

## **A system for transcribing talk-in-interaction: GAT 2**

**translated and adapted for English by  
Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten**

**Margret Selting, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzluft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, Susanne Uhmann<sup>1</sup>**

### *Abstract*

This article presents a revised version of GAT, a transcription system first developed by a group of German conversation analysts and interactional linguists in 1998. GAT tries to follow as many principles and conventions as possible of the Jefferson-style transcription used in Conversation Analysis, yet proposes some conventions which are more compatible with linguistic and phonetic analyses of spoken language, especially for the representation of prosody in talk-in-interaction. After ten years of use by researchers in conversation and discourse analysis, the original GAT has been revised, against the background of past experience and in light of new necessities for the transcription of corpora arising from technological advances and methodological developments over recent years. The present text makes GAT accessible for the English-speaking community. It presents the GAT 2 transcription system with all its conventions and gives detailed instructions on how to transcribe spoken interaction at three levels of delicacy: minimal, basic and fine. In addition, it briefly introduces some tools that may be helpful for the user: the German online tutorial GAT-TO and the transcription editing software FOLKER.

*Keywords:* Conversation analysis, transcription, prosody.

### *German Abstract*

Dieser Beitrag stellt die englische Übersetzung und Adaption der überarbeiteten Version des Transkriptionssystems GAT vor. GAT wurde 1998 von einer Gruppe deutscher Konversations- und Diskursanalytiker und interaktionaler Linguisten entwickelt. GAT versucht, so viele Prinzipien und Konventionen der in der CA genutzten Jefferson'schen Transkriptionsweise wie möglich zu übernehmen, schlägt jedoch einige Symbole vor, die auf die linguistische und phonetische, insbesondere die prosodische, Analyse gesprochener Sprache ausgerichtet sind. 10 Jahre später wurde GAT auf der Grundlage der bisherigen Erfahrungen mit seinem Gebrauch und vor dem Hintergrund neuer Anforderungen, die sich aus jüngeren technischen Entwicklungen sowie methodologischen Überlegungen heraus ergaben, überarbeitet. Dieser Beitrag macht GAT für die internationale Forschungsgemeinschaft zugänglich. Er beschreibt GAT 2 mit all seinen Konventionen und gibt detaillierte Anweisungen für die Transkription gesprochener Interaktion in drei Genauigkeitsstufen: Minimal-, Basis- und Feintranskript. Außerdem stellt er

---

<sup>1</sup> We would like to thank Thomas Schmidt for his contributions to the discussions that led to this text.

kurz für die Transkription hilfreiche Werkzeuge vor: das deutsche Online-Tutorial GAT-TO und den Transkriptionseditor FOLKER.

*Keywords:* Konversationsanalyse, Gesprächsanalyse, Diskursanalyse, Transkription, Prosodie.

1. Introduction
2. Principles of transcription
3. Transcription conventions (GAT 2)
  - 3.1. Parts of the transcript
    - 3.1.1. Transcript header
    - 3.1.2. General structure of the transcript
  - 3.2. The minimal transcript
    - 3.2.1. Segments and wording
    - 3.2.2. Sequential structure
    - 3.2.3. Other segmental transcription conventions
    - 3.2.4. Further conventions for the minimal transcript
  - 3.3. The basic transcript
    - 3.3.1. Prosody
    - 3.3.2. Further conventions for the basic transcript
  - 3.4. The fine transcript
4. Suggestions for the representation of more complex phenomena
  - 4.1. Autosegmental representation of intonation
  - 4.2. Notation of rhythm
5. Interlinear translations
6. Summary of the GAT 2 transcription conventions
7. Sample transcript of an excerpt from natural conversation
8. References

Appendix 1: The German internet tutorial GAT-TO

Appendix 2: The transcription editor FOLKER

## 1. Introduction

GAT – an acronym which stands for *Gesprächsanalytisches Transkriptionssystem* (discourse and conversation-analytic transcription system) – is a transcription system for notating, first and foremost, the wording and prosody of natural everyday talk-in-interaction. It can be used both for the compilation of working transcripts of talk-in-interaction for research purposes and for transcripts in linguistic publications.

A first version of GAT – mainly devised for German – was published ten years ago (see Selting et al. 1998). In the meantime, a range of discourse- and conversation-analytic publications in German-speaking countries and research communities have employed it for the transcription of data excerpts. This suggests that the original goal of establishing a unified system for the transcription of spoken interaction has been reached, at least among German-speaking scholars. Ten years later, recent developments in the fields of prosody and multimodality as well as the spread of Interactional Linguistics, new computational possibilities for recording and storing data and, last but not least, many years of experience in using GAT have made it advisable to adapt the original GAT (henceforth GAT 1) to new demands. GAT 2 (Selting et al. 2009) provides a revised version of GAT 1 which clarifies ambiguities and makes small amendments where shortcomings

have become evident in daily use. It contains all the conventions needed for producing publishable transcripts and can be read and appreciated without knowledge of the earlier system. The present paper is a translation and adaptation of Selting et al. (2009), which aims to make GAT 2 accessible for the English-speaking community.

What are the advantages of GAT 2?

- It can be used specifically for research in the fields of Conversation Analysis, Discourse Analysis and Interactional Linguistics.
- It is easily accessible for novices to transcription.
- It offers a simple level of transcription at the outset: the minimal transcript. The minimal transcript is sufficient for simply notating the wording of discourse, as is often done, for instance, in sociology and psychology.
- It is suitable for transcription with ordinary text processing software in which the output resembles the input as far as possible. For transcribing with specialized transcription editors, the conventions may have to be modified or stated more precisely. GAT 2 is intended to be compatible with transcription conventions for specialized editors (it defines a possible output), but does not aspire to defining these conventions.
- The minimal transcript in particular is compatible with sustainable, i.e. platform-independent, electronic transcription corpora and with efficient computerized search algorithms.<sup>2</sup> It permits the exact notation of breathing, pausing and lengthening. Interactional phenomena such as laughter, crying and overlap are represented in ways that are unambiguous and formalizable with transcription editors and other processing tools.
- The system includes suggestions for the notation of more complex phenomena in separate lines. For the detailed description of intonational phenomena, autosegmental representations are discussed. There are notation conventions for rhythm and explicit instructions for interlinear translations.

GAT 2 acknowledges the fact that the analysis of visual aspects in multimodal interaction is developing rapidly. In this area, both theory and technology are constantly changing and there are as yet no established conventionalizations. Moreover, the complexity of the field goes far beyond what is manageable in the limited space of this paper. For this reason, suggestions for the notation of visual aspects in multimodal interaction have been excluded here. Interested readers are referred to recent work published by Charles Goodwin (e.g. 2007a, 2007b, 2010), Lorenza Mondada (e.g. 2007a, 2007b, 2008) and Jürgen Streeck (e.g. 2003, 2009) (see also Heath/Hindmarsh/Luff 2010).

Apart from this, additional tools have been developed which are described in more detail in the appendix to this paper:

- On the basis of GAT 2, an online tutorial (GAT-TO) is now available in German. It introduces novices to the practice of transcribing and provides them

---

<sup>2</sup> "Efficient" in the sense of avoiding both underselection (i.e. neglecting relevant instances) and overselection (i.e. including irrelevant instances) to the extent possible (see, e.g., Edwards 1995).

with online learning modules on the challenging aspects of transcribing with GAT (<http://paul.igl.uni-freiburg.de/GAT-TO/>; see also the description in appendix 1).

- For transcribing with GAT, Thomas Schmidt (Hamburg) has developed a transcription editor, FOLKER, at the *Institut für Deutsche Sprache* (IDS, Mannheim), in the current version available with a user interface in German, English or French respectively. This editor automatically checks the validity of the transcript against the conventions for a GAT minimal transcript. FOLKER has been introduced in order to facilitate the compilation of a corpus of spoken German for research and teaching purposes (*Forschungs- und Lehrkorpus gesprochenes Deutsch* (FOLK)). The editor can be downloaded free of charge from the website of the archive of spoken German (*Archiv für Gesprochenes Deutsch* (AGD)) at the IDS (<http://agd.ids-mannheim.de>; see also the description in appendix 2). It can be used for the transcription of audio data and as a basis for importing into ELAN, the software most commonly used for video analysis.

In the age of the internet, GAT 2 would be unthinkable without its own homepage and links to the tools mentioned above (GAT-TO and FOLKER). This is the URL:

[http://agd.ids-mannheim.de/html/gat\\_en.shtml](http://agd.ids-mannheim.de/html/gat_en.shtml)

Finally, it should be stressed that a transcript is always dependent on the specific – and potentially variable – purposes of analysis and description. While GAT can be used as a guideline for transcription at all levels of granularity, from working transcript to refined version with elaborate detail, it is first and foremost intended to establish minimal standards for linguistic publications. Notation conventions for the latter are set out under the heading of *basic transcript*. Depending on researchers' specific purposes, the basic transcript can be reduced to the level of *minimal transcript* (for internal data sessions, for instance), or further expanded to the level of *fine transcript* and beyond, for special-purpose publications.

## 2. Principles of transcription

GAT has been developed on the basis of the following principles:<sup>3</sup>

- Expandability of the notation for increased granularity ("onion skin" or multi-layer principle): A transcript with low granularity should be expandable without revision to a higher degree of granularity.
- Readability of the transcript: The transcript should be understandable for linguists and non-linguists. This entails not using any special mode of representation for speech, such as, e.g., the phonetic alphabet. The latter may of course be added for specific research purposes.

---

<sup>3</sup> See the contributions in Edwards/Lampert (1993), which present a selection of discourse-analytic transcription systems in use at the time. For more recent suggestions, see Jefferson (2004), Schegloff (2007:265-269) and Du Bois (2006).

- Unambiguousness: The transcription system is intended to be explicit in providing for the representation of auditory phenomena at the specified level of granularity. Transcription symbols are defined unambiguously; each represents one (and only one) phenomenon.
- Iconicity: Transcription symbols are intended to be as non-arbitrary as possible, i.e. they should follow iconic principles.
- Relevance: The transcription system should permit the notation of those phenomena which previous research has shown to be relevant for the interpretation and analysis of verbal interaction.
- Form-based parametricization: The notation conventions aim for form-based transcription. That is, instead of interpretive comments such as e.g. "surprised", the specific parameters on which such interpretations are based should be represented individually on the basis of their form.

### **3. Transcription conventions (GAT 2)**

#### **3.1. Parts of the transcript**

##### **3.1.1. Transcript header**

When a transcript or part of a transcript is cited in a publication, it should be preceded by a transcript header containing the following metadata:

- source corpus and recording,
- indication of the beginning and end of the excerpt,
- where necessary, a short description of the interactional context in double round brackets (( )).

At times, transcripts are simplified for the sake of the argument (with conversational schisma, for instance, the irrelevant part of the conversation may be omitted). This kind of simplification should be mentioned in the transcript header.

If parts of the original transcript are omitted when it is reproduced in a publication, this should be signaled by ((...)). Larger deletions can be indicated as follows:

- 
- ((20 sec omission))
- 

##### **3.1.2. General structure of the transcript**

A GAT transcript represents the linguistic and non-linguistic actions and events in an audio or video recording in their temporal order. These can be linked to specific time frames in the recording by giving absolute time stamps. The latter provide hours, minutes and seconds, i.e. {1:01:04} for 1 hour, 1 minute and 4 seconds. Optionally, such a time stamp can be provided in a column to the left of each line of transcription.

Choosing an equidistant font such as Courier 10 pt, as has been done for the examples in this text, will minimize problems resulting from conversion between operating systems. The line spacing should normally be 1.5. Working with tabs should be avoided; instead, the appropriate number of blanks should be inserted. Separating words by hyphen is not allowed; in general, there is no hyphenation at all in GAT. The entire transcript is written in small letters. (Capital letters are needed in order to notate accents.)

In the running transcript, as in standard writing, the sequence of items on the screen or paper is an iconic reflection of the temporal sequence of events in real time. The usual direction of reading from left to right and from top to bottom captures – with only a few exceptions – the linear temporal sequencing of talk-in-interaction.

