

List of CLAN Header Tiers

Hidden Headers

There are three hidden headers that appear before the @Begin header. These are the @Font header, the @UTF8 header, and the @ColorWords which appear in that order.

Header	Function
@Font:	This header is used to set the default font for the file.
@UTF8	All files in the database use this header to mark the fact that they are encoded in UTF8. If the file was produced outside of CLAN and this header is missing, CLAN will complain and ask the user to verify whether the file should be read in UTF8. Often this means that the user should run the CP2UTF program to convert the file to UTF8.
@ColorWords	This header is used to set the color words for the file.

Initial Headers

CHAT has seven initial headers. The first six of these – @Begin, @Languages, @Participants, @Options, @ID, and @Media – appear in this order as the first lines of the file. The last one @End appears at the end of the file as the last line.

Header	Function
@Begin	This header is always the first visible header placed at the beginning of the file. It is needed to guarantee that no material has been lost at the beginning of the file. This is a “bare” header that takes no entry and uses no colon.
@Languages:	This is the second visible header; it tells the programs which language is being used in the dialogues. Here is an example of this line for a bilingual transcript using Swedish and Portuguese. @Languages: sv, pt
@Participants:	This is the third visible header. Like the @Begin and @Participants headers, it is obligatory. It lists all of the actors within the file. @Participants: SAR Sue_Day Target_Child, CAR Carol Mother
@Options:	This header is not obligatory, but it is frequently needed. When it occurs, it must follow the @Participants line. This header allows the checking programs (CHECK and the XML validator) to suspend certain checking rules for certain file types.
@ID:	This header is obligatory. It is used to control programs such as STATFREQ, output to Excel, and new programs based on XML. To facilitate typing of these headers, you can run the CHECK program. Then the program fills in the missing information automatically. The form of this line is: @ID: language corpus code age sex group SES role education custom
@Media:	This header is used to tell CLAN how to locate and play back media that are linked to transcripts. The first field in this header specifies the name of the media file. Extensions should be omitted. If the media file is abe88.wav, then just enter “abe88”. Then declare the format as “sound” or “video”. It is also possible to add the terms “missing” or “unlinked” after the media type. So the line has this shape: @Media: abe88, sound, missing

@End Like the **@Begin** header, this header uses no colon and takes no entry. It is placed at the end of the file as the very last line. Adding this header provides a safeguard against the danger of undetected file truncation during copying.

Participant-Specific Headers

The third set of headers provides information specific to each participant. Most of the participant-specific information is in the **@ID** tier. That information can be entered by using the ID headers option in CLAN's Tiers menu. The exceptions are for these tiers:

@Birth of XXX: 01-JUL-1965

@Birthplace of XXX: Xxxx

@L1 of XXX: Xxxx

Constant Headers

Currently, the constant headers follow the participant-specific headers. However, once the participant-specific headers have been merged into the **@ID** fields, the constant headers will follow the **@Media** field. These headers, which are all optional, describe various general facts about the file.

Header	Function
@Exceptions:	This allows for special word forms in certain corpora.
@Interaction Type:	The possible entries here include: constructed computer phonecall telechat meeting work medical classroom tutorial private family sports religious legal face-to-face
@Location:	This header should include the city, state or province, and country in which the interaction took place. Here is an example of a completed header line: @Location: Boston, MA, USA
@Number:	The possible entries here include: two three four five more audience
@Recording Quality:	Possible entries here are: poor, fair, good, and excellent.
@Room Layout:	This header outlines room configuration and positioning of furniture. This is especially useful for experimental settings. The entry should be a description of the room and its contents. Here is an example of the completed header line: @Room Layout: Kitchen; Table in center of room with window on west wall, door to outside on north wall
@Tape Location:	This header indicates the specific tape ID, side and footage. This is very important for identifying the tape from which the transcription was made. The entry for this header should include the tape ID, side and footage. Here is an example of this header: @Tape Location: tape74, side a, 104
@Time Duration:	It is often necessary to indicate the time at which the audiotaping began and the amount of time that passed during the course of the taping, as in the following header: @Time Duration: 12:30-13:30 This header provides the absolute time during which the taping occurred.
@Time Start:	If you are tracking elapsed time on the %tim tier, the @Time Start header can be

used to indicate the absolute time at which the timing marks begin. If a new @Time Start header is placed in the middle of the transcript, this “restarts” the clock.

