

Transcription standard for single-word utterance study. Version 1.6n, 7/24/00.

WARNING: This document supplements *The CHILDES Project: Tools for Analyzing Talk, Second edition*, by Brian MacWhinney. It does not substitute for reading the corresponding sections of that book.

The following is a list of differences between sequential versions of this document. If you have not seen any previous versions, please skip this section.

Changes from version 1.6 to version 1.6n

1. The letter “n” after the standard version number indicates that lines containing only the metalinguistic tags **babblers**, **nonspeech**, **vocalizes**, and **speechplay** are not being transcribed. All lines that contain actual speech are identical in both version 1.6 and version 1.6n.

Changes from version 1.5 to version 1.6

2. For diminutives, we will be using a hyphen, e.g. sweet-ie, chick-ie, NOT sweetie or chickie.
3. For pauses within a single utterance, we will use one # for each 100msec, e.g. uhoh ##### you bump-ed your head .

Changes from version 1.4 to version 1.5

1. @b, most instances of @f, @p all conflated and replaced by simply @., which indicates all nonsense syllables and idiosyncratic word forms. @l (for letters) and @o (for onomatopoeia used as words, but not for realistic imitations) are retained. Do not use @ for nicknames and variant nicknames, such as *Maggie+moo* and *Maggie+mooskie+moo*. Use @f only if there is a word that has a clear meaning and is used on a regular basis by the family, such as if they often say *poonie@f* as a joke variant of *spoon*. Use @f very sparingly, only when the word is frequent or important. If it only occurs only a few times, consider treating it like a nonsense syllable and using @. Do not use @f for nicknames.
2. For compound names, capitalize only the first letter of the first part. E.g., *Peter+rabbit* and *Mickey+mouse* and *Star+wars*, *Winnie+the+pooh*. Further clarification of when to use +.
3. For diminutives, we will use “ie” rather than “y”, as in *sweetie*, *doggie*, and *chickie*, not *sweety*, *doggy*, and *chicky*. However, for now we not hyphenate these as morphemes. Later we will go back and add hyphens.
4. A table of all currently accepted standardized spellings was added to notes on Chapter 5.
5. A new voluntary paralinguistic tag, **speechplay**, was added to cover all cases of adult cooing, babbling, imitation, and other attempts that to entertain or engage the child with spoken syllables that do not fall under any of the other paralinguistic tags or standardized

spellings. The transcriber has the option to transcribe these syllables and follow them with an @ special-form marker. For the child, a new tag, **vocalizes**, was also introduced. It includes **babbling** (which may still be transcribed as such) and all other forms of nonsense syllable vocalization.

Changes from version 1.3 to version 1.4

1. Emphasized that commas should not be used for pauses, only appositives and restrictive relatives.
2. Spelled out more detail on the use of + for compounds.
3. Emphasized that the stem to the left of the inflection-marker “-“ should be the normal spelling of the uninflected form and not the inflected form stripped of the inflection — e.g., *like-ing* and not *lik+ing*.

Changes from version 1.2 to version 1.3

1. Add @o for onomatopoeia used as a word (not for actual imitations of sounds).
2. Added + for compounding, including names like Mickey+mouse or Star+wars.
3. Allowed paralinguistic tags for special vocal characteristics to occur on the same line with tags for nonlinguistic vocalization. Use sparingly. Also clarified that all lines must end in a final punctuation, even if they only contain a paralinguistic tag for nonlinguistic vocalization.

Changes from version 1.1 to version 1.2

4. Paralinguistic Material tags occur only at the end of the line, without angle brackets, no matter whether they apply to the whole or only part of it.
5. All other scoped symbols now have the previous word as default scope.
6. New paralinguistic tags: **whistles**, and **laughed**.
7. Paralinguistic tags are now always one per line. Use a new turn record after any silent pause.
8. Speaker overlap is no longer marked explicitly. The time-alignment handles it.
9. For material that is not transcribed because it is not addressed to the child, use **%add not CHI** instead of **%exp**.

Changes from version 1.0 to version 1.1

10. Added notation for quoted material from section 6.7 of the book.
11. Added @p for consistent child forms whose meaning is unclear. (Use sparingly.)
12. Removed %gls tier for glossing child forms.
13. Removed [:= text] for translations.
14. Defined a specific set of permissible tags to put inside [=!] for nonspeech vocalization and speech with special vocal characteristics.
15. The default scope of scoped symbols is now the utterance.

3. File Headers

We will be using all and only the following headers. However, the current version of segmentation-editor may choke on some of them, especially @Coder and @Location. The best thing might be to use only @Begin, @Participants, @Warning, @Comment, and @End, until the next release of segmentation-editor. But do not consider the first pass transcription complete until all of these headers are in place.

@Begin

Header placed at the beginning of the file. Guarantees that no material has been lost at the beginning of the file. Takes no entry and uses no colon.

@Participants

Lists all the persons present during the recording. Persons are separated by a comma. Includes:

- 3 letter Speaker ID. Rather than names, use MOT for the mother and CHI for the child.
- Name of the Speaker (can be omitted). For more than one word use the underscore symbol to join them (ex. Maggie_Wynn);
- Role of the Speaker. The fixed roles are: Child, Mother, Father, Non_Human.(ex. Child for Maggie)

An example of a complete @Participants header would look like this:

```
@Participants: MAG Maggie_Wynn Child, EVA Eva_Wynn Mother
```

@End

Header placed at the end of the file. Guarantees that no material has been lost at the end of the file. Takes no entry and uses no colon. The only header that is not placed at the beginning of the file.