The transcript is subdivided into segments. The segments are numbered.<sup>4</sup> If more space is needed for notating a segment than is provided by a single line, the segment is continued on the next line. The latter, however, is usually indented and is not assigned a number. The number of the segment is followed (after three blanks) by the speaker ID.<sup>5</sup> Speaker IDs are not repeated in the following segment when the same participant continues. After three more blanks, the notation of the utterance itself follows. The basic transcript can be extended by adding further lines below the wording, e.g., in order to give a more detailed representation of prosody and/or visual phenomena and for translations into another language (these additional lines do not have numbering; see also section 5). An arrow '→' to the left of the line can be used to highlight a phenomenon of relevance for the analysis. Excerpt (1) illustrates the general structure of a transcript.<sup>6</sup> Other conventions will be explained in section 3.2.

**Example (1):** ((fictitious conversation, 1:04-2:05 sec))

	{1:04}	01	A:	this is where the transcript starts
		02	B:	yes exactly
		03		(--)
		04	A:	if you interrupt [me ]
→		05	B:	[i'm] not interrupting
				you
	{2:00}	06	A:	yes you are
		07		(.)
		08		you just did
		09		you just (.) interrupted me
		10	B:	okay
	{2:05}	11		i'm sorry

<sup>4</sup> Excerpts from a longer transcript reproduced in a publication start either with the relevant segment number of the original transcript or with the segment number 01.

<sup>5</sup> For unidentifiable speakers, IDs such as X1, X2, ..., or F1, F2, ... and M1, M2, ... for female and male speakers respectively can be used.

<sup>6</sup> In the following, we will work first with fictitious examples to illustrate the phenomena we would like to introduce. An example from natural everyday conversation will then be presented at the end of the paper.

A wide right-hand margin (at least 5 cm, or 2 in) will allow the transcriber to expand the minimal transcript into a basic transcript (see section 3.3).

### 3.2. The minimal transcript

The minimal transcript is the simplest form of discourse representation within GAT. While it will be sufficient for a range of purposes in the social sciences (such as content analysis in interviews), it is usually not detailed enough for research purposes in Conversation Analysis and Interactional Linguistics. In the context of the latter, the minimal transcript is first and foremost a working tool and is not intended for publication. It can be used, e.g., in determining which passages are to be notated in more detail. Example (1) (see above) shows such a minimal transcript.

In the minimal transcript, the wording of the participants' contributions and their segmentation into intonation phrases are notated (see sections 3.2.1.-3.2.2., 3.3.1.). The minimal transcript also includes an indication of overlaps, hesitation markers, pausing, inbreaths and outbreaths, laughter and non-linguistic actions and events as well as stretches of speech which are unintelligible (see sections 3.2.2.-3.2.4.). The segment endings are marked by line breaks.

In non-linguistic research areas, minimal transcripts may also be used in publications. When representing longer monologic passages, for instance in social science interviews, it may be advantageous to mark the segment boundaries by a vertical bar | (U+007C)<sup>7</sup> and to number the lines consecutively, in order to save space. It should be kept in mind, however, that this method makes it more difficult to expand the transcript at a later stage. Example (2) shows the continuous notation of segments.

#### Example (2):

```
01  INT:  and when did you last have this problem |
02  NAR:  well | uh mm | (0.35) | it was maybe about | i mean |
03        do you really | if you want to know an exact point in
04        time | (0.50) | i would say probably | let me think |
05        (1.23) | it might have been about thirteen months ago
06        | or so |
```

#### 3.2.1. Segments and wording

##### Spelling

In general, the actual wording of an utterance is notated according to the standard conventions of orthography for a given language, which are basically rules for the conversion of its sound segments into writing. Usually, the national standard variety of a language is taken as the norm, unless a regional variety is chosen by

<sup>7</sup> Brackets with "U+" and combinations of letters and ciphers following a sign indicate that sign's character encoding in the Unicode standard.

the participants and explicitly introduced as such in the transcript header or commentary. Where several national standard varieties exist, the one adopted by the participants should be specified in the transcript header. Pronunciation variants within the adopted norm are notated according to its standard orthography, i.e. they are not represented separately. Only departures from the adopted norm are captured as departures from its standard orthography.<sup>8</sup> For English this implies, for instance:

Received Pronunciation as norm<sup>9</sup>

/ka:/ [ka:]# = *car*  
           [ka:r]# = *carr*<sup>10</sup>

General American English as norm<sup>11</sup>

/kar/ [kar] = *car*  
           [ka:] = *ca'*<sup>12</sup>

### *Cliticizations*

With cliticizations, the cliticized word loses its phonetic integrity. Often, cliticizations involve several reductional processes at once, such as deletions and/or assimilations and/or reduction of full vowels and consonants. Frequently, such cliticizations occur with combinations of function words such as, e.g., personal pronoun + auxiliary or auxiliary + negative particle. In English, many of these have become conventionalized and are accepted in standard orthography. Conventionalized cliticizations are notated in English transcripts as follows:<sup>13</sup>

[PersPron + Aux]: *i'm* – i am  
                           *he's* – he is  
                           *they'll* – they will  
                           *we've* – we have  
                           *you'd* – you would, you had

<sup>8</sup> If more phonetic detail is needed in the transcript, this can be notated in a separate line, for instance according to the IPA conventions. But see also the section on modified orthography below.

<sup>9</sup> In Received Pronunciation postvocalic "r" is not pronounced when it occurs before a consonant or at the end of an utterance (indicated here with #).

<sup>10</sup> If the post-vocalic "r" rule is followed, the word would be written in standard orthography, e.g., *car*. If an RP speaker deviates from the norm by pronouncing a postvocalic "r" before a consonant or in utterance-final position, this would be captured in transcription by altering the standard orthography of the word, e.g., *carr*.

<sup>11</sup> In General American English "r", whether prevocalic or postvocalic, is always pronounced.

<sup>12</sup> If a General American speaker were to drop an "r", this departure from the norm would be captured in orthography by notating, e.g., *ca'* rather than *car*. An apostrophe can be used as a sign of omission. For the general use of the apostrophe in English GAT transcripts, see footnote 13.

<sup>13</sup> We consider the apostrophe a part of standard English orthography and have retained it in this representation of cliticization. For GAT transcribers who wish to use a transcription editor such as FOLKER, the use of an underscore instead of an apostrophe is recommended.

[Aux + NegPart]:     *don't* – do not  
                          *couldn't* – could not

[Aux + Aux]:         *would've* – would have  
                          *they'd've* – they would have

Less conventionalized in spelling, but still observable, are combinations of auxiliary + infinitival or prepositional 'to', such as:

*wanna* – want to  
*gonna* – going to  
*hafta* – have to

Similarly, content words + function words or other high-frequency word combinations can be cliticized. These have only rarely become conventionalized in English orthography, with the exception of:

*cuppa* – cup of (esp. BE)

In cases of doubt as to whether a particular cliticization is conventionalized, dictionaries such as the Longman English Dictionary Online for British English (BE) (<http://www.ldoceonline.com/>) and the Merriam-Webster OnLine for American English (AE) (<http://www.merriam-webster.com/dictionary/>) can be consulted.

Depending on the researcher's interest and the purpose of the transcript, several further levels of granularity for representing the wording of utterances are possible in GAT.<sup>14</sup> If one of these is chosen, it should be referred to in the transcript header.

#### *Modified standard orthography*

For transcribers of all languages in which there is a significant gap between standard orthography and casual pronunciation, a decision must be made as to how to represent actual phonetic realizations. It may also be desirable to convey the special flavor of spoken language or to capture the way words 'sound' for argumentative purposes. In these cases, modified standard orthography, often referred to as "eye dialect", may be used. "Eye dialect" has the advantage of being truer to the reality of spoken language but the disadvantage of (i) being difficult for non-native speakers to decipher, (ii) possibly caricaturizing the speakers as social types and (iii) compromising the accuracy of automatic searches.

For the transcripts presented here, the translators of GAT 2, after careful consideration, have opted for a moderate and systematic form of modified standard orthography. A comprehensive application of this can be found in the sample transcripts of "Sweet guy from Laguna" in section 7 of this paper. In deciding

---

<sup>14</sup> This and the following subsections depart from the German version of GAT 2 because the translators felt a particular need to adapt this section to the English-speaking audience and their presumed needs and backgrounds.

when and how to modify standard orthography in these transcripts we have been guided by the following principles:

- Use standard orthography where regular (obligatory) phonetic and phonological processes have applied:

*years* – not: *yearz* (line 10)<sup>15</sup>  
*political* – not: *polidical* (line 22)  
*abortion* – not: *aborshun* (line 35)  
*explain* – not: *iksplain* (line 48)  
*was* – not: *wuz* (line 66)

- Indicate noticeable, non-regular (optional) phonetic and phonological variants by modifying the orthography in such a way that the actual realizations will be retrievable:<sup>16</sup>

*twenny*<sup>17</sup> – twenty (line 10)  
*ril* – real (line 22)

- For deletions at the beginning or the end of a word, use an apostrophe to represent the omission:

*an'*, *'n'* – and (lines 8 and 65)  
*jus'* – just (line 6)  
*'at* – that (line 48)  
*don'* – don't (line 49)

- For sound and syllable deletions within the word, leave out the corresponding letter(s):

*evrything* – everything (line 19)  
*sd* – said (line 48)  
*cathlic* – catholic (line 41)

- Orthographic modifications notwithstanding, attempt to preserve the identifiability of the words in question.

For all modifications, the use of a controlled set of forms is recommended and a list of these forms used should accompany each transcript. This could include

<sup>15</sup> Line numbers here refer to the sample transcripts in section 7.

<sup>16</sup> Cliticized realizations can be notated depending on the degree of reduction, ranging from univerbation, apostrophization, linking with underscore '\_' to simple separation. Writing with an underscore rather than using univerbation or apostrophization improves the automatic searchability of cliticized items.

<sup>17</sup> Since representing some realizations can involve the use of stigmatized spellings (see also *unnerstand*, line 49), a disclaimer to the effect that the representations are not intended discriminatorily may be in order.

reduced forms frequently used in spoken language in their quasi-standardized spelling:

*cuz, yiknow, gotta, gimme, lotta*

#### *Other spelling variants*

More detailed phonetic spelling variants can be used if the transcript serves to make a particular phonetic point. As an illustration, see the following possible levels of granularity for segment 8 of the sample transcript in section 7:

a) Standard orthography

8 i and i just talked to this (.) asian guy

b) Modified standard orthography (moderate)

8 i an' i jus' talked to this (.) asian guy

c) "Eye dialect"

8 i en i jes' talk' to this (.) asian guy

Even more refined levels would be the partial and/or full use of IPA symbols to represent the entire utterance.

#### **Regional dialect forms**

When regional forms are relevant for the research question, a line with phonetic transcription can be added below the respective line of transcript. (Alternatively, a comment in the transcript header can indicate the variety used by one or several of the speakers.)

#### **Foreign-language words**

Foreign-language words are written as prescribed by the orthography of the original language. For example:

*gestalt*  
*angst*<sup>18</sup>

However, when the pronunciation diverges notably from the standard, this is notated. For instance, *gemutlich* (rounded high back vowel) instead of *gemütlich* (rounded high front vowel).

---

<sup>18</sup> German orthography requires initial capitalization of nouns; however, this is avoided in GAT, where capital letters are reserved for indicating primary and secondary accents.

## Compounds

In GAT, hyphenated compounds (*self-conscious*, *ice-cream*) are represented without a hyphen, as the latter is needed for the notation of intonation (see 3.3.1):

*selfconscious*  
*icecream*

## Abbreviations

When speakers use letters, abbreviations and acronyms, they are spelled out in their long form in GAT, so that accents can be notated where needed. Thus:

*jay ou en ee es* – j-o-n-e-s  
*yu en* – U.N.

## Numbers

Numbers are transcribed using words, i.e., for example,

*twenty two*  
*one hundred and ten*  
*one point three*  
*one and a half*  
*twelve twenty ei em*

## Other

Extensions of the alphabet which are not used in the language in question are not notated in transcription: *a la carte*, *a propos* (not *à la carte*, *à propos*); *aperitif*, *apres ski* (not *apéritif*, *après ski*). If necessary, a comment can be provided, such as <<French pronunciation> > (see 3.2.3.).

