Transcription Guidelines: main tier

Authors:
Rasmus Steinkrauss (University of Groningen)
Cornelia Lahmann (University of Groningen)

Project:
L1 attrition and L2 acquisition: a perspective from *Kindertransport* survivors

Date: 01-12-2014

Content:

Introduction

The aim of our project is to analyze the level of language proficiency of oral history witnesses. To measure their language proficiency we will work within the CAF framework (Housen & Kuiken 2009; Van Daele, Housen, Kuiken, Pierrard & Vedder 2007) and analyze the following four aspects: the language the speaker uses (German or English), complexity, accuracy and fluency. To be able to do this it is necessary that the data is correctly transcribed in these regards and that the data is immediately annotated during the transcription process. Therefore, aspects such as codeswitching, errors of various kinds and the marking of pauses as well as questions such as what does in fact constitute a word and what not, what are one or two words, and what is actually considered an utterance are important.

All these aspects will already be decided on and annotated accordingly in the speaker line which is called the main tier line. The main tier line always starts with the speaker header (e.g., *XYZ) and gives the basic transcription of what the speaker said. Apart from pauses and other codes the main tier line contains only material that was actually said. Since we are mostly using data which was already transcribed by historians, the main tier lines might contain both too little and too much information. For example, they might contain material that was not actually uttered but added later e.g. to facilitate a historical understanding. For our needs such additional information should be marked as a comment within the main tier line since it is not supposed to be included in the analysis later on. At the same time, some words that were actually uttered might have been left out or abbreviated. It is crucial that these words are added or written out in order to analyze what was actually said.

Transcripts already made available by historians were automatically transformed into preliminary CHAT versions. The transformation is blindly generated by the computer and depends on the original transcripts. Therefore, be prepared to find many errors in the first version of each transcript. It might be, e.g., that utterances were transcribed either too long or too short or that they were accidentally marked as a comment because they did not start with a speaker header in the original transcript or because the transformation macro did not recognize a certain speaker header. It might also occur that the wrong language was given as the main language in the transcript. These errors must be changed manually.

General remarks for transcribing

- These guidelines mainly apply to the way the interviewee (XYZ) should be transcribed during the first 30 minutes of each interview. For the time after that, and for the interviewer and other persons, other rules apply – see below on p. 7.

- In case you do not know how to transcribe something, you should first insert a @Comment line with some information and the signs $$$. These sections can later on be easily tracked.

- Concerning the codes in square brackets (e.g., [//] or [*]): these usually apply only to the immediately preceding word. If the code is meant to apply to several words then you should insert the pointed brackets < > . Otherwise pointed brackets do not need to be used. To give you an example, if the transcript says "I am [x 3]" it means the speaker actually said: "I am am am". However, if the transcript says "<I am> [x 3]", then the speaker originally said "I am I am I am".

- Concerning written material within square brackets: Comments on the main tier line following [% do not have to correspond to CHAT's minimum requirements; thus, they could also contain *Umlaute*, numbers, etc. Correct replacement forms after [: however need to correspond to CHAT standards because they replace other material and will be analyzed.

- We advise you to read the entire transcript once before actually starting to transcribe. In case you notice that the interviewee/informant left the UK or the United States or any other English-speaking country to return to Germany or to settle elsewhere (e.g., Palestine), please let us know! These transcripts do not need to be transcribed further for the time being! Also, please indicate this in the corresponding excel file on the server.

Minimum standards for CHAT files

- CHAT only recognizes three **utterance terminators**: period (.), question mark (?) and exclamation mark (!).
These must appear only once in a main tier line, namely at the end.
Each main tier line has to end with one of those three utterance terminators.

    - Three periods (…) are often used in the original transcripts to indicate pauses or trailing off. They have to be changed into (.) or +… (see below).
    - multiple utterance terminators such as ?! need to be reduced to one

    Besides these utterance terminators only , and ; are allowed on the main tier line.

    - Brackets need to be erased and their content removed or changed into a comment (see below)
    - Colons (:) need to be resolved, e.g. to a new utterance if an enumeration begins
    - Hyphens (-) are either transformed into plus signs (+) or the hyphenated word is written as one word (see below)
    - Commas (,) appearing in numbers will be spelled out (see below)

- Words are separated by **spaces**.

- **Numbers** are not permitted as words or as part of words and are written out as one single word (see below).

- **Umlaute and sharp s (β)** on the main tier line (while for comments it does not matter) should be transcribed as ae, ue, ss etc., so that CLAN can automatically create a %mor line later.

- Only nouns (alle Nomen) in German, proper nouns (Eigennamen) in English and the pronoun I are **typed in upper case**. The beginning of an utterance should otherwise always be typed in **lower case**.

- A main tier line always only contains one (sub)clausal unit; several main tier lines can combine into one utterance (see below).

@-lines

- In the initial @-lines at the beginning of each transcript you should insert an @Transcriber line with the name of the person transcribing

- The @Participant line lists all the speakers in the following format:
  Speaker ID   name   role

  The following speaker ID's and roles are fixed:

XYZ	Subject	the interviewee
QQQ	Interviewer	

  All other persons:

ZZZ, ZZ2, ZZ3 etc.	Interviewer	another interviewer
	Partner	wife / husband of the interviewee
	Child	son / daughter of the interviewee
	Brother, Sister, Mother, Father	of the interviewee
	Family	other family of the interviewee
	Family_Friend	friend/companion of the interviewee
	Male, Female, Unknown	for all others

  If during the automatic generation of transcripts a speaker was forgotten, it should get assigned a new speaker ID and be added to the @Participants line.

  Some emigrants were interviewed at the same time (e.g., siblings) – in this case a single transcript suffices which should though be saved twice under different names and in which the headers and roles are differently arranged (e.g., the ZZZ brother becomes the XYZ subject).

- In the @Languages line the language primarily used in the interview is listed first, then the second language. In our case this should be either eng or deu.

  The order is important since CLAN assumes that everything was uttered in the language listed first and everything marked as a codeswitch was uttered in the second language (see below).
  The order might be wrong in the first transcript version and should in this case be changed (then please update the Excel-file on the server and save it there).
  Already marked @s should then of course be changed as well.

Minimum standards for QQQ and transcribing after the first 30 minutes

- Transcribing the interviewer (QQQ):

We are only going to analyze the utterances of the witnesses, not of the interviewers. Thus, the interviewers' utterances do not need to be transcribed as precisely – meaning that you do not need to **split the utterances** into (sub)clauses and that you may leave out all marks for **pauses, retracings, trailing offs, errors, spelling, etc.** However, whenever an utterance of the interviewer in some way affects the interviewee's utterance or vice versa, you should indicate this. Possible examples of such cases could be **completions or interruptions**. **Codeswitches** do not need to be marked either, but please do mark **numbers and names** (= everything on page 11 in this manual). Apart from this, the interviewer's utterances only need to correspond to the **general CHAT requirements** (e.g., speaker header, punctuation, lower-case letters etc.).

Feedback signals from the interviewer (such as *mhm*) should only be transcribed when they occur between two utterances of the interviewee, or interrupt an interviewee's utterance. If a feedback signal of the interviewer does not affect the interviewee's flow of speech, just disregard it.

- Transcribing the interviewee (XYZ) after the first 30 minutes:

For the time being, we are only interested in analyzing the first 30 minutes of each interview. Therefore, only the first 30 minutes of the interviewee (XYZ) need to be transcribed according to these guidelines. However, we do ask you to transcribe XYZ (and QQQ) in the remaining interview according to the minimum standards we have defined for transcribing the interviewer (see above). In addition, we would like you to mark stretches with interesting content with @Bg and @Eg (see p. 25) and add all relevant information to the Excel file on the server. All this may be done without listening to the interview – you may reshape the transcript on basis of the text alone.

After the first 30 minutes, please insert an @Comment-line stating that the detailed transcription of XYZ ends here.

- Remarks about the London Data:

The London Data may contain a lot of utterances that are erroneously marked as comments. This does not need to be changed after the first 30 minutes.

Also, it seems that during the last part, the informants often show the interviewer photographs and other pieces from the past and describe those. There is no need to mark locations (@z:geo), names (@z:name), or people's names (@z:per) as well as dates (@z:num) in this part, you only need to look out for interesting passages (in which the informants "judge" their English or German skills or discuss identity issues), marking those with @Bg & @Eg.. Please mark the start of this final part of the interview with an @Comment-line stating that the part describing photographs follows.

What belongs to a main tier line?

First of all: what is a main tier line? A main tier line is a line starting with *, thus a line containing what a speaker says - such as:

```
*ABC:  there we go !
```

When transcribing, you will often be faced with the question when to end a line, and when to begin a new one. Note that line here refers not to one line on the screen, or to one line in your document, but to everything until the next line starting with *. This section wants to assist you in deciding what belongs to one main tier line and what belongs to the next main tier line.

- **In general, one main tier line consists of one clause**. A clause is defined as everything governed by one finite verb. The simplest case is a simple, independent clause:

  ```
  *ABC:  I went to Hamburg.
  *ABC:  I don't understand that !
  ```

- If there is a **subordinate clause** contained in another clause it should be **transcribed on the next line**:

  ```
  *ABC:  I don't understand +/.
  *ABC:  +, what he is saying !
  ```

  So even if, strictly speaking, *what he is saying* is also governed by *understand* (it is the direct object of *understand*, just like *that* in the utterance above), it is transcribed on a line of its own because it contains a verb of its own.

- **Subordinate clauses have to contain at least one element besides the verb.**[1] Therefore, the first utterance is transcribed on one main tier line while the second one is split:

  ```
  *ABC:  ich hatte keine Gelegenheit zu fragen .

  *ABC:  ich hatte keine Gelegenheit +/.
  *ABC:  +, ihn zu fragen .
  ```

- **If two main tier lines belong to the same utterance** like in the cases above, **they have to be connected** so CLAN can see that they belong together. This is achieved by using the interruption and self-completion markers **+/. and +,** at the end and the start of a main tier line, respectively. Note that +/. does not mark any interruption here - it is simply required by CLAN for technical reasons.

- Apart from clauses, there might be phrases that do not contain any verb. Those **subclausal units** also go on one main tier line.

  ```
  *YYY:  wie lange warst du da ?
  *ABC:  fuenf@z:num Jahre .
  ```

  Again, if there is some subordinate clause associated with the subclausal unit, the subordinate clause goes on a main tier line of its own and +/. and +, are used to mark that the two main tier lines together form one utterance:

  ```
  *ABC:  five@z:num years +/.
  *ABC:  +, while he was sick .
  ```

---

[1] I.e. subject, object, adverbial or complement (in effect: everything that is not (part of) a verb).

If two main tier clauses belong to the same utterance, they have to be connected using +/. and +, (see above). But how do you decide where one utterance ends and the next one starts? Use **intonation and pause as indications**:

- If the intonation suggests that something has been planned as a single utterance / is meant to belong to the previous material, connect the main tier lines using +/. and +, .

- If there is a noticeable pause and a different tone unit starts, don't connect the main tier lines.

Possible examples are the previous utterance and utterances such as:

```
*ABC:  and then she talked to him +/.
*ABC:  +, and then looked in the other direction .

*ABC:  this was really dangerous +/.
*ABC:  +, because they could have heard him .
```

Similarly, when someone continues after having been interrupted, use intonational cues to decide whether the utterance following the interruption is a completion (like below) or not:

```
*ABC:  oh that (i)s a big problem +/.
*YYY:  oh@i no !
*ABC:  +, because I don't have the keys either .
```

Please be aware of the fact that **one utterance may consist of more than just one clause with its associated subordinate clauses**. For example, the following utterance consists of a main clause, a clause associated with that main clause, and another main clause starting with "and" :

```
*ABC:  and then the neighbours told us +/.
*ABC:  +, to pack our stuff +/.
*ABC:  +, and we left shortly after that .
```

Whether the last main clause is connected to the preceding clauses using +/. and +, or not is decided purely on basis of intonation and pause.

Special cases:

- If a speaker uses a **parenthesis (Schaltsatz, eingeschobenen Satz)** in the middle of an utterance, the way you should transcribe this depends on whether the part after the parenthesis is a new utterance or not.

  To determine if it is a new utterance, use syntactic cues: Is the last part of the utterance a completion of the first or does it start a new clause? In cases of doubt, look at the intonation: Is there a clear intonation break / pause or not?

  If the last part continues the first, simply link everything into one utterance:

  ```
  *ABC:  das war +/.
  *ABC:  +, wann war das +/.
  *ABC:  +, 1957 .
  ```

  If the last part starts a new utterance, mark the first part as a self-interruption (+//.) and all following parts as utterances of their own:

  ```
  *ABC:  das war +//.
  *ABC:  wann war das ?
  *ABC:  das war 1957 .
  ```

Here are some more examples of parentheses where everything should be linked:

```
*ABC: ah@fp but ah@fp I suppose +/.
*ABC: +, the feeling of the Nazis coming to power +/.
*ABC: +, I mean for such a small child +/.
*ABC: +, was not real .

*ABC: ah@fp and so finally ah@fp America worked you_know
+/.
*ABC: +, that we got an affidavit +/.
*ABC: +, to come to the States .

*ABC: weil ich selber Kinder interviewt [///] +/.
*ABC: +, ah@fp was heisst das +/.
*ABC: +, gesprochen habe .
```

Note that the last example also contains a reformulation. ABC's utterance is not simply continued after *interviewt*, but *interviewt* is reformulated as *gesprochen*.

[Unlike the German "wann war das" example, the "was heisst das" example is a word search. Because this can be seen as a kind of dysfluency or error, we will mark this later on, on an extra tier (thus not on the main tier line).]

- When speakers **construct an utterance together (scaffolding)**, you should mark the incomplete utterance as interrupted (+/.) and the "help" of the other speaker as an other-completion (++).

```
*ABC: ich hab(e) mit ihr +/.
*YYY: ++ gesprochen .
```

When the first speaker then takes up the suggestion of the second speaker, the way you transcribe this depends on whether that turn continues the original utterance (1st example below) or starts a new one (2nd example). Again, use syntactic (and intonational, if necessary) cues:

```
*ABC: ich hab(e) mit ihr +/.
*YYY: ++ gesprochen .
*ABC: +, gesprochen auf der Strasse .

*ABC: ich hab(e) mit ihr +/.
*YYY: ++ gesprochen .
*ABC: ich habe mit ihr gesprochen .
```

Taken together, this means that we define as an utterance
- one or several independent clause(s) <u>or</u> subclausal unit(s)
- <u>together with</u> any subordinated clauses associated with either.
One utterance may be distributed over several main tier lines, with one (sub)clausal unit per main tier line, connected by +/. and +, . Intonation and pauses indicate the end of an utterance.
This definition is based on Foster et al.'s (2000) definition of the *AS-unit*.

In the end, what goes on one main tier line and what comprises one utterance is often a matter of interpretation. No matter how you decide in cases of doubt, it is important that you try to be consistent across similar cases.

Numbers

@z:num Numbers are always written out and should in addition be marked with @z:num. Only mark numbers as such if they could be written as numbers in the text, not if they are a part of other words (e.g., "dreissiger Jahre"). Ordinal numbers such as first, second, or third do not need to be marked either.

While German numbers may be one word, English numbers should be split up in several words. Only transcribe what was in fact said.

- If, e.g., *hundred* in 1939 was not pronounced, it will not be transcribed:
  `nineteen@z:num thirty@z:num nine@z:num`

- Parts of numbers such as a comma and percent are written out: 3,4% is
  `drei@z:num Komma vier@z:num Prozent`
  `three@z:num point four@z:num percent`

- 2 ½: `zweieinhalb, two@z:num and_a_half`

Names and Anonymization

Anonymization serves to prevent that someone might recognize real persons or place names in the data. The anonymization itself will be established later. However, words that have to be anonymized should already be marked now. Generally, you can just use common sense when deciding what should be marked for anonymization and what not.

@z:geo Added to names of locations (cities, streets):

- `Gänsemarkt@z:geo, Wien@z:geo, Lower_Stowe@z:geo`

Names of countries will not be anonymized later on and need not be marked with @z:geo.

@z:per Added to a name of a person mentioned:

- `Edith@z:per, Martin_Meier@z:per, Onkel Peter@z:per`
  (note: no _ in the last example!)

Famous people's names do not need to be marked:

- `Heinz_Ruehmann, Theodor_Herzl`

@z:name Added to other names if they could unveil the identity of real persons:

```
*ABC: mein Vater war Journalist beim
      Hamburger_Fremdenblatt@z:name .
```

If such names do not serve identification purposes the @z:name can be dropped:

```
*ABC: this was mentioned in the Washington_Post .
```

Compound words, phrases, collocations

The question about what can be considered one (single) word is relevant for two reasons:
- The length of utterance in words is a possible measure for the complexity of utterances.
- CLAN will automatically create the %mor line for words it can recognize, i.e. it will supply morphological information about all the words of an utterance.

To make it recognizable for CLAN that something is to be counted as one word there are two options: the plus sign (+) and the underscore (_).

+   Single words that are made up of several words, such as compounds, should generally be written as single words. In some cases however, the plus sign may be inserted to indicate boundaries within such words:

- for unusual / newly invented compounds in German:
  `Kindertransport+Zug, Kristallnacht+s+Chaos`

  English compounds should always be written as one word.

- if usually there would be a hyphen (-):
  `S+Bahn, Bahnhof Berlin+Friedrichstrasse,`
  `Horst+Wessel+Lied, blue+green`

  However, hyphenated words should rather be written as one word if possible. English numbers are written as several words (see above).

_   Use the underscore for several words which are so closely related that they are considered a fixed phrase or collocation (like one long word). These are expressions such as

- `you_know, french_fries, Big_Ben, in_god_we_trust,`
  `wie_gesagt, Bar_Mizwa, Voelkischer_Beobachter`

- `Rita_Kornbach@z:per, Adolf_Hitler`
  (however, NOT: aunt_Hilde, Herr_Kleinschmidt: these are considered as two separate words!)

Also, use the underscore for acronyms where each letter is pronounced:

- `U_S_A, N_S_D_A_P, U_Bahn, O_P_Saal`

Acronyms which can be pronounced as a single word (e.g., *UNO, NATO*) are transcribed as normal words.

The decision to mark something as a phrase with _ may be made on a speaker-by-speaker basis: if you feel that a certain speaker has a favorite phrase and uses it so often that for her/him it constitutes a fixed phrase, you may use _ .

Because such decisions will strongly influence the results for measuring the length of an utterance the underscore should be moderately used. In case of uncertainty, please mark the utterance with @Comment: $$$ (see above).

Epistemic phrases (Ich glaube, I think, etc.)

In spoken speech, speakers often indicate their level of certainty about what they are saying by adding phrases like *I think, I guess, Ich glaube, Ich denke, etc.* to their utterance. The stance someone takes vis-à-vis the reliability/truth of his or her proposition is called epistemic stance, and phrases indicating this stance are called epistemic phrases.

An example of an epistemic phrase is *I think / Ich glaube* in the following example:

- I think it happened in 1934 / Ich glaube, das ist 1934 passiert.

In the example, the speaker simply adds *I think / Ich glaube* to the main clause *It happened in 1934 / Das ist 1934 passiert* to indicate the degree of certainty of the information in the main clause.

It is important to make a difference between epistemic phrases and other, similar (or even identical) phrases that are used differently. Compare the following sentences to the ones above:

- I think that it happened in 1934 / Ich glaube, dass das 1934 passiert ist

In this example, *I think / Ich glaube* are not added to a main clause like the epistemic phrases in the first example are. Rather, *I think / Ich glaube* are main clauses of their own, and they each start a subordinate clause beginning with *that / dass*. Epistemic phrases however are never followed by *that / dass*.

Another difference is that as epistemic phrases always indicate a speaker's stance, they always contain *I / ich*. *You think / Du glaubst* is therefore never an epistemic phrase, even in cases where it is not followed by *that / dass*:

- You think it happened in 1934 / Du glaubst, das ist 1934 passiert.

Thirdly, epistemic phrases always combine with another proposition. This is because they indicate the epistemic stance of the speaker vis-à-vis that proposition. A simple answer like *I think so* or *Ich glaube auch* does not contain an epistemic phrase.

To sum up, epistemic phrases

- **always contain *I / ich***
- **are in present tense / Präsens**
- **are never followed by a complementizer like *that / dass*
  (or, e.g., *if, whether, ob, falls*...)**

The following examples are to be treated as epistemic phrases, but only if they are not followed by the *that* complementizer:

- I (don't) think, I believe, I guess, I suppose, I know, I mean, etc.
- ich glaube, ich meine, ich finde, ich denke

This list should cover everything. If you encounter a phrase that you think is an epistemic phrase that is not on the list, please contact us!

Epistemic phrases are
- **to be transcribed with an underscore**
- **to be marked with @z:ep**
- **not to be split from the remaining utterance with +/.**

Examples:

```
*ABC: I_think@z:ep I met her in Sydney@z:geo.
*ABC: I believe +/.
*ABC: +, that she was already retired then.
*ABC: du denkst auch +/.
*ABC: +, du weisst alles, ne [: nicht] ?
*ABC: I think so, yes .
```

Spelling, Acronyms

@l   When something is spelled out, append @l to each single letter. If the material that is spelled out needs to be anonymized, don't spell it out but insert a comment instead such as : [% spells out his name, Cohen].

_   For words which consist of individually pronounced letters:

U_S_A etc. (see previous page)

In case spelling takes place in the second language, an @s to indicate codeswitching should be added (with @s last, yielding @l@s). Titles such as Dr. or Mrs. have to be written out: `Doktor, Missus`

Incomplete, contracted and dialectal forms

() and '   When speaking rapidly speakers often drop sounds. In the transcript these sounds appear in parentheses to restore the complete word:

`(ei)n, (den)n, he (ha)s, I (woul)d`

For transcribing in English you may also use apostrophy ` if it is commonly used and not ambiguous:

`I'm` for *I (a)m*, `don't` for *do n(o)t*
(but not `he'd`, because that could mean both *he (ha)d* and *he (woul)d* !)

There is no need to use parentheses sparingly, they may appear often!

[:]   Non-standard forms and dialectal forms cannot be automatically recognized and annotated by CLAN. Therefore, they are transcribed as heard and repeated in the complete, standard form in square brackets:

```
*ABC: ick [: ich] hab(e) (eine)n Appel [: Apfel]
      gegessen .
*ABC: whyncha [: why don't you] just be quiet !
```

Uh, Oh, Ääh, Mhm, Boing

ah@fp	filled pause	Pauses filled with expressions such as "uh", "oh", etc. are transcribed as ah@fp no matter what was said. E.g., "This was ehmmmmm Amsterdam" becomes `*ABC: this was ah@fp Amsterdam .`

oh@i	exclamations / interjections	Exclamations such as Ah!, Oh!, Aua!, Hah! are all transcribed as oh@i. "Mmm delicious!" and "Aua das tut weh!" thus become
		
		`*ABC: oh@i delicious !`
		`*ABC: oh@i das tut weh !`

hm@i	feedback / communicators	Hm@i will be used throughout to indicate any sort of feedback to the listener, independent of the form actually used. YYY's "Uhhhuh" becomes:
		
		`*ABC: he always worked a lot .`
		`*YYY: hm@i .`
		
		Hm@i does not need to be used for the interviewer as his utterances are not analyzed.
		Standard words such as yes, well, Nein, etc. are transcribed as standard words without special marking.

@o	onomatopoeia (Lautmalerei)	Expressions which imitate natural sounds are marked with @o: `*ABC: and splash@o he fell in the water .`

Stutters, babbling, non-finished words and non-words

   &   If the spoken material is acoustically recognizable but does not make sense, e.g. due to babbling, stuttering, stumbling, incomplete words etc., then use & for transcribing:

Stutter:

```
*ABC: and then I &f &f fell .

*ABC: and &i I_mean@z:ep mother was ill .
```

Stumbling and retracing:

```
*ABC: and then I &fet [//] fell .
```

In the second utterance "fet" is obviously a failed attempt to say "fell" – therefore it has to be marked as retracing as well (see below).

Acoustically difficult or unintelligible material

[?]   If you think you recognize a word but are not entirely sure, e.g., because the word was not clearly pronounced then you should mark the word as unclear [?]:

```
*ABC: I have two [?] brothers .
```

[=?]   If you are not sure whether the one or the other was said then you type into [=? Text] the alternative transcription:

```
*ABC: we want <one or two> [=? one too].
*ABC: I habe zwei [=? drei] Brueder.
```

xx, xxx   If something is completely unintelligible, e.g., because it was too soft-spoken or due to noise in the background then you transcribe with xxx or xx:

```
*ABC: and then xx I fell.

*ABC: xxx.
```

xxx either implies several unintelligible words or an entire utterance. However, if it is clear that we are dealing with several unintelligible words, then you should use xx for each word.

Nonverbal sound events

If someone cries, coughs, or sneezes we can distinguish whether someone does so during speaking or instead of speaking.

[=!]   Added if sound events accompany speaking. Thus, if someone whines "I don't know" or screams "Oh my god!" we transcribe it as:

```
*ABC: <I don't know> [=! cries].
*ABC: oh_my_god [=! yells]!
*ABC: <oh@i my god> [=! yells]!
```

The last two options depend on whether you consider „Oh my god" as a collocation (i.e. one word) or not.

&=   Added if sound events occur instead of speaking.

```
*ABC: &=pants .
*ABC: &=coughs .
```

To denote such sound events an English verb in the 3rd person singular should be used. Here's a list that you can use – if you feel this list does not cover your case, use a new word and send it to me so we can add it to this list:

coughs (hustet), cries (weinerlich), gasps (schnauft), groans (stöhnt), hums (summt), inhales (atmet ein), laughs (lacht), moans (jammert), mumbles (murmelt), sneezes (niest), sniffs (schnieft), sighs (seufzt), sings (singt), swallows (schluckt), whistles (pfeift), whispers (flüstert), yells (schreit), vocalizes (lallt)

Material that is not to be transcribed/ comments on the main tier line

www   Something that was said but is unimportant for both the conversation and the analysis can be transcribed as www. Examples are real names (see for anonymization above) or comments made by the interviewer about the camera guy etc. However, in case something like this has already been transcribed it is not necessary to transform it first into www.

[%]   Sometimes it is useful to insert a comment immediately in the main tier line and not only afterwards an @Comment line. This might be the case if comments made by the interviewee or the historians were added later which were in fact not uttered. These could in principle be left out but since they are already there, they can simply be treated as comments. This can be done with [%]:

```
*ABC: er war im Santa_Fu [% Gefängnis Fuhlsbüttel]
      inhaftiert .
```

Just like other info in [square brackets], a [%] comment refers to the preceding word or <several preceding words> . Therefore, [%] may not occur at the beginning of a main tier line.

Within a main tier line: repetitions and reformulations

[/]	Repetition	*Simple Repetition*
		
		The [/] symbol is used in those cases when a speaker begins to say something, stops and then repeats the earlier material without change:
		
		`*ABC: <I wanted> [/] I wanted to ask her .`
		
		In case there is a pause marked by (.), the utterance may neverthereless be considered a simple repetition, e.g. :
		
		`*ABC: he is [/] (.) is pausing .`
		
		Don't mark something as a repetition if it is repeated purely for rhetorical effect, such as:
		
		`*ABC: I was very very angry !`
		
		Also, don't mark something as a repetition if the material following [/] is different from the material preceding [/], for example when there is an ah@fp between the repeated word(s). In these cases, use the retracing symbol instead (see below):
		
		`*ABC: I [//] ah&fp I wanted to ask her.`

[x N]	Multiple Repetition	*Multiple repetitions*
		
		If the same word is repeated several times, instead of transcribing
		
		`*ABC: <it's it's it's> [/] it's like a dog .`
		
		use the [x N] marker, with N being the number of times the word is repeated:
		
		`*ABC: it's [x 4] like a dog .`
		
		This form indicates the fact that a word has been repeated four times. The [x N] marker should be used for a minimum of 3 repetitions.

Within a main tier line (continued)

[//]   Retracing   *New attempt for the same utterance / Neuer Anlauf*

This symbol is used when a speaker starts to say something, stops, repeats the basic phrase, changes the syntax but maintains the same idea. In retracing with correction, **it is necessarily true that the retraced material is different from what follows the retracing symbol**:

```
*ABC: <I wanted> [//] ah@fp I thought I
      wanted to ask Margie .
```

Retracing with correction can combine with retracing without correction, as in this example:

```
*ABC: <the fish is> [//] the [/] the fish
      are        swimming .
```

[///]   Reformulation   *Full and complete reformulation*

Sometimes retracings involve full and complete re-formulations of the message without any specific corrections:

```
*ABC: <all of my friends had> [///] ah@fp
      we had decided to go home for lunch .
```

Please remember to put <> around the material which is dropped, i.e. reformulated.

For a better readability of the transcript, you might consider beginning a new line after [///] and linking the two lines with +/. and +, where this is appropriate:

```
*ABC: <all of my friends had> [///] +/.
*ABC: +, we had decided to go home for
      lunch .
```

When none of the material being corrected is included in the retracing, it is better to use [///] than [//].

At the end of a main tier line: trailing off and interruptions

+…	Trailing off	The trailing off or incompletion marker is the terminator for an incomplete, but not interrupted, utterance **where the speaker is not aware of the fact that the utterance is incomplete**. Therefore, trailing offs cannot normally be followed by a self-completion (+,) .
		Trailing off thus occurs when speakers • shift attention away from what they are saying, sometimes even forgetting what they were going to say • do not finish an utterance but feel that it is finished.
		Trailing off does not occur when a speaker can't find a word and another speaker completes the utterance (scaffolding). This is marked with +/. and +, (see below).
		When a speaker completes an utterance and only the intonation indicates that the words are left "hanging in the air", transcribe it as a normal utterance (with a period, not as trailing off).
		Please make sure Word does not change your three separate periods into one sign with three periods.

+/.	Interruption	The interruption symbol is used for an utterance that is incomplete because one speaker is interrupted by another speaker:
		
		`*ABC: what did you +/.`
		`*YYY: Mommy .`
		`*ABC: +, with your spoon .`

		Note that the last utterance completes the first and is therefore marked as a self-completion (see next section).
		Apart from this, the interruption symbol is used simply to chain two main speaker lines together into one utterance, without any real interruption occurring (see "What belongs to a main tier line").
+//.	Self-Interruption	A self-interruption occurs if someone interrupts his own current utterance to continue with a next utterance that does not complete the interrupted utterance. The next utterance is therefore not marked as a self-completion:
		*ABC: smells good enough for +//.
		*ABC: what is that ?
		If a speaker interrupts his current utterance only to continue with it in a different way, this is marked as a retracing or reformulation (see previous section).

At the beginning of a main tier line: continuations

+,	Self-Completion	The self-completion symbol is used at the beginning of a main tier line to link an utterance to a preceding main tier line of the same speaker.

In effect, the symbol is used in two kinds of situations:
- After an interruption, when the interrupted speaker completes the interrupted utterance (see Interruption above)
- To indicate that a clausal unit (such as a subordinate clause or parenthesis) belongs to the same utterance as the preceding material of the same speaker (see "What belongs to a main tier line").

The part of the utterance that is being completed needs to end with +/.

The symbol +, is not normally used after trailing off. |
| ++ | Other-Completion | The ++ symbol marks "latching" or the completion of another speaker's utterance as in the following example:

```
*ABC: if Bill had known +...
*YYY: ++ he would have come .
```

It also marks scaffolding, thus if one speaker misses a word and another speaker helps in constructing the utterance:

```
*ABC: das war wohl in +/.
*YYY: ++ Berlin .
```

The preceding utterance of the other speakers ends with +... if the other speaker is not aware of the fact that his/her utterance is incomplete. Otherwise it ends with +/. such as in the case of scaffolding. |

Codeswitching

Of special importance for our data are codeswitches (CS). How CLAN interprets a marked CS later on depends on which language was listed as first and second in the @Languages line at the beginning of each transcript (see above). How a CS is marked depends on whether it applies to one word only or an entire utterance.

@s	Codeswitches of single words:

```
@Languages: deu, eng
[...]
*ABC: Wir waren damals, well@s, sehr arm .
*ABC: Es war wirklich schwierig, you_know@s .
```

[- lan]	In case of a complete utterance in the language listed as second, please insert [- lan] in front of the utterance (before the first word but after the tab) where lan refers to the language in which the utterance was produced (deu for German, eng for English):

```
@Languages: eng, deu
[...]
*ABC: check Wir waren damals sehr arm .
```

Within such utterances words in the language listed first are once again considered codeswitches:

```
@Languages: eng, deu
[...]
*ABC: [- deu] Wir waren damals, well@s,  sehr
      arm .
```

Special cases: what counts as a codeswitch, what doesn't?

No codeswitch:

- Words for which there is no corresponding word in the other language.

  E.g., the use of the word *Reichstag* in English is not a codeswitch. In contrast, the use of *The_White_House* in German is a codeswitch as the White House is commonly referred to as *weisses_Haus* in German. Most of these instances are names, see next point

- Names that are pronounced in their respective „native" language

- Deliberal translations, such as *Rassenkunde* in the following example:

```
*XYZ:  I couldn't attend the Rassen@s class, the
       Rassenkunde .
```

Codeswitch:

- Letters or numbers uttered in the second language, e.g. acronyms such as `D_V_D@s` (which in this case was pronounced /di-wi-di/ in a otherwise German utterance) or during spelling (then we add @l@s)

In cases of doubt please add a @Comment: $$$ line.

If someone has a clearly pronounced German accent while speaking English or the other way around, please add this to our Excel file on the server and save the file on the server.