@Birth

Gives date of birth of the participant. The entry is day-month-year. Months should be uppercase and abbreviated as follows: JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC.

Example of an @Birth header:

```
@Birth of MAG: 11-MAR-1996
```

Note that we are not using @Age. @Birth combined with @Date provides this info.

@Warning

This header should contain the name and version number of the standard used for the transcription (SWU for the standard described in this document). The header should also contain

information on the status of the transcription. For the moment, the status of all files will be either 1st pass incomplete or 1st pass complete. If a file was transcribed earlier it may have no time records, incomplete time records, or inaccurate time records, and this should be noted. Later, we will have additional status marks to indicate various stages of checking and resolution of differences.

@Coder

Identifies people who transcribed the file and what their role was (i.e. initial transcriber, post-hoc time-aligner, etc.) Useful for organizational purposes as well as for answering questions that may arise later.

@Sex

Indicates gender of the speaker, male or female. Example:

@Sex of MAG: female

@Comment

This header is to be used as an all-purpose comment line. It would be useful to include the baby's age in days and the window number (corresponding to recording target dates) here.

@Date

The header indicates the date of the interaction. Entry given in Day-Month-Year, abbreviated the same way as the @Birth date. Example:

@Date: 1-JUN-96

@Location

This header should include the city, state, and country of where the interaction took place. This information can be obtained from the family cards that are located in the upstairs lab. DO NOT put the whole address of the participants in the file. Example:

@Location: Baltimore, MD, USA

4. Transcribing Words

4.1 Special form markers

@, @b, @p

Use @ with nothing following instead of @b, most case of @f, @p, and other ways of marking speech that does not consist of regular dictionary words or names, nicknames, or special words listed in the CHILDES book or the supplement below.

Use @ to indicate:

1. Genuine babbling,
2. Nonsense syllables used in speech play, songs, and rhymes,
3. Family-specific forms (can be derived from another language spoken in the family, or can be a made - up or modified word that has been adopted by the family.)
4. A child form that appears consistently but whose meaning you can't decipher.

Do not use @ for nicknames and variant nicknames, such as *Maggie+moo*, *Maggie+mooskie+moo*, and *Boboone*.

@f

Use @f only if there is a word that has a clear meaning and is used on a regular basis by the family, such as if they often say *poonie@f* as a joke variant of *spoon*. Use this very sparingly, only when the word is frequent or important. If it only occurs a few times, consider treating it like a nonsense syllable and using simply @. Do not use @f for nicknames.

@l

This marker is used to indicate letters, for example when something is being spelled out, or if the speaker is reciting an alphabet, etc. Example:

*MAG: my name is spelled m@l a@l g@l g@l i@l e@l.

@o

Use this after onomatopoeia when it is used as a word. Use [!= nonspeech] for actual imitations, such as making the noise of an animal or a train.

4.3 Codes for unidentified material

Unintelligible speech **xxx**

The code used to denote that it cannot be heard or understood what the speaker is saying. If you can distinguish the number of words, you can use the marker several times. Example:

*EVA: let's go xxx xxx milk.

Untranscribed material **www**

The code is used to indicate material that you do not know how to transcribe, or material that is irrelevant to the recording, for example when the mother talks on the phone. The www marker is used in conjunction with a %exp line, on which the problem itself is described. However, in the case of speech that is not transcribed because it is not addressed to the child, we are using the %add tier instead of %exp (see below).

*EVA: Your www is calling!

%exp: word in a different language, meaning unknown.

*EVA: www.
 %add: not CHI.

Best Guess at a Word: [?]

This code is used to indicate that the previous word or word group are your best guess, but that there is some doubt as to what the word(s) actually is/are.

*MAG: I want a banana[?] to eat.

Noncompletion of a word: () marker

the code is used when the word is not found in a list of abbreviations commonly used in speech, and if the meaning is clear, for example:

*MAG: gimme ba(na)na!

See pages 30, 31, 32, 33, and 34 in the CHILDES book for a list of standard spellings of marginal words such as Kinship forms (ex. Da(da)), Shortenings (ex. (a)bout), Assimilations (ex. gotta) Exclamations (ex. yumyum), Interactional markers (ex. huh), and Colloquial forms (ex. lookit!), Babytalk (ex. doggy). We will NOT be using Dialectical variants (ex. sumpin) on page 35, or subsequent tables on page 36.

NOTE: Don't go hog wild with parentheses or waste a lot of time deciding whether to use them. When in doubt, transcribe the word in the whole standard spelling.

4.5 Conventions for Standardized Spellings

In addition to the words listed in Section 4.5 of MacWhinney, the following words are officially approved.

Marker	Function
ahah	forbidden
alrightie	okay
blah	iconic nonsense speech
boo	startle or disapproval
brr	cold, shivering
ew	disgust
hiya	“hi” or “how are you”
icky	disgust
mkay	“okay”
neato	“neat”, interesting

Marker	Function
onna	assimilation of “on the”
patty	shortening of “patty+cake”
poopie	diminutive of “poop”
tada	ecstatic display
wazat	assimilation of “what’s that”
whatcha	assimilation of “what do you”
whoopie	delight
yay	cheer, alternate “yea”
yuck	opposite of “yum”

oops	“whoops”, blunder
oopsie	“whoopsie”, blunder
oh	surprise, address, or acknowledgment
ooh	“wow”

heