der prominenten (wortbetonten) Silbe platziert (s. Abbildung 4, „wird“ – H* und „von“ – <). → < / >
Dabei zeigt das Label auf den zugehörigen Ton:
 Zielpunkt erscheint in der nachfolgenden Silbe → H* <
 Zielpunkt erscheint in der vorhergehenden Silbe → > H*
 - Diakritika II: Downstep „!“ und Upstep „^“ werden vor den Tönen notiert
 → z.B. !H* ^H*
- **Nicht-Akzenttöne**
- Zur Identifizierung dieser Töne dienen die Wendepunkte in der F0-Kurve vor und nach einem Akzentton → H / L (ggf. inklusive ! oder ^)

Hinweis zu Akzent- und Nicht-Akzenttönen:

1. Es werden keine mikroprosodischen Perturbationen annotiert; Zielpunkte ggf. unmittelbar vor oder nach Perturbationen setzen. Abbildung 7 illustriert eine mikroprosodische Perturbation im Wort, das den Akzentton trägt.

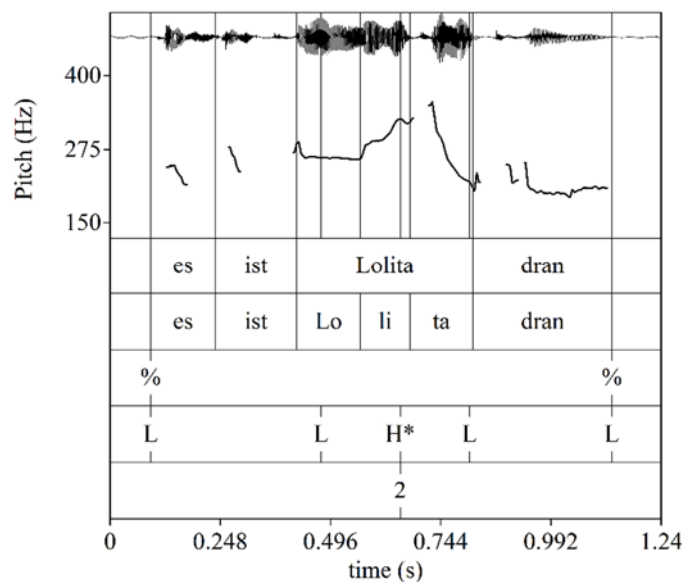
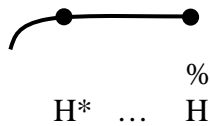


Abbildung 7. Illustration mikroprosodischer Perturbationen im Wort „Lolita“, das den Akzentton auf der wortbetonten Silbe „li“ trägt. Der Akzentgipfel wird in der wortbetonten Silbe annotiert, obwohl die F0 zu Beginn der folgenden Silbe „ta“ höher ist. Die höhere F0 ist nur durch den stimmlosen Plosiv hervorgerufen und wird deshalb nicht als Akzentgipfel in der nachfolgenden Silbe annotiert.

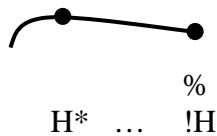
Sonderfall – Plateaukonturen:

Es gibt drei verschiedene Plateaukonturen, die in DIMA entsprechend annotiert werden können (vgl. für Plateaukonturen Niebuhr 2013, Peters 2014).

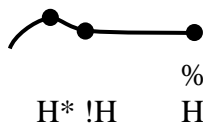
1a. Hochplateau, das auf gleicher, hoher F0 Ebene bleibt.



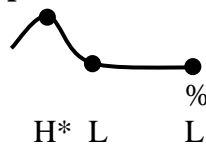
1b. In Abgrenzung zu einem Plateau gibt es eine Kontur mit gleichmäßig fallender F0 von einem hohen Akzentton bis zu einem herabgestuften finalen Grenzton; die Perception der Grenze ist eindeutig *nicht* tief.



2. Abgestuftes Hochplateau nach einem hohen Akzentton; F0 bleibt konstant auf abgestuften F0 Niveau bis zum Phrasenende.



3. Tiefplateau, das auf gleicher, tiefer F0 Ebene bleibt.



- **Kommentarebene**

Gegebenenfalls werden hier Kommentare wie z.B. „Unterbrechung“ oder andere Auffälligkeiten, die nicht vom Prosodieinventar erfasst werden, festgehalten. Unklarheiten oder Zweifel der Transkription sollen hier am entsprechenden prosodischen Label mit „?“ markiert werden.

Hinweis zu Zweifelsfällen: Resynthese durch Praat erzeugen und Unterschiede perceptiv identifizieren; ein Unterschied liegt vor, wenn potentiell eine andere Funktion wahrnehmbar ist (t' Hart, Collier & Cohen, 1990).

NB: Der **Nuklearakzent** korrespondiert in der Regel mit dem letzten Tonakzent vor einer Phrasengrenze, der ein Prominenzlabel 2 oder 3 enthält. DIMA verzichtet allerdings bewusst auf eine strikte Definition des Konzepts „Nuklearakzent“, um so theorie-neutral wie möglich zu sein. Jedoch lässt sich aus der Kombination von Prominenzebene und Tonebene schlussfolgern, welcher Akzentton dem eines Nuklearakzentes im klassischen Ansatz entspricht.

4. Übersetzbarkeit in phonologische Systeme

Es soll grundsätzlich möglich sein, die DIMA-Annotation mit Hilfe von Regeln in eines der o.g. phonologischen Systeme zu übersetzen. Beispiele zur Übersetzung von DIMA zu GToBI befinden sich bereits auf der DIMA-Homepage (<http://dima.uni-koeln.de>).

Literatur

- Beckman, M. E., & Ayers-Elam, G. (1997). Guidelines for ToBI Labelling, Version 3. Ohio State University. http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- van den Berg, R., Gussenhoven, C. & Rietveld, T. (1992). Downstep in Dutch - implications for a model. In: G. J. Docherty & D. R. Ladd (Eds) *Papers in Laboratory Phonology II* (pp. 335–359). Cambridge: Cambridge University Press.
- Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Retrieved from <http://www.praat.org/>
- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.
- Féry, C. (2012). Die Annotation der deutschen Intonation: Wie kann GToBI die Syntax und die Semantik integrieren? *Linguistische Berichte*, 229, 39–97.
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford: Oxford University Press.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- t' Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Kohler, K. J. (1991). A Model of German Intonation. In K. J. Kohler (Ed.), *Arbeitsberichte (AIPUK): Vol. 25. Studies in German Intonation* (pp. 295–360). Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Mayer, J. (1995). *Transcription of German intonation: the Stuttgart System*, Ms., University of Stuttgart.
- Niebuhr, O. (2013). Resistance is futile - The intonation between continuation rise and calling contour in German. *Proceedings of 14th InterSpeech*, Lyon, 225–229.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Berlin: Mouton De Gruyter.
- Peters, J. (2005). Intonation. In *Duden-Die Grammatik* (pp. 95–128). Mannheim: Dudenverlag [7. Auflage].
- Peters, J. (2014). Intonation. Heidelberg: Winter.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (PhD dissertation). MIT, Cambridge, Massachusetts.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.