Errors

[*]	Morphological, syntactic and lexical errors (e.g., interferences from the other language) should be marked with [*]. If the error is not confined to one word, pointed brackets (< >) need to be used to mark the „area" within which the error occurs:

```
*ABC: We had, you_know, Quartettabends@s [*]
        every week.
*ABC: Und er hat mir <angeboten eine Stelle>
        [*].
```

The exact type of error will be coded later on in the %err tier. The area marked by < > can then still be changed.

[* p:w] [* p:n]	Both symbols mark phonological errors. If someone pronounces a word erroneously or unnaturally (independent of a possible accent which should be mentioned in the Excel file, see previous section) this word will be marked accordingly.

The word itself (if possible) should be transcribed as heard and resolved afterwards in [:]. Then follows [* p:w] or [* p:n] depending on whether the uttered word was a word or a non-word:

```
*ABC: I was ferry [: very] [* p:w] happy .
*ABC: It was not keer [: clear] [* p:n] to me .
```

Whether the wrong pronunciation is transcribed in "German" or "English" does not matter (e.g., for the second utterance you could either transcribe it as kiehr or keer).

[* d]	This symbol marks a dysfluency within a word. Just transcribe the word as you hear it and add the correct form afterwards in [:] and add [* d]:

```
*ABC: I was insuhside [: inside] [* d] .
```

0	If an error occurs due to a missing word, fill in the gap for the missing word beginning with a 0 (zero) and followed by the error marker [*]:

```
*ABC: many in Germany I couldn't keep in touch
        0with [*], they had all disappeared .
```

If you are not sure whether a word is indeed missing you should nevertheless transcribe it as missing.

Pauses

(.)   Pauses between words are marked with (.), (..) or (...) depending on their length (just use what you think is best). The exact length will be measured later with PRAAT and the periods substituted by exact measures, e.g. (2.4).

Pauses between utterances only need to be marked when they are clearly longer than usual. In that case, you can place a pause marking at the end of the utterance that is followed by the pause:

```
*ABC: I don't know (..) .
```

^   Pauses within words, i.e. between syllables could also be an indicator for dysfluencies. They are marked with ^:

```
*ABC: Das war in Berlin+Schoene^feld !
```

Lengthening

:   If syllables within words are lengthened set the colon accordingly: „Waaas" becomes `wa:s` .

Stress

[!]   If a single word or several words are especially stressed they should be marked with [!].
„Das hab ich dir **schon tau-send Mal** gesagt!!" becomes:

```
*ABC: das hab(e) ich dir <schon tau^send Mal>
        [!] gesagt !
```

Quotations

["]   In case the speaker does not make up his words himself but quotes
        someone or something verbatim, such quotes should be marked
        accordingly:

```
*ABC: <fest gemauert in der Erden> ["] +...
*ABC: und als er <raus hier> ["] sagte bekam ich
      Angst .
```

        In the second sentence, it is very questionable if the speaker actually
        cites someone word by word or just imitates the other speaker. As a
        rule of thumb, something should be marked as a quotation when the
        speaker has pulled his utterance as a made-up string from memory and
        not constructed the utterance himself word by word.

        Be careful not to insert a typographic quotation mark " but a normal
        quotation mark ". (In Word, go to *Insert > Symbol* and chose the third
        sign in the "Font: normal text / Subset: Basic Latin" list.)

Stretches with interesting content

@Bg   Stretches of conversation that are particularly interesting for analysis
       are termed „gems" in CHAT. They start with @Bg (begin gem) and end
@Eg   with @Eg (end gem).

       In our case, interesting stretches are those that contain relevant
       content (they need not be particularly interesting linguistically). A
       stretch is relevant when the speakers talk about their linguistic
       experiences, how they felt towards the new or the old language, how
       they learnt to speak English, how they felt about their own identity, etc.
       If you encounter such a stretch, insert a line containing just @Bg (no
       tab) before the stretch, and a line containing just @Eg after the stretch.
       You should also add a @Comment line containing *** and maybe add a
       comment if you feel like it.

Standardized spellings

In the German transcripts, please use the following spelling standards for these
words (mostly related to Jewish culture and religion):
*Aschkenasen / Aschkenasim, aschkenasisch, Babi_Yar, Bar_Mizwa, die
Chanukka, Chanukka-Leuchter, Chassiden / Chassidim, Chassidismus, Ghetto,
der Goi, die Gojim pl., Holocaust, Iwrit, Jahwe, Jeckes, jeckisch, Jehova,
jiddisch, das Jiddische, der Jom_Kippur, Kabbala, Kaddisch, Kibbuz, Kibbuzim,
Kibbuznik, koscher, Makkabiade, Masel Tow, Menora, das Pessach, Pogrom,
das Purim, Rabbi, Rabbinat, Rabbiner, Rebbe, das Rosh ha-Shanah,
Schabbat, Schabbes, Schalom, Sepharden / Sephardim, Shoah, Sintezza, Sinti
und Roma, Sinto, die Tora, shivah (seven-day period of the mourning),
minyan (a prayer service with the minimum quorum of ten worhsipers)*

In the English transcripts, use the spelling that is given, but take care to insert an
underscore if two words belong together:
Tom_Yov