### 3.2.2. Sequential structure

#### *Speaker contributions*

A speaker contribution includes all linguistic and non-linguistic events which immediately follow one another and which are produced by that one speaker. A speaker contribution can be made up of one or several segments. GAT places each segment on a separate line. A segment can consist of :

- an intonation phrase<sup>19</sup> (see section 3.3.1.) including initial or segment-internal delay (e.g. 'stuttering')
- an intonation phrase and elements prosodically dependent on it (incremented or prefaced) (see section 3.2.3.)
- an incomplete intonation phrase
- a pause
- a visible bodily action

---

<sup>19</sup> The term 'unit of phrasing' (*Phrasierungseinheit*) used in GAT 1 has been replaced by the term 'intonation phrase' in GAT 2.

Pauses and events which cannot be attributed to any one particular speaker are also notated on a separate line.

*Overlaps and simultaneous speech*

[        ]        opening square brackets are inserted at exactly the point in speaking where the overlap starts, and closing square brackets, where it ends. The respective brackets are aligned with each other.

**Example (1')**: ((fictitious conversation, 1:04-2:00 sec))

```

01  A:  this is where the transcript starts
02  B:  yes exactly
03      (--)
→ 04  A:  if you interrupt [me ]
→ 05  B:                               [i'm] not interrupting
      you
      ((...))
08  A:  you just did
09      you just (.) interrupted me

```

Excerpt (1') begins with short contributions by A and B that consist of only a single intonation phrase. Between the contribution by B in segment 2 and that by A in segment 4, a pause ensues that cannot be attributed to either one of the speakers. It is neither a part of A's nor of B's turn and is hence notated as a separate segment (segment 3). In segment 5, B begins a new turn in overlap with A, who has not yet come to the end of her turn. In segments 8 and 9, A produces a turn which consists of two intonation phrases, the second one showing an internal micro pause.

After each overlap, further contributions by the speaker overlapped/interrupted are notated as a new segment. The only exceptions to this are continuers and short vocalizations which do not lay a claim to the floor. In this case a reader, whose eye is moving from left to right and from top to bottom, must briefly return from a later to an earlier segment in the transcript:

```

01  A:  i just [wan]ted [to] say
02  B:           [hm ]   [hm]

```

When several segments with overlaps follow each other, the rare problem of identifying respective pairs of overlapping passages may occur. In this case, the segments belonging to each other can be marked by narrower line spacing (1 instead of 1.5):

```

01  A:  [i just wanted to say] something else
02  B:  [i also want to say something]

03  A:  [who goes first ]
04  B:  [i always have to] fight for the floor

```



06 A: yes you are  
 07 (.)  
 08 A: you just did  
 09 you just (.) interrupted me  
 10 B: oh  
 11 did i really  
 12 A: (.) yes  
 13 B: okay  
 14 i'm sorry  
 15 A: never mind

A pause within an intonation phrase can be found in segment 9. Here the intonation phrase begun before the pause is continued after the pause.

Segment 12 shows a pause at the beginning of a turn which can be attributed to an individual speaker. Similarly, pauses produced within the extended turn of a single speaker, such as during projected stories, are placed within the relevant transcript segment (and not on a separate line): pauses separating intonation phrases are placed at the end of the preceding segment, hesitation pauses at the beginning of the following one.

The pauses in segments 3 and 7 cannot be unambiguously attributed to any one speaker. Therefore they are notated as separate segments. When such a non-attributable pause occurs, the next segment must be attributed to one of the speakers again, even if it is the same speaker who spoke before the pause, as in segment 8.

### 3.2.3. Other segmental transcription conventions

#### *Hesitation markers*

uh, er, uhm, erm etc. hesitation markers, so-called "filled pauses"<sup>20</sup>

#### *Laughter and crying*

hahaha }  
 hehe } short "syllabic" laughter, according to the vowel quality  
 hihi } and number of pulses or syllables,<sup>21</sup> respectively

((laughs)), ((cries)) characterization of a non-linguistic event

<<laughing> so> }  
 <<crying> so> } speaking interspersed with laughter or crying is notated  
 using a descriptor within a set of inner angled brackets  
 (see 3.2.4.). A set of outer angled brackets indicates the  
 extension of the phenomenon.<sup>22</sup>

<sup>20</sup> The representation of these must be adapted to the specificities of each individual language. Care should be taken to avoid homography with lexical items.

<sup>21</sup> There is still little known about the status of prominent and less prominent syllables in laughter. If relevant, more prominent laughter particles can be notated like accents, i.e. with capital letters.

<sup>22</sup> Inserting laugh particles into words such as, e.g., *so(h)o* should be avoided, as this makes automatic searches more difficult.

<< :-) > so>                      smile voice

*Continuers*

hm, mm	}	monosyllabic tokens
yes, yeah, yah, yep, yeh		
no, nope, naw		

m_hm, mm_hm	}	bi-syllabic tokens
uh_huh		
nhn		

ʔhmʔhm                                      reduplicated token, usually for negating  
 (ʔ = U+0294 (on the use of ʔ see section 3.3.2.))

*Unaccented turn beginnings and tags*

Unaccented turn beginnings<sup>23</sup> such as *yeah, oh, well, so* and unaccented tags such as *eh, huh* are usually not independent intonation phrases, but prosodically dependent elements. I.e., they attach to, or join, the following or preceding complete intonation phrase respectively (for details see 3.3.1.). When this happens, they are notated on the same line as the relevant intonation phrase and connected to it by an equal sign '=' (U+003D).

X:        so=that'll work=huh

**3.2.4. Further conventions for the minimal transcript**

*Non-verbal vocal actions and events*

Only those non-verbal vocal actions and events are notated which are relevant for the interaction.

((sniffs))	}	characterization of non-verbal vocal actions and events
((coughs))		

((coughs, 10 sec))                      action or event with an indication of its duration

Non-verbal vocal actions and events can occur within a turn or in the place of one or more turns. E.g.:

```
01    A:    i didn't mean it ((sniffs)) like that
02    B:    ((sighs))
03            (.)
04    B:    but you said it like that
```

---

<sup>23</sup> 'Beginnings' is used here to include what more precise analysis might reveal to be instances of 'pre-beginnings', 'beginnings' or 'turn-prefacing' discourse markers (see also Schegloff 1996).

They can also accompany verbal actions:

01 A: i didn't <<crying> mean it like that>  
 02 B: ((sighs))  
 03 (. )  
 04 B: <<getting up> but you said> it like that

Descriptions and comments in angled brackets, such as <<*crying*> > and <<*getting up*> >, are generally to be used in the following way: the inner angled brackets separate the descriptor, or the comment, from the wording of the turn. The outer angled brackets indicate its scope. The descriptor/comment is placed at that point in the transcript where the relevant phenomenon becomes observable. The outer angled bracket is placed at that point in the transcript where the relevant phenomenon ceases to be observable.

Extra-linguistic events which cannot be attributed to any one speaker are notated in a line without a speaker ID, for example ((*phone rings*)), ((*recording device beeps*)). If necessary, simultaneous events can be notated in a fashion similar to overlaps. Events which are attributed to several participants are notated in the same way: ((*Pat and John shake hands*)). In contrast to this, an action which is attributed to one specific participant is notated as a segment with the speaker ID for that speaker.

When non-verbal vocal events and verbal events occur simultaneously, the duration of the non-verbal vocal action or event can be split up in order to indicate precisely when each event begins/ends, e.g.:

04 B: <<getting up> but you said it> like that  
 05 A: ((sighs 2 sec)) [((sighs 1 sec)) ]  
 06 B: [do you think i] can make up  
 for it

### *Intelligibility*

( )	unintelligible passage
(xxx xxx xxx)	unintelligible passage with an indication of duration in syllables, with each 'xxx' representing one syllable
(may i)	assumed wording, uncertain
(may i say/let us say)	uncertain sounds, or syllables, at word level, with possible alternatives
((unintelligible, appr. 3 sec))	unintelligible passage with an indication of its duration (notated like a non-verbal vocal action or event)

Theoretically, uncertainties/alternatives can also apply to several words in a sequence. These are then notated in the same way: (*take Grey to London/take greater London*).

### 3.3. The basic transcript

At the level of the basic transcript, the minimal transcript is expanded to include prosodic information which is necessary to avoid misinterpretation of the segments in their interactional context (in terms of semantic structure and pragmatic function). This section will introduce the notions of intonation phrase, focus accent and pitch movement at the segment ending. In addition, notation conventions for these phenomena are presented. Moreover, conventions for other phenomena such as latching, lengthening, glottal closure and interpretive comments are introduced.

#### 3.3.1. Prosody<sup>24</sup>

##### *Intonation phrases*

Depending on their complexity, speaker contributions can consist of one or more intonation phrases.<sup>25</sup> Although the segmentation of a speaker contribution into intonation phrases is not entirely independent of syntax, the relation between intonational and syntactic units is not one-to-one. This is because participants can employ the segmentation of their turns for stylistic, interactional and semantic purposes (focus-background-distinction, see Uhmann 1991, Gussenhoven 2004). For the division of the transcript into segments only prosody, i.e. segmentation into intonation phrases, is relevant.

##### *Characteristics and boundaries of intonation phrases*

The intonation phrase is perceived as a single unit because of its cohesive pitch contour ("a stretch of speech uttered under a single coherent intonation contour", Du Bois et al. 1992:17). Of special importance is the phrase-final configuration of pitch. The final pitch movement of a segment is relevant both for the perception of the intonation phrase boundary as well as for the interpretation of the interactive function of the intonation phrase in question. An intonation phrase features at least one accented syllable (the nucleus), i.e. one syllable that is phonetically prominent due to pitch movement and/or loudness and/or length and that crucially determines the meaning of the utterance. This semantically and pragmatically most relevant pitch accent is referred to as *focus accent* in the following.

Intonation phrases are separated from each other by a number of optional boundary signals. These can occur in combination with one another or on their own. The individual signals themselves can be phonetically more or less salient. In the literature (see Couper-Kuhlen 1986, Du Bois et al 1992, Cruttenden<sup>2</sup>1997, etc., also Selting 1995), the following signals are mentioned:

---

<sup>24</sup> "Prosody is a superordinate term encompassing those suprasegmental aspects of speech which result from the interaction of the acoustic parameters fundamental frequency (F<sub>0</sub>), intensity and duration at the level of the syllable or in larger domains. It comprises auditory phenomena such as intonation, i.e. the pitch configuration of speech over time, volume, length, pause, as well as the related, more complex phenomena of speech rate/tempo and rhythm" (Selting 1995:1, transl. DB-W/EC-K). For a survey of the role of prosody in organizing talk-in-interaction see also Couper-Kuhlen/Selting (1996), Ford/Couper-Kuhlen (2004), Couper-Kuhlen (2009), Selting (2010).

<sup>25</sup> On the notion of 'intonation phrase' see Fox (2000) and Cruttenden (1997).

- pitch reset at the beginning of the phrase
- faster speech rate on the syllables preceding the first accented syllable (anacrusis)
- final lengthening at the end of the phrase
- creaky voice (glottalization) at the end of the phrase
- pitch movement on the (un)stressed syllables at the end of the phrase
- pauses

The relevant concepts will be explained in more detail in the following.

### *Focus accent*

Usually intonation phrases contain both accented as well as unaccented syllables, but they have at least one accented syllable. Accented syllables are most often made prominent by using pitch movement (and possibly volume and duration). The perception of accent is relational, i.e. a syllable is accented or unaccented in comparison to surrounding syllables.

The accents in an intonation phrase are usually placed on syllables that are lexically stressed. The only exception to this is contrastive accenting, as with: *no, it isn't baBY but BABy* or *i said fiFTY not fiFTEEN*. However, not every lexically stressed syllable in an intonation phrase is accented.<sup>26</sup> Stress is one kind of lexical information, i.e. it is part of the entry for a word in the lexicon (*cor'rect*, *'problem*, *'africa*, *ti'betan*). In some instances it even differentiates lexemes (*'detail - de'tail*; *'import - im'port* etc). In some words stress may be variable (*'hello - hel'lo*).

Yet even though an intonation phrase can have several accents, usually only one of them is especially relevant for the meaning of the utterance in question. This accent is referred to as *focus accent*. The focus accent is the semantically and pragmatically most relevant of the actual phonetic prominences in the intonation phrase. It indicates, and foregrounds, the focus of the utterance (see below). In the basic transcript, focus accent syllables are indicated by capital letters.<sup>27</sup> Examples for a correct notation of the accented syllable are:

1. with long vowels/diphthongs: *SEE*, *CRYing*, *disSOLVE*, *enLIGHTening*
2. with short vowels: *SIT*, *FOOTball*, *PROBlem*,<sup>28</sup> *COULDn't*,
3. with simple ambisyllabic consonants, i.e. after a short vowel: *BATtle*, *ROTten*, *PARrot*, *FUNny*, *SPANish*, *imPECcable*

Clitic extensions can also belong to the syllable:

*i CAN'T, he WON'T*

<sup>26</sup> This distinction between 'stress' (abstract potential for prominence) and 'accent' (phonetically realized prominence) goes back to Bolinger (1964/1972); see also Couper-Kuhlen (1986), Uhmann (1991:21-22).

<sup>27</sup> This convention contrasts with that in CA and other transcription systems. However, it is preferred here, as capital letters (a) do not need any specific formatting command and therefore take little time to write and (b) are generally convertible, i.e. they can also be used in html.

<sup>28</sup> Since short (or lax) vowels in English do not occur in open syllables, the first syllable here should be represented as *prob-*, not *pro-*.

When a pitch peak is shifted (phonetically) to the syllable before or after the lexically stressed syllable, it is still the lexically stressed syllable that forms the (phonologically and semantically relevant) accent syllable: we perceive these shifts not as shifted accents but only as variant pitch contours.

An outstandingly prominent accent – signalled by particularly extensive pitch movement or loudness – is additionally marked by exclamation marks before and after the accented syllable, as in segment 12 of the example ‚Sweet guy from Laguna‘ (see section 7), for instance:

08 JEFF: i an' i? (0.2) jus' talked to this (0.3) Asian guy,  
h°  
09 u:m,  
10 °hh who:'s twenny: h° six years OLD;  
11 and  
→12 (0.9)((click)) he's a !VE:R!y: sweet guy!=  
13 =he jus' moved to laGUna.

The position of the focus accent in the intonation phrase has repercussions for the semantics of the utterance. This becomes intuitively clear from the following set of examples:

- a. i see this text on my SCREEN
- b. i see this TEXT on my screen
- c. i see THIS text on my screen
- d. I see this text on my screen

All of these accentuations produce a different focus, i.e. they suggest different meanings and interpretations which we can grasp intuitively – even without knowing the rules for accent placement. In the above set, example (a) has a wide focus, i.e. the entire sentence is focussed. Examples (b) to (d), in contrast, all have a narrower focus, which in each case suggests that there is an alternative to the focussed element. Thus, in (b), "(this) TEXT" suggests an alternative such as "not (that) PICTURE", for instance. In (d), focussed "I" suggests "and not YOU" or "not HE" as possible alternatives. This explains why it is essential to notate accentuation in the transcript: otherwise misunderstandings can arise over the meaning expressed by the utterance transcribed.<sup>29</sup>

<sup>29</sup> In case of doubt, the focus accent syllable can be identified with the help of the following test, which excludes possible alternative accentuations: produce different possible accentuation variants aloud, compare them with each other as well as with the recording of the utterance transcribed and decide which one fits the auditory impression in the relevant sequential context best. Doing this with segment 19 of the story 'Sweet guy from Laguna' (see section 7) will show that the appropriate focus position is:

you jus' talk about `EVrything

rather than any of the possible alternatives, such as e.g.:

YOU jus' talk about evrything  
you JUS' talk about evrything  
you jus' TALK about evrything

The latter transcriptions suggest foci for the utterance that do not correspond to the auditory impression or to the context of the narrative and are therefore less plausible.

Other accents within an intonation phrase (which are by definition not focus accents) do not need to be indicated in the basic transcript. Should they need to be marked, only the vocalic element of the syllable is written as a capital letter (i.e.: *this tExt*). Intonation phrases with several focus accents are rare, but they can occur. An example of this would be the answer to the question: *who invented what?*:

the chiNESE invented ACupuncture  
but the euROPEans invented SURgery

Here there are two focus accents in each intonation phrase.<sup>30</sup>

Monosyllabic continuers are notated without an accent, i.e. in small letters. If they are louder than surrounding talk, this is indicated via the conventions described in section 3.4., such as, for instance, <<*f*> *hm*,>. With bi-syllabic continuers, the accented syllable is notated in capital letters, such as, e.g., *hm\_HM*.

### *Pitch movements at the end of intonation phrases*

In the basic transcript, every complete intonation phrase is provided with a symbol indicating its last pitch movement. Only incomplete intonation phrases are left without any such symbol.

Following CA notation conventions, the pitch movement of the intonation phrase is notated with one of the following symbols:

?	high rising
,	rising
–	level <sup>31</sup>
;	falling
.	low falling

The symbol for final pitch movement is placed directly after the last word in the intonation phrase; there is no blank in between.

The last pitch movement can extend across a variable number of syllables. In most cases, the pitch movement begins on the focus accent syllable and continues to the end of the phrase. In these cases, the pitch steadily falls, rises or stays level from the focus accent syllable onwards. This is the case with the falling pitch movement in the following example from the narrative 'Sweet guy from Laguna' (see section 7):

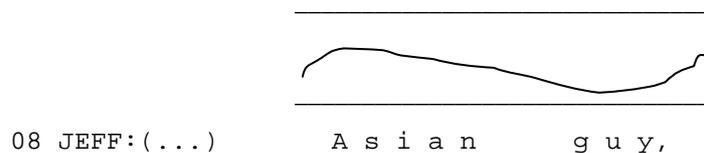
01 JILL:                    ( ... )   Y O U   b e e n   u p   t o : .



<sup>30</sup> Cf. Couper-Kuhlen (1986:47-48), Ladd (2008:7-8), also Uhmman (1991:221ff.).

<sup>31</sup> This symbol '–' is not a minus sign or a hyphen, but an en-dash (U+2013). Marking level pitch is necessary because the final pitch movement symbols at the same time indicate the end of the intonation phrase. Therefore, no complete intonation phrase should be left without such a symbol, even if the pitch does not change, as with level tone. In such cases, the end of the intonation phrase is usually marked by other parameters such as intensity and duration.

In the case of a falling-rising contour, i.e. when the pitch falls from the focus accent syllable onwards and then rises again, only the final pitch movement is notated:



In this example the pitch first falls from a peak on the syllable *A-* over the following syllable *-sian* before reversing direction on the word *guy* and rising to mid over the last part of the segment. In other words, even though overall the entire last pitch movement from the focus accent syllable is falling and then rising, only rising pitch is notated, as the basic transcript only records the final pitch movement at the segment ending.<sup>32</sup>

It is important to note that the symbols used here are not punctuation marks. Hence, they do not indicate the function of clauses or utterances, such as, e.g., statement or question: they only refer to the perceived final pitch movement of the intonation phrase.

Cut-off phrases which do not end with a glottal closure, but simply with trail-off, are not given any final intonation symbol:

x:      i thought i can

Here it is the notation without a symbol for final intonation that indicates that the intonation phrase was left incomplete.<sup>33</sup>

The symbol '=' represents final intonation remaining level in the middle of the speaker's range.<sup>34</sup>

Note that turn-beginning elements and tags are sometimes realized as separate, prosodically independent intonation phrases, with their own final pitch movement. In the case of tags with a separate pitch movement, both the pitch movement at the end of the preceding intonation phrase as well as that on, or at, the end of the tag are notated. Tag and intonation phrase are usually connected by latching, which is notated with '=' (see 3.3.2.):

01      A:      THIS is where the transcript starts;=isn't it?

At the same time, realizing the tag as an independent intonation phrase is possible:

01      A:      THIS is where the transcript starts.  
02              ISn't it?

<sup>32</sup> To notate the pitch movement more precisely, notation conventions from fine transcription can be added (see 3.4.). This makes it possible to represent the falling and rising movement in its distribution over the accented syllables:

    `Asian ˈgŭy,

<sup>33</sup> If necessary for specific research purposes, the special symbol '\ (U+005C) can be inserted when intonation phrases are broken off without glottal closure.

<sup>34</sup> For a more detailed notation of level intonation at the end of an intonation phrase, see the conventions for fine transcription.

Or both elements can be fully integrated into one intonation phrase:

01 A: THIS is where the transcript starts isn't it?

Beginning elements and turn-prefacing discourse markers such as *so*, *yes*, *no*, *well*, *okay*, can also be notated (a) as unaccented items belonging to the following intonation phrase, (b) as prosodically dependent intonation phrases or (c) as prosodically independent intonation phrases, according to their realization. Compare the turn-prefacing *so* and the tag *isn't it* in the following three versions with increasingly stronger boundaries between *so* and the following syntagm, as well as between *isn't it* and the preceding syntagm:<sup>35</sup>

a. 01 A: so THIS is where the transcript starts isn't it?

b. 01 A: so;=THIS is where the transcript starts;=isn't it?

c. 01 A: SO;  
02 THIS is where the transcript starts;  
03 ISn't it?

### 3.3.2. Further conventions for the basic transcript

#### *Latching of intonation phrases*

= latching, i.e. immediately moving into a new intonation phrase (by the same or by another speaker) without the usual micropause

The latching symbol is placed at the end of the preceding and at the beginning of the next intonation phrase. For instance:

01 A: me TOO;=  
02 B: =me TOO;

Within intonation phrases, words are usually produced directly following one another. Hence, latching is the rule here and is consequently not notated.

---

<sup>35</sup> In general, transcribers can train their perception by repeating the utterance aloud to themselves, by self-observation and, in particular, by comparing their own production with the utterance as recorded. Also, some simple tests can help in identifying the pitch movement for transcription: In order to clarify whether the pitch at the end of the intonation phrase in segment 1 of our example "Sweet guy from Laguna" is falling, as notated,

01 JILL: what have YOU been up to:.

producing the utterance aloud with rising instead of falling pitch and comparing this with the original in the recording will usually settle the matter quickly. Advanced students with knowledge of acoustic phonetics may also want to use an acoustic-phonetic program for speech analysis such as PRAAT (URL: <http://www.praat.org>).

### *Lengthening*

: :: ::: according to the duration of the lengthening, e.g.: *so:* or *so::*, *a::nd*, etc.

The colons are used as follows:

: lengthening of about 0.2-0.5 seconds  
 :: lengthening of about 0.5-0.8 seconds  
 ::: lengthening of about 0.8-1.0 seconds

The perception of duration in lengthened sounds or syllables depends, in turn, on accentuation, tempo and rhythm.<sup>36</sup>

### *Glottal closure*

? cut-off with glottal closure or syllable onset  
 with glottal closure (U+0294)<sup>37</sup>

The use of a glottal closure in accordance with the pronunciation rules of a language (see section 3.2.1.) is not notated. (See continuers, section 3.2.3.)

### *Interpretive comments*

<<angry> > } interpretive comments with an indication of their scope  
 <<surprised> > }

Auditory phenomena that the transcriber cannot represent in terms of form but that are presumed to be relevant for the interpretation, can be notated initially with interpretive comments. In these cases, comments such as '<<surprised> >', '<<upset> >', '<<hesitant> >', etc. are placed in front of the relevant utterance (or part of utterance). The outer angled bracket is placed where the scope of the feature described by the comment ends. The scope of a comment can vary. It can range from phonetic segments to syllables, words, phrases, intonation phrase or phrases and finally to entire turns. If a comment is valid for more than one turn, it is repeated at the beginning of the new turn:

```
01 A: are you <<hesitant> possibly a policeman
02 do i know you from the police station>
03 B: sorry
04 A: <<hesitant> are you from the police>
```

Interpretive comments can be complemented or replaced by more form-oriented descriptors at a later stage.

<sup>36</sup> Shortening is not notated by a separate symbol. If needed, a shortened realization of /ru:f/ 'roof', for instance, as /rʊf/ can be indicated with *ruf*.

<sup>37</sup> Note that the phonetic symbol for glottal closure is raised as a superscript. If this symbol is not available or not reliable, the symbol \* (U+002A) can be used instead. If a symbol for break-off without glottal closure is needed for special purposes, \ (U+005C) can be used.



pitch peaks or valleys which are very high or very low in comparison to the preceding accents. Such pitch jumps are local phenomena, in contrast to the shifts in pitch register discussed below. Their notation symbol is placed in front of the syllable(s) it applies to.