@Time Start: 12:30

@Transcriber: * This line identifies the people who transcribed and coded the file. Having this indicated is often helpful later, when questions arise. It also provides a way of acknowledging the people who have taken the time to make the data available for further study.

@Transcription: The possible entries here are: eye_dialect partial full detailed coarse checked

@Warning: This header is used to warn the user about certain defects or peculiarities in the collection and transcription of the data in the file. Some typical warnings are as follows:

1. These data are not useful for the analysis of overlaps, because overlapping was not accurately transcribed.
2. These data contain no information regarding the context. Therefore they will be inappropriate for many types of analysis.
3. Retracings and hesitation phenomena have not been accurately transcribed in these data.
4. These data have been transcribed, but the transcription has not yet been double checked.
5. This file has not yet passed successfully through CHECK.

Changeable Headers

Changeable headers can occur either at the beginning of the file along with the constant headers or else in the body of the file. Changeable headers contain information that can change within the file.

Header

Function

@Activities: This header describes the activities involved in the situation.

@Bck: It is sometimes impossible to decide whether background material occurs before, during, or after the utterance. In order to avoid having to make these decisions after the fact, one can simply enter it in an @Bck header.
@Bck: Rachel was fussing and pointing toward the cabinet where the cookies are stored.
*RAC: cookie [/] cookie.

@Bg and @Bg: These headers are used to mark the beginning of a “gem” for analysis by GEM. If there is a colon, you must follow the colon with a tab and then one or more code words.

@Blank This header is created by the TEXTIN program. It is used to represent the fact that some written text includes a blank line or new paragraph. It should not be used for transcripts of spoken language.

@Comment: This header can be used as an all-purpose comment line. Any type of comment can be entered on an @Comment line. When the comment refers to a particular utterance, use the %com line. When the comment refers to more general material, use the @Comment header. If the comment is intended to apply to the file as a whole, place the @Comment header along with the constant headers before the first utterance.
@Comment: Gestational age of MAR is 7 months
@Comment: Birthweight of MAR is 6 lbs. 4 oz

- @Date:**..... This header indicates the date of the interaction. The entry for this header is given in the form day-month-year. The date is abbreviated in the same way as in the **@Birth** header entry. Here is an example of a completed **@Date** header line:
@Date: 01-JUL-1965
- @Eg and @Bg:**..... These headers are used to mark the end of a “gem” for analysis by the GEM command. If there is a colon, you must follow the colon with a tab and then one or more code words. Each **@Eg** must have a matching **@Bg**.
- @G:**..... This header is used in conjunction with the GEM program, which is described in the CLAN manual. It marks the beginning of “gems” when no nesting or overlapping of gems occurs. Each gem is defined as material that begins with an **@g** marker and ends with the next **@g** marker.
- @New Episode**..... This header simply marks the fact that there has been a break in the recording and that a new episode has started. It is a “bare” header that is used without a colon, because it takes no entry. There is no need to mark the end of the episode because the **@New Episode** header indicates both the end of one episode and the beginning of another.
- @New Language:**..... This header is used to indicate the shift from the initially most frequent language listed in the **@Languages** header to a new most frequent language. This header should only be used when there is a marked break in a transcript from the use of one language to a fairly uniform use of another language.
- @Page:**..... This header is used to indicate the page from which some text is taken. It should not be used for spoken texts.
- @Situation:**..... This changeable header describes the general setting of the interaction. It applies to all the material that follows it until a new **@Situation** header appears. The entry for this header is a standard description of the situation. Try to use standard situations such as: “breakfast,” “outing,” “bath,” “working,” “visiting playmates,” “school,” or “getting ready to go out.” Here is an example of the completed header line:
@Situation: Tim and Bill are playing with toys in the hallway.
There should be enough situational information given to allow the user to reconstruct the situation as much as possible.