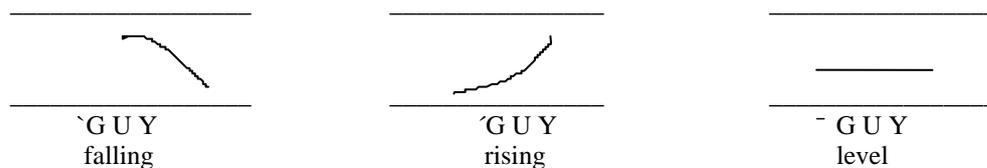
#### *Change in pitch register*

<<l>            >     lower pitch register  
<<h>            >     higher pitch register

These symbols are used when the speaker shifts to a pitch register that is different from the previous one or from the speaker's usual register – for one or more intonation phrases, or for parts of an intonation phrase, such as in marking parenthetical matter. The scope of a register change can vary, but it usually applies to larger chunks in a turn (on the use of angled brackets see section 3.2.4. et passim).

#### *Pitch accents*

The usual notation of pitch accents indicates the pitch movement on and after the accented syllable. It is usually assumed to take one of the following shapes:



When the pitch peaks and valleys are produced on the core of the accented syllable, e.g. within the phonetic segment /αI/ ("uy") in this example, the accent notation indicates the pitch movement that starts here and is continued throughout the remainder of the syllable, often spreading over the word and any following unaccented syllables. Hence, the pitch accents on the word "guy" shown above are falling, rising and level respectively. The relevant symbol is placed in front of the accented syllable:<sup>39</sup>

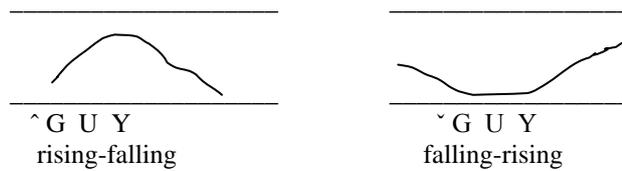
`SO                    falling (U+0060)  
^SO                    rising (U+00B4)  
-SO<sup>40</sup>                level (U+00AF)

A parameter change in comparison to the preceding syllable such as, e.g., a step up or down to the accented syllable or a glide up or down to the peak or valley of the pitch accent, is indicated by arrows (see below).

Apart from that, there are cases in which noticeable falling-rising or rising-falling glides are produced within a single accented syllable, which is then also often longer. Such cases are schematized in the following pictures:

<sup>39</sup> The symbols to be used here are the acute and grave accent marks on the keyboard; they are not apostrophes or single quotation marks.

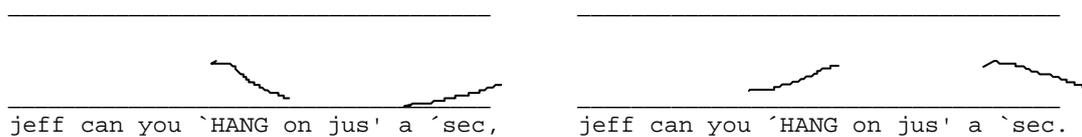
<sup>40</sup> Alternatively: –SO (with U+2013)



Such rising-falling and falling-rising pitch accents are represented as follows:

$\wedge$ SO<sup>41</sup> rising-falling (U+02C6)  
 $\checkmark$ SO<sup>42</sup> falling-rising (U+02C7)

In yet other cases the first part of a rising-falling or falling-rising pitch accent is realized on the accented syllable, but the second part only later on an unaccented syllable:



In these cases, the pitch movements are notated where they occur, i.e. the first part of the pitch movement on the accented syllable and the second part on the relevant unaccented syllable.

Combinations of the notation of accent type and pitch accent movement then look as follows, e.g.:

Focus accents:	Secondary accents:
$\checkmark$ FALling	$\checkmark$ fAlling
$\wedge$ RIsing	$\wedge$ rIsing
$\bar{\text{L}}$ EVel	$\bar{\text{L}}$ Evel

Upsteps and downsteps at the onset of an accented syllable as well as differences between prior unaccented syllables and the peak or valley of an accented syllable can be represented as follows:

$\uparrow$	small pitch upstep to the peak of the accented syllable (U+2191 and U+0060)
$\downarrow$	small pitch downstep to the valley of the accented syllable (U+2193 and U+00B4)
$\uparrow\uparrow$ SO or $\downarrow\downarrow$ SO	noticeable high or low pitch jump to the peak or valley of the accented syllable
$\uparrow\bar{\text{L}}$ SO or $\downarrow\bar{\text{L}}$ SO	pitch jumps to noticeably higher or lower accents with level pitch <sup>43</sup>

<sup>41</sup> Alternatively:  $\wedge$ SO (with symbols for rising and falling accents combined, i.e. U+00B4 and U+0060)

<sup>42</sup> Alternatively:  $\checkmark$ SO (with symbols for falling and rising accents combined, i.e. U+0060 and U+00B4)

In combination with the notation of the pitch movement at the end of the intonation phrase (see 3.3.1.), the conventions suggested above capture how high or low the last pitch movement of the intonation phrase goes: a final falling pitch movement can fall to mid (notated with ';') or low ('.'), a rising final pitch movement can rise to mid ('.') or high ('?'). See:

27            'at's\_a 'BASketball player,  
73            jEff can you ↑`HANG on jus' a 'sEc?

In the first instance the pitch rises to mid on the final unaccented syllables, in the second it rises to high on the final secondary accent.

### *Loudness and tempo changes*

Notating loudness and tempo only becomes relevant when there is a shift in comparison to the loudness and tempo chosen for, or established by, previous bits of talk. If a speaker habitually speaks loud or fast, this is indicated in the transcript header. In this case the speaker's habit is the reference point for transcription. Changes in loudness and tempo are notated with as much differentiation as possible by means of the following parameters:

<<f>	>	forte, loud <sup>44</sup>
<<ff>	>	fortissimo, very loud
<<p>	>	piano, soft
<<pp>	>	pianissimo, very soft
<<all>	>	allegro, fast
<<len>	>	lento, slow
<<cresc>	>	crescendo, becoming louder
<<dim>	>	diminuendo, becoming softer
<<acc>	>	accelerando, becoming faster
<<rall>	>	rallentando, becoming slower

The indication of a change in loudness or tempo refers to a stretch of the turn enclosed in the brackets: the descriptor is positioned (in angled brackets itself) just before the syllable on which the parameter change can be observed; the outer bracket is placed where the scope of the parameter ends.<sup>45</sup>

<sup>43</sup> If necessary, the symbol for level pitch at the end of an intonation phrase ( – = U+2013, see basic transcript, 3.3.) can be lowered with a subscript in order to represent level pitch in a lower register ( \_ = U+005F) or raised with a superscript ( ^ = U+00AF) for level pitch in a higher register.

<sup>44</sup> GAT cannot use the CA convention for notating loudness with capital letters, as the latter are used for indicating accents. The prosodic phenomenon of accenting is in fact due to more than just loudness: in order to identify accents, it is necessary to take duration and in particular pitch as well as loudness into account.

<sup>45</sup> If parameters with differing scope are employed, the angled brackets can be used with indices, such as <<f, h>            f>            h>.

*Changes in voice quality and manner of articulation*

- <<creaky> > glottalized, "vocal fry"  
 <<whispery> > examples of change of voice quality as stated, e.g. to  
 whispery phonation

Such parameters can also be used for stretches of the turn of variable length. Like the loudness and tempo parameters described above, they are surrounded by angled brackets. Other common descriptors for voice quality shifts include, for instance, breathy, harsh, lax, tense, falsetto, rounded lips, spread lips, nasalized, denasalized, etc. (see Laver 1980).

*Other phonetic phenomena*

- |                    |   |  |
|--------------------|---|--|
| ((click))          | } | phonological, para- and non-linguistic actions and events<br>(descriptors are used in as detailed a way as possible) |
| ((smack))          |   |  |
| ((labial click))   |   |  |
| ((alveolar click)) |   |  |

*Boundary signals*

In order to notate the boundary signals of intonation phrases as described above, the following symbols can be used:

- |             |   |
|-------------|---|
| =           | latching of intonation phrases  |
| ↑ ↓         | pitch reset at the beginning of the intonation phrase                 |
| <<all> >    | faster speech rate on the first syllables of the phrase (anacrusis)   |
| :           | final lengthening at the end of the phrase                            |
| <<creaky> > | creaky voice at the end of the phrase                                 |
| . ; - , ?   | final pitch movements on the last accented (and unaccented) syllables |
| (.) etc.    | micro or other pause after or between phrases                         |
| °h h° etc.  | in- or outbreath after or between phrases                             |

## 4. Suggestions for the notation of complex phenomena

In this section, we present suggestions for extending the fine transcription, in particular in terms of intonation and rhythm, for special research questions.

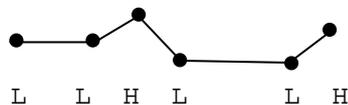
### 4.1. Autosegmental representation of intonation<sup>46</sup>

As an alternative to the auditory/impressionistic record of pitch as presented in section 3, GAT 2 allows for the integration of an autosegmental representation of intonation contours into basic and fine transcripts for specific research purposes.<sup>47</sup> Such representations typically require an a priori analysis of intonation and the establishment of an inventory of tones and tonal sequences for each individual language variety investigated. Specific tones and tonal sequences from this inventory are then assigned to individual transcript segments. As a rule, the autosegmental representation takes a phonological approach, i.e. the pitch contours observed are understood to be instantiations of abstract patterns in the tonal inventory of a given language. But autosegmental notation can also be used for varieties whose intonational inventory is not yet known. Then the symbols used represent first and foremost a strongly abstracting phonetic notation; they are only potentially phonological. For an exclusively phonetically oriented notation, see Grabe et al. (1998) and also Gilles (2005).

In the following, the integration of autosegmental representation into GAT is illustrated by means of ToBI notation conventions for the intonation of standard varieties of British, American and Australian English (Beckman/Ayers 1994, also Beckman/Ayers Elam 1997). It is beyond the present paper to introduce ToBI in any detail, let alone reflect current trends in the discussion of it (cf. Ladd 2008). Instead, the aim is to point to the integratability of (some of) its relevant concepts.

A 'proper' ToBI notation includes several tiers (cf. Grice 2006, Ladd 2008). We will focus here on the tonal layer for illustration. Pitch contours are understood to be sequences of local tonal events (targets) and transitions between these. The pitch targets are specified in terms of high tones (H) and low tones (L). The transitions usually evolve from the linear interpolation of the pitch targets.

An intonation contour is a sequence of any such H and L tones. Schema (1) shows such a tonal sequence; the black dots represent the phonetic targets of the individual tones.

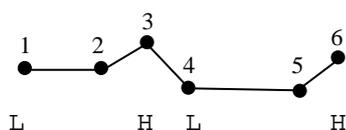


Schema 1

Neighboring targets which are both low or high respectively can be associated with a single low, or high, tone (*tone spreading*). Thus, in (2) the first L tone specifies the first and second target and the second L the fourth and fifth target.

<sup>46</sup> We are grateful to Nicole Dehé and Anne Wichmann for editorial help with this section.

<sup>47</sup> For an introduction to the autosegmental representation of intonation see Ladd (2008), Gussenhoven (2004) and Beckman/Ayers Elam (1997).



Schema 2

There are four types of tone: accent tones, leading/trailing tones, phrasal tones and boundary tones. Accent tones ( $H^*$ ,  $L^*$ ) constitute the core of pitch accents. They are associated with prominent syllables, the association being signalled by a star. Monotonal pitch accents consist of one accent tone only. Leading and trailing tones ( $H$ ,  $L$ ) occur only in combination with accent tones. Together with an accent tone, they form bitonal pitch accents. Depending on whether they precede or follow the accent tone they belong to, they are referred to as leading or trailing respectively. Phrasal and boundary tones are edge tones. Phrasal tones ( $H^-$  or  $H^-$ ,  $L^-$  or  $L^-$ ) are associated with the edge of a prosodic constituent called intermediate phrase, boundary tones are associated with the edge of intonation phrases. Boundary tones indicating a markedly high beginning of an intonation phrase are notated as  $\%H$ , those at the end of an intonation phrase are notated as either  $L\%$  or  $H\%$ .

Depending on the specific version of ToBI,<sup>48</sup> there are five types of pitch accent, which can appear as nuclear or pre-nuclear accents: a high or 'peak' accent ( $H^*$ ), a downstepped accent ( $!H^*$ ), a low accent ( $L^*$ ), a 'scoped' accent ( $L^*+H$ ) and a rising 'peak' accent ( $L+H^*$ ). In ToBI, the nuclear accent is by definition the last pitch accent of an intonation phrase. Any and all preceding accents are called pre-nuclear accents. Specified combinations of one of the five pitch accents  $H^*$ ,  $!H^*$ ,  $L^*$ ,  $L^*+H$  and  $L+H^*$  in nuclear position together with one of the phrasal tones  $L^-$ ,  $H^-$  and one of the final boundary tones  $L\%$ ,  $H\%$  result in tone sequences which roughly correspond to the falling, rising, level, falling-rising and rising-falling accents described for GAT in section 3.3.1 and 3.4 above.<sup>49</sup>

Here is a table of correspondences for the inventory of nuclear accents (tonetic stress marks that are final) in (Southern) British English (Roach 1994:96, see also Ladd 2008:91):

Tone name	Pitch accent	Phrase accent	Boundary tone
Low level	$L^*$	$L^-$	$L\%$
High level	$H^*$	$H^-$	$L\%$
(High) rise-fall	$L^* + H$	$L^-$	$L\%$
High fall-rise	$H^*$	$!H^-$	$H\%$
High fall	$H^*$	$L^-$	$L\%$
Low fall	$!H^*$	$L^-$	$L\%$
High rise	$H^*$	$H^-$	$H\%$
Low rise	$L^*$	$L^-$	$H\%$
Low fall-rise	$!H^*$	$L^-$	$H\%$

<sup>48</sup> Here we follow MAE\_ToBI (Mainstream American English ToBI, cf. Beckman/Hirschberg/Shattuck-Hufnagel 2005, Ladd 2008:105).

<sup>49</sup> For more English examples the reader is referred to Cruttenden (<sup>2</sup>1997:61ff.). For problems inherent in this 'transferral', see the discussion in Ladd (2008:90-92).

In unmarked cases, nuclear accents correspond to focus accents in GAT and pre-nuclear accents to secondary accents in GAT (see section 3). Accordingly, nuclear accented syllables are notated with capital letters in the orthographic notation line. With pre-nuclear accented syllables, only the syllable core is capitalized; with pre-nuclear accented syllables containing diphthongs or several vowels, only the first vowel is capitalized. The ToBI intonation contour is notated in a separate line below the relevant orthographic GAT transcript line. Both these lines together form a transcript segment. Therefore, the line indicating the intonation is not assigned a separate segment number.

In font and font size the intonation line resembles the orthographic notation line. The starred tones are aligned notationally as far as possible with the vowels of the accented syllables. The boundary tones are placed such that the letter indicating the tone (not the percentage sign) is aligned with the beginning or the end of the relevant verbal notation line respectively. For an illustration see the following excerpt from "Sweet guy from Laguna":

```

08  JEFF:  I an' I? (0.2) jus' tAlked to this (0.3)
          H*    !H*                !H*

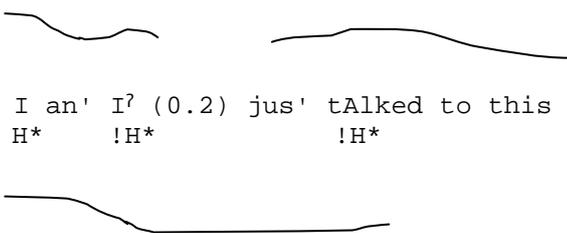
          `Asian <<creaky> ´gUy,> h°
          !H*L                L* H- H%

09          u:m,

10          °hh whO:'s twEnny: h° sIx years `OLD;
          L*                H*                L*                H* L- L%

```

For illustrative purposes, transcript segments consisting of one line of orthographic notation and one line of intonational notation can also be complemented by impressionistic pitch curves as indicated below. The latter are placed above the relevant line of verbal transcript. Such pitch curves are not part of the transcript segment.



```

08  JEFF:  I an' I? (0.2) jus' tAlked to this (0.3)
          H*    !H*                !H*

          `Asian <<creaky> ´gUy,> h°
          !H*L                L* H- H%

```

The pitch curve serves to facilitate understanding of the intonational notation, not to indicate additional phonetic detail. For the latter, F0 extractions should be used; these, however, are not integrated into the transcript.<sup>50</sup>

<sup>50</sup> It should be pointed out in conclusion that ToBI, like GAT, needs to be adapted to the particular language or language variety under investigation.

## 4.2. Notation of rhythm

Rhythm can emerge either when accented syllables are produced isochronously, i.e. follow each other at perceptually equal intervals of time, or when two immediately adjacent, equally long intonation phrases exhibit recurrent accent locations. The different principles resulting in rhythmic structure necessitate different notations.<sup>51</sup>

A basic rhythmic unit is the *foot* (cf. Nespor/Vogel 2007). It includes an accented syllable, perceived as a rhythmic beat, and any following unaccented syllables up to the next accented syllable (the latter is part of the next foot). For representing rhythmic feet produced isochronously, we recommend the notation developed by Couper-Kuhlen (1993, see also Auer et al. 1999). Using this system, the accented syllables in the intonation phrase /*thEy*: /*brOught it up*/On *them*/SELVES (segment 71 in the English GAT example "Sweet guy from Laguna"; see section 7), which are perceived as isochronous, would be notated as follows:

```
71  JEFF:  /thEy:           /
          /brOught it up  /
          /On them       /
          /↑`SELVES.
```

Each foot is notated on a separate line. The left- and right-hand slashes, which are aligned with each other, indicate the perceptually equal duration of each foot. The actual measured length<sup>52</sup> of each foot and the number of syllables per foot (here in parentheses) can also be indicated:

```
71  JEFF:  /thEy:           /      0.28 sec. (1)
          /brOught it up  /      0.34 sec. (3)
          /On them       /      0.34 sec. (2)
          /↑`SELVES.
```

If the rhythm is broken and a next foot follows faster or slower, i.e. if it starts earlier or later than expectable on the basis of the rhythm established so far, this is indicated by a re-positioning of the slashes. The following is an example of the notation of a late beat (on *they*), after which a new faster rhythm emerges:

<sup>51</sup> Given the current state of research on speech rhythm, only the rhythmic principle of accent isochrony will be presented here. A discussion of the rhythmic principle of recurrent accent distribution in Turkish German as well as a proposed representation based on metrical grids (see Selkirk 1984) can be found in Kern (2008). Further research on the occurrence and function of this principle in other language varieties and styles of speech remains to be done.

<sup>52</sup> The acoustic analysis programme PRAAT ([www.praat.org](http://www.praat.org)) can be recommended for actual measurements of the temporal duration of feet. What is measured is the interval between accented syllables from vowel onset to vowel onset (see Couper-Kuhlen 1993:54).

70 JEFF: tha'ʔ (0.3)/cAlls for the / 0.55 sec.  
 /˘^DEATH penalty; (0.1) 0.81 sec.  
 ((next beat late))  
 /thEy: / 0.28 sec.  
 /brOught it up / 0.34 sec.  
 /On them / 0.34 sec.  
 /˘^SELVES.

The following is an example of the notation of an early beat (on *moved*), after which the tempo accelerates:

12 JEFF: he's a /!`VE:Ry:!! sweet / 0.69 sec.  
 /gUy;==he jus' 0.43 sec.  
 ((next beat early))  
 /mOved to la / 0.50 sec.  
 /`GUna.

## 5. Interlinear translation of turns

If the language used for publication is different from the language transcribed, translation becomes necessary. To this end, a free translation of each line or turn into the target language is inserted just beneath the verbal transcript segment in the original language. This translation line should be set in Times New Roman. If the study takes typological aspects into account, it may be desirable to include an interlinear gloss between the verbal transcript segment in the original language and the free translation. The interlinear gloss usually indicates the morphological structure of the original (see Croft 2003, Leipzig Glossing Rules 2008).

An example from Finnish will illustrate this (Sorjonen 2001:413). The first line is the original transcript line, the second is the interlinear gloss and the third is the free translation:

5 E: Meinaa-k se nyt jää-hä kotti-i sitte.  
*intend-Q it now stay-INF home-ILL then*  
 Is she gonna stay at home then.  
 6 M: Joo::, Joo::,  
*PTCL PTCL*

In the interlinear gloss, each morpheme of the original utterance (L1, here Finnish) is represented by an appropriate morpheme in L2 (here English). The lexical morphemes of L1 are translated into lexical morphemes of L2 and displayed in italics in the line below (e.g. in segment 5 *meinaa* = 'intend'; *jää* = 'stay'). The grammatical morphemes of L1 are glossed by reference to a set of metalinguistically determined grammatical categories (see the table below) represented with capitalized abbreviations (e.g. *k* = Q (interrogative particle), *i* = ILL (illative 'into')) etc. Morpheme boundaries are indicated by the minus sign (-), word boundaries

by blanks ( ). Particles and other words which are hard to translate, such as, e.g., in segment 6 *joo*, can remain untranslated, in particular when they are the object of study.

The following is a list of common grammatical categories with their abbreviations (Croft 2003). This list may need to be supplemented depending on the language in question:

<i>Grammatical category</i>	<i>Abbreviation</i>
Definite	DEF
Indefinite	INDEF
Masculine	M
Feminine	F
Neuter	N
Nominative	NOM
Genitive	GEN
Dative	DAT
Accusative	ACC
Singular	SG
Plural	PL
Diminutive	DIM
Familiar (T) form	FAM
Direct object	DO
Indirect object	IO
Present	PRS
Past	PST
First person	1
Second person	2
Third person	3
Auxiliary verb	AUX
Particle	PTCL
Infinitive	INF

Participle	PART
Subjunctive	SUBJ
Passive	PASS
Negation	NEG
Complementizer	COMP
Quotative	QUOT

## 6. Summary of the most important GAT 2 transcription conventions

### Minimal transcript

#### Sequential structure

[ ] overlap and simultaneous talk  
 [ ]

#### In- and outbreaths

°h / h° in- / outbreaths of appr. 0.2-0.5 sec. duration  
 °hh / hh° in- / outbreaths of appr. 0.5-0.8 sec. duration  
 °hhh / hhh° in- / outbreaths of appr. 0.8-1.0 sec. duration

#### Pauses

(.) micro pause, estimated, up to 0.2 sec. duration appr.  
 (-) short estimated pause of appr. 0.2-0.5 sec. duration  
 (--) intermediary estimated pause of appr. 0.5-0.8 sec. duration  
 (---) longer estimated pause of appr. 0.8-1.0 sec. duration  
 (0.5)/(2.0) measured pause of appr. 0.5 / 2.0 sec. duration  
 (to tenth of a second)

#### Other segmental conventions

and\_uh cliticizations within units  
 uh, uhm, etc. hesitation markers, so-called "filled pauses"

#### Laughter and crying

haha } syllabic laughter  
 hehe }  
 hihi }  
 ((laughs)) } description of laughter and crying  
 ((cries)) }  
 <<laughing> > laughter particles accompanying speech with  
 indication of scope  
 <<:-)> so> smile voice

#### Continuers

hm, yes, no, yeah monosyllabic tokens  
 hm\_hm, ye\_es, bi-syllabic tokens  
 no\_o  
 ʔhmʔhm with glottal closure, often negating

### Other conventions

((coughs))	non-verbal vocal actions and events
<<coughing>>	...with indication of scope
( )	unintelligible passage
(xxx), (xxx xxx)	one or two unintelligible syllables
(may i)	assumed wording
(may i say/let us say)	possible alternatives
((unintelligible, appr. 3 sec))	unintelligible passage with indication of duration
((...))	omission in transcript
→	refers to a line of transcript relevant in the argument

### **Basic transcript**

#### Sequential structure

=	fast, immediate continuation with a new turn or segment (latching)
---	--

#### Other segmental conventions

:	lengthening, by about 0.2-0.5 sec.
::	lengthening, by about 0.5-0.8 sec.
:::	lengthening, by about 0.8-1.0 sec.
?	cut-off by glottal closure

#### Accentuation

SYLlable	focus accent
!SYL!lable	extra strong accent

#### Final pitch movements of intonation phrases

?	rising to high
,	rising to mid
–	level
;	falling to mid
.	falling to low

### Other conventions

<<surprised>>	interpretive comment with indication of scope
---------------	---

## **Fine Transcript**

### Accentuation

SYLlable	focus accent
sYllable	secondary accent
!SYL!lable	extra strong accent

### Pitch jumps

↑	smaller pitch upstep
↓	smaller pitch downstep
↑↑	larger pitch upstep
↓↓	larger pitch downstep

### Changes in pitch register

<<l>	>	lower pitch register
<<h>	>	higher pitch register

### Intralinear notation of accent pitch movements

`sO	falling
˘sO	rising
ˉsO	level
ˆsO	rising-falling
˘sO	falling-rising
↑	small pitch upstep to the peak of the accented syllable
↓	small pitch downstep to the valley of the accented syllable
↑ˉsO bzw. ↓ˉsO	pitch jumps to higher or lower level accented syllables
↑↑˘sO bzw. ↓↓˘sO	larger pitch upsteps or downsteps to the peak or valley of the accented syllable

### Loudness und tempo changes, with scope

<<f>	>	forte, loud
<<ff>	>	fortissimo, very loud
<<p>	>	piano, soft
<<pp>	>	pianissimo, very soft
<<all>	>	allegro, fast
<<len>	>	lento, slow
<<cresc>	>	crescendo, increasingly louder
<<dim>	>	diminuendo, increasingly softer
<<acc>	>	accelerando, increasingly faster
<<rall>	>	rallentando, increasingly slower

### Changes in voice quality and articulation, with scope

<<creaky>	>	glottalized
<<whispery>	>	change in voice quality as stated

## 7. Sample transcript of an excerpt from natural, everyday talk-in-interaction

In the following we present a conversational storytelling episode in basic transcript and in fine transcript form with intralinear prosodic notation and in modified standard orthography. It comes from a conversation in the Corpus of Spoken American English (sbc 028, Du Bois 2000). Jeff and Jill are a young couple who are in different cities at the time and who are communicating via long-distance telephone. The relevant audio file can be accessed through the GAT home page ([http://agd.ids-mannheim.de/html/gat\\_en.shtml](http://agd.ids-mannheim.de/html/gat_en.shtml)).

The storytelling episode "Sweet guy from Laguna" in basic transcript form (modified standard orthography):<sup>53</sup>

01 JILL: ((click)) what have YOU been up to:.  
 02 JEFF: (0.3) NOThing.  
 03 JILL: (0.3) NOThing?  
 04 JEFF: jus' HOMEwork.  
 05 JILL: (0.3) YE::AH?  
 06 JEFF: i jus' played BASKetball at main beach.  
 07 JILL: OH:::.  
 08 JEFF: i an' i? (0.2) jus' talked to this (0.3) Asian guy, h°  
 09 u:m,  
 10 °hh who:'s twenny: h° six years OLD;  
 11 and  
 12 (0.9) ((click)) he's a !VE:R!y: sweet guy;=  
 13 =he jus' moved to laGUna.  
 14 (.)°h[h]  
 15 JILL: [U]Nhu[nh.]  
 16 JEFF: [A:N]D-  
 17 (.) we jus' (0.7)  
 18 <<laughing> whh° we\_jus' got in one of those? (0.2)  
 conversations where; >  
 19 °hh you jus' talk abou' EVrything.=  
 20 =?ih heh heh heh huh,  
 21 JILL: un[HU:NH, ]  
 22 JEFF: [like like] r:il poLITical:\_an';  
 23 °hh a:nd h°

<sup>53</sup> Among others, the following modified spellings have been used: 'twenny' (=twenty), 'ril' (=real), 'evrything' (=everything), 'yiknow' (=you know), 'cuz' (=cause), 'cathlic' (=catholic), 'sd' (=said), 'unnerstand' (=understand), 'gonna' (=going to).

24 EVrything from norm:ʔ (.)  
 25 um: h°  
 26 norm NIXon.  
 27 huh 'at's\_a BASKetball player,  
 28 °hh hn heh huh;=  
 29 JILL: =ehh hnh  
 30 JEFF: i said NORM instead of like (.) PRESIDen' (0.3)  
 NIX[on],  
 31 JILL: [oh] RI::GHT;  
 32 JEFF: heh heh: huh  
 33 JEFF: from NIXon-  
 34 to like rush LIMbaugh-  
 35 to aBORTion-  
 36 to capital PUNishmen'- h°  
 37 JILL: ((click)) WO[::W.]  
 38 JEFF: [like] h° (.)  
 39 YEAH- =  
 40 =°h yiknow an' it's so FUNny;  
 41 cuz he'sʔ (0.2) he's a CATHlic,  
 42 (.) °h (.) ((click)) A:ND;  
 43 °hh YIknow like;  
 44 i n:ailed him on the contraDIction yiknow;=  
 45 =he's like pro capital PUNishmen'-  
 46 (0.2) but an:' pro LIFE? ʔehhh°  
 47 JILL: ʔunHUNH[:? ]  
 48 JEFF: [i sd] try to exPLAIN 'at to me.=  
 49 =i don' <<laughing> unnerSTA:ND<sup>54</sup>. heh;>  
 50 JILL: °hh WAIT;=  
 51 =he said he's: (.) pro: capital PUNishmen'-  
 52 JEFF: an' pro LI[FE,]  
 53 JILL: [an'] pro LIFE.  
 54 RI::GH[T. ]  
 55 JEFF: [that's] likeʔ (.)  
 56 to me that's like a hu:ge contraDIction?

---

<sup>54</sup> This representation is not intended discriminatorily.

57 °hh ((click)) a:nd [he] said  
58 JILL: [m:]  
59 JEFF: he\_d he: JUStified it as;  
60 °hh well SEE;=  
61 =the<sup>?</sup> (0.2) th<sup>?</sup> (0.2) those little EN<sup>?</sup> (.) the little<sup>?</sup> (0.5)  
u:m (.) EMBryos? (0.4)  
62 ((click)) they don' have a (0.2) deCIS<sup>?</sup> (.) they  
can't make a deCISion.  
63 (0.5)  
64 JEFF: ((click)) YIknow;=  
65 ='n' they're gonna DIE;=  
66 JILL: =this's what HE was saying.  
67 JEFF: YEAH;  
68 an' HE said that yiknow people who (.)  
°hh commit: CRI:MES;;  
69 °hh yiKNOW; (.)  
70 tha<sup>?</sup> (0.3) calls for the DEATH penalty;  
71 (0.1) they: brought it upon themSELVES.  
72 JILL: (0.4) oh WAIT; (.)  
73 jeff can you HANG on jus' a sec?  
74 JEFF: YE:S.

The same storytelling episode "Sweet guy from Laguna" in fine transcript form (modified standard orthography).<sup>55</sup>

01 JILL: ((click)) whAt have ↑`YOU been Up to:.  
 02 JEFF: (0.3) `NOthIng.  
 03 JILL: (0.3) ´NOthIng?  
 04 JEFF: ↑jus' `HOMework.  
 05 JILL: (0.3) <<h> ´YE::AH?>  
 06 JEFF: ↑I jus' played `BASketball at main bEach.  
 07 JILL: `OH:::.  
 08 JEFF: I an' I? (0.2) jus' tAlked to this (0.3) `Asian  
     <<creaky> ´gUy,> h°  
 09       u:m,  
 10       °hh whO:'s twEnny: h° sIx years `OLD;  
 11       and  
 12       (0.9) ((click)) he's a !`VE:R!y: sweet gUy;=  
 13       = <<all> he jus' mOved to la`GUna.>  
 14       (.)°h[h]  
 15 JILL:       [U]Nhu[nh.]  
 16 JEFF:       [A:N]D-  
 17       (.) wE jus' (0.7)  
 18       <<laughing> whh° we\_jus' gOt in one of those? (0.2)  
     cOnver`SAtions where; >  
 19       °hh <<all> you jus' talk abou' `EVrything.>=  
 20       =?ih heh heh heh <<ingressive> huh>,  
 21 JILL: un[´HU:NH, ]  
 22 JEFF: [like like] r:Il po`LITical:\_an';  
 23       °hh a:nd h°  
 24       `EVrything from nOrm:?' (. )  
 25       <<creaky> um:> h°  
 26       <<l, dim> nOrm `NIXon.>  
 27       huh <<l, all> 'at's\_a ´BASketball player,>  
 28       °hh hn heh ↑huh;=  
 29 JILL: =ehh hnh

<sup>55</sup> Among others, the following modified spellings have been used: 'twenny' (=twenty), 'ril' (=real), 'evrything' (=everything), 'yiknow' (=you know), 'cuz' (=cause), 'catholic' (=catholic), 'sd' (=said), 'unnerstand' (=understand), 'gonna' (=going to).

- 30 JEFF: <<dim> †i said `NORM instEad of like (.) `PRESiden' (0.3)  
 †NIX[on],>
- 31 JILL: [Oh] ^RI::GHT;
- 32 JEFF: heh heh: <<ingressive> huh>
- 33 JEFF: <<one declination unit> from /†NIXon-  
 34 to like †rUsh /↓-LIMbaugh-  
 35 to a/†BORTion-  
 36 to cApital /†PUNishmen'-> h°
- 37 JILL: ((click)) <<nasalized> ^WO[::W.]>
- 38 JEFF: [like] h° (.)
- 39 -YEAH- =
- 40 =°h †yiknOw an' it's so `FUNny;  
 41 cuz he's<sup>?</sup> (0.2) he's a `CATHlic,  
 42 (.) °h (.) ((click)) `A:ND;  
 43 °hh <<creaky>`YIknow like;  
 44 I n:Ailed him on the contra`DICTion> <<all>you\_know;=>  
 45 =<<creaky> he's like prO capital †PUNishment-  
 46 (0.2) but<sup>?</sup> An:' pro `LIFE? ?ehhh°
- 47 JILL: ?un`HUNH[:? ]
- 48 JEFF: [i sd] †trY to ex`PLAIN 'at to me.=  
 49 =I don' <<laughing> unner`STA:ND<sup>56</sup>. †heh;>
- 50 JILL: °hh †`WAIT;=  
 51 =he said hE's: (.) /†`prO: / `cApital /`PUNishmen'-
- 52 JEFF: En pro ↓`LI[FE,]
- 53 JILL: [An'] pro `LIFE.  
 54 †`RI::GH[T. ]
- 55 JEFF: [that's] like<sup>?</sup> (.)  
 56 to †mE that's like a hU:ge contra`DICTion?  
 57 °hh ((click)) <<f> a:nd> [hE] said
- 58 JILL: [m:]
- 59 JEFF: he\_d he: `JUSTified it as;  
 60 °hh †well `SEE;=  
 61 =the<sup>?</sup> (0.2) th<sup>?</sup> (0.2) those little †EN<sup>?</sup> (.) the little<sup>?</sup>  
 (0.5) <<l> u:m> (.) †EMBryos? (0.4)  
 62 ((click)) thEy don' have a (0.2) de`CIS<sup>?</sup> (.) they  
 can't make a de†`CISion.

<sup>56</sup> This representation is not intended discriminatorily.

63 (0.5)  
64 JEFF: ((click)) YIknow;=  
65 ='n' thEy're gonna <<creaky> `DIE;=>=  
66 JILL: =↑thIs's what `HE was <whispery> saying.>  
67 JEFF: `YEAH;  
68 en `HE said that <<all, p, l> yIknow> pEople who  
(.) °hh commIt: ↑`CRI:MES;;  
69 °hh ↑yi`KNOW; (.)  
70 thA? (0.3) <<creaky> cAlls for the `↑DEATH penalty;  
71 (0.1) <<dim, acc> /thEy: /brOught it up/On  
them/↑`SELVES.>>  
72 JILL: (0.4) <<h> oh `WAIT; (.)  
73 jEff can you ↑`HANG on jus' a `sEc?>  
74 JEFF: <<h> ^YE:S.>

## 8. References

- Auer, Peter / Couper-Kuhlen, Elizabeth / Müller, Frank Ernst (1999): *Language in time. The rhythm and tempo of spoken interaction*. Oxford: Oxford University Press.
- Beckman, Mary E. / Ayers, Gayle M. (1994): *Guidelines for ToBI labelling, Version 2.0*. Columbus, OH, Ohio State University, Department of Linguistics.
- Beckman, Mary / Ayers Elam, Gayle (1997): *Guidelines for ToBI labelling; version 3.0*.  
[http://www.ling.ohio-state.edu/~tobi/ame\\_tobi/labelling\\_guide\\_v3.pdf](http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf)
- Beckman, Mary E. / Hirschberg, Julia / Shattuck-Hufnagel, Stefanie (2005): *The original ToBI system and the evolution of the ToBI framework*. In: Jun, Sun-Ah (ed.), *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 9-54.
- Bolinger, Dwight (1964/1972): *Around the edge of language: Intonation*. In: *Harvard Educational Review* 34, 282-293. [Reprinted in: Bolinger, Dwight (ed.) (1972): *Intonation*. Harmondsworth: Penguin, 19-29].
- Couper-Kuhlen, Elizabeth (1986): *An introduction to English Prosody*. Tübingen: Max Niemeyer and London: Edward Arnold.
- Couper-Kuhlen, Elizabeth (1993): *English speech rhythm. Form and function in everyday verbal interaction*. Amsterdam: Benjamins.
- Couper-Kuhlen, Elizabeth (2009). *Prosody*. In: D'hondt, Sigurd / Östman, Jan-Ola / Verschueren, Jef (eds.), *The pragmatics of interaction (Handbook of Pragmatics Highlights, vol. 4)*. Amsterdam: Benjamins, 174-189.
- Couper-Kuhlen, Elizabeth / Selting, Margret (1996): *Prosody in conversation*. Cambridge: Cambridge University Press.
- Croft, William (2003): *Typology and universals*. Cambridge: Cambridge University Press, xix-xxv (2nd edition).
- Cruttenden, Alan (1997): *Intonation*. Cambridge University Press (2nd edition).
- Du Bois, John W. (2000): *Santa Barbara Corpus of Spoken American English*. CD-ROM. Philadelphia: Linguistic Data Consortium.  
[\[www ldc.upenn.edu/Publications/SBC/\]](http://www ldc.upenn.edu/Publications/SBC/)
- Du Bois, John W. / Cumming, Susanna / Schuetze-Coburn, Stephan / Paolino, Danae (1992): *Discourse transcription*. Santa Barbara Papers in Linguistics, Vol. 4, Department of Linguistics, UCSB.
- Du Bois, John W. (2006): *Representing discourse. Transcription in action*. Resources for the representation of linguistic interaction.  
[\(http://www.linguistics.ucsb.edu/projects/transcription/ last accessed 17.04.09\)](http://www.linguistics.ucsb.edu/projects/transcription/).
- Edwards, Jane (1995): *Principles and alternative systems in the transcription, coding and mark-up of spoken discourse*. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (eds.), *Spoken English on computer: Transcription, markup and application*. Harlow: Longman, 19-34.
- Edwards, Jane A. / Lampert, Martin D. (1993): *Talking data. Transcription and coding in discourse research*. Hillsdale: Lawrence Erlbaum.
- Ford, Cecilia E. / Couper-Kuhlen, Elizabeth (2004): *Conversation and phonetics: Essential connections*. In: Couper-Kuhlen, Elizabeth / Ford, Cecilia E. (eds.), *Sound patterns in interaction. Cross-linguistic studies from conversation*. Amsterdam: Benjamins: 3-25.

- Fox, Anthony (2000): Prosodic features and prosodic structure. Oxford: Oxford University Press.
- Gilles, Peter (2005): Regionale Prosodie im Deutschen. Variabilität in der Intonation von Abschluss und Weiterweisung. Berlin / New York: de Gruyter.
- Goodwin, Charles (2007a): Interactive footing. In: Holt, Elizabeth / Clift, (eds.), Reporting talk: Reported speech in interaction. Cambridge: Cambridge University Press, 16-46.
- Goodwin, Charles (2007b): Participation, stance, and affect in the organization of activities. In: Discourse and Society 18, 53-73.
- Goodwin, Charles (2010): Constructing meaning through prosody in aphasia. In: Barth-Weingarten, Dagmar / Reber, Elisabeth / Selting, Margret (eds.), Prosody in interaction. Amsterdam: Benjamins, 373-394.
- Grabe, Esther / Nolan, Francis / Farrar, Kimberley (1998): IViE - a Comparative transcription system for intonational variation in English. In: The 5th International Conference on Spoken Language Processing (ICSLP-1998), Sydney, Australia, 1259-1262.  
[[http://www.isca-speech.org/archive/icslp\\_1998/i98\\_0099.html](http://www.isca-speech.org/archive/icslp_1998/i98_0099.html)].
- Grice, Martine (2006): Intonation. In: Brown, Keith (ed.), *Encyclopedia of Language and Linguistics, vol. 5, 2nd Edition*, Elsevier: Oxford, 778-788.
- Gussenhoven, Carlos (2004): The phonology of tone and intonation. Cambridge: Cambridge University Press.
- Heath, Christian / Hindmarsh, Jon / Luff, Paul (2010): Video in qualitative research. London: Sage.
- Jefferson, Gail (2004): Glossary of transcript symbols with an introduction. In: Lerner, Gene (ed.), *Conversation Analysis. Studies from the first generation*. Amsterdam: Benjamins, 13-31.
- Kern, Friederike (2008): Das Zusammenspiel von Prosodie und Syntax am Beispiel von Türkendeutsch. Habilitationsschrift, Universität Potsdam.
- Ladd, D. Robert (2008): Intonational phonology. Cambridge: Cambridge University Press (2nd revised edition).
- Leipzig Glossing Rules (2008): Conventions for interlinear morpheme-by-morpheme glosses. ([http://www.eva.mpg.de/lingua/pdf/LGR09\\_02\\_23.pdf](http://www.eva.mpg.de/lingua/pdf/LGR09_02_23.pdf), last accessed on 30.4.09)
- Mondada, Lorenza (2007a): Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. In: Discourse Studies 9, 195-226. [reprinted in Dijk, Teun van (ed.), *Discourse Studies*, London: Sage, vol. IV, 126-157].
- Mondada, Lorenza (2007b): Commentary: Transcript variations and the indexicality of transcribing practices. In: Discourse Studies 9, 809-821.
- Mondada, Lorenza (2008): Doing video for a sequential and multimodal analysis of social interaction: Videotaping institutional telephone calls. In: Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research 9 (3) ([www.qualitative-research.net/](http://www.qualitative-research.net/)).
- Nespor, Marina / Vogel, Irene (2007): Prosodic phonology. Berlin / New York: de Gruyter (2nd edition).
- Roach, Peter (1994): Conversion between prosodic transcription systems: "Standard British" and ToBI. In: Speech Communication 15, 91-99.

- Schegloff, Emanuel A. (2007): Sequence organization in interaction. Cambridge: Cambridge University Press, Appendix 1: 265-269.
- Selkirk, Elisabeth (1984): Phonology and syntax. The relation between sound and structure. Cambridge, Mass.: MIT Press.
- Selting, Margret (1995): Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation. Tübingen: Niemeyer.
- Selting, Margret (2010): Prosody in interaction: State of the art. In: Barth-Weingarten, Dagmar / Reber, Elisabeth / Selting, Margret (eds.), Prosody in interaction. Amsterdam: Benjamins, 3-40.
- Selting, Margret / Auer, Peter / Barden, Birgit / Bergmann, Jörg / Couper-Kuhlen, Elizabeth / Günthner, Susanne / Quasthoff, Uta / Meier, Christoph / Schlobinski, Peter / Uhmman, Susanne (1998): Gesprächsanalytisches Transkriptionssystem (GAT): In: Linguistische Berichte 173: 91-122.
- Selting, Margret / Auer, Peter / Barth-Weingarten, Dagmar / Bergmann, Jörg / Bergmann, Pia / Birkner, Karin / Couper-Kuhlen, Elizabeth / Deppermann, Arnulf / Gilles, Peter / Günthner, Susanne / Hartung, Martin / Kern, Friederike / Mertzlufft, Christine / Meyer, Christian / Morek, Miriam / Oberzaucher, Frank / Peters, Jörg / Quasthoff, Uta / Schütte, Wilfried / Stukenbrock, Anja / Uhmman, Susanne: Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung 10, 353-402 ([www.gespraechsforschung-ozs.de](http://www.gespraechsforschung-ozs.de)).
- Sorjonen, Marja-Leena (2001): Simple answers to polar questions: The case of Finnish. In: Selting, Margret / Couper-Kuhlen, Elizabeth (eds.), Studies in Interactional Linguistics. Amsterdam: Benjamins, 405-431.
- Streeck, Jürgen (2003): The body taken for granted: Lingerin dualism in research on social interaction. In: Glenn, Phillip / LeBaron, Curtis D. / Mandelbaum, Jenny S. (eds.), Studies in Language and Social Interaction. In Honor of Robert Hopper. Mahwah, N.J.: Lawrence Erlbaum, 427-440.
- Streeck, Jürgen (2009): Gesturecraft. The manu-facture of meaning. Amsterdam: Benjamins.
- Uhmman, Susanne (1991): Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Tübingen: Niemeyer.

## **Appendix 1:**

### **GAT-TO: GAT – Tutorial Online**

URL: <http://paul.igl.uni-freiburg.de/GAT-TO/>

GAT-TO is a multimedial, interactive platform in German introducing users to notating talk-in-interaction according to the GAT 2 conventions. GAT-TO is aimed at novices and advanced GAT transcribers alike.

GAT-TO covers the entire teaching and acquisition process: learners are introduced to the notation conventions and at the same time are provided with multimedial sequences for practice – an up-to-date training situation. Moreover, a simple restructuring of content allows users to employ GAT-TO in other contexts, such as in academic teaching.

Since it is webbased, GAT-TO can be used in a temporally and spatially flexible fashion. It is divided into 8 sets of modules, i.e. learners are introduced to individual parts of the system step by step.

The structured online material, which is specially adapted to teaching, also allows the supervisor to use it freely. Each module can be used recursively for nearly unlimited practice. Users are not forced to cover the modules in any fixed order. Each set of modules starts with a survey of the contents and goals of the module as well as with the prerequisites for covering it, which allows the learner to check his/her knowledge and switch to another module if needed.

## Appendix 2:

### Transcription editor FOLKER

URL: <http://agd.ids-mannheim.de>

One motivation for revising GAT was to change the conventions in a way that would facilitate their optimal and complete conversion into formal rules for digital processing. This is a basic pre-requisite for developing two kinds of tools applicable in the research process:

- editors for compiling transcripts which are sustainable in terms of archiving and which can be exchanged between researchers without loss of information,
- tools for automatic searches of transcripts (for instance in data bases).

The new transcription editor FOLKER is such a tool. It was developed by Thomas Schmidt, the developer of EXMARaLDA, for the project "*Forschungs- und Lehrkorpus Gesprochenes Deutsch*" [Research and teaching corpus of spoken German] (FOLK) and can be downloaded free of charge from the website of the *Archiv für Gesprochenes Deutsch* [Archive of spoken German] (AGD).

#### FOLKER

- optimally supports transcribing with GAT by automatically checking that the transcription conventions have been complied with,
- guarantees the compilation of transcripts in a format that is sustainable for archiving and that can be exchanged between researchers without loss of information and can be used with new program versions and on new platforms without necessitating adaptations.
- is therefore an ideal tool for the compilation of large, automatically searchable data bases of spoken language.

For more information on FOLKER (quick-start instructions, manual) and download (for Windows and MAC OS X) after free-of-charge registration, see the website of the *Archiv für Gesprochenes Deutsch* [Archive of spoken German] (AGD) at the address given above.

For all authors:

Prof. Dr. Margret Selting  
Universität Potsdam  
Institut für Germanistik  
Lehrstuhl Kommunikationstheorie und Linguistik  
Am Neuen Palais 10  
14469 Potsdam  
Germany  
selting@rz.uni-potsdam.de

For the translation and adaptation into English:

Prof. Dr. Elizabeth Couper-Kuhlen  
Finland Distinguished Professor  
Faculty of Arts  
University of Helsinki  
00014 Helsinki  
Finland  
elizabeth.couper-kuhlen@helsinki.fi

Dr. Dagmar Barth-Weingarten  
Hermann Paul School of Language Sciences  
Albert-Ludwigs-Universität Freiburg  
Starkenstr. 44  
79104 Freiburg  
Germany  
dagmar.barth-weingarten@hpsl.uni-freiburg.de

Veröffentlicht am 1.9.2011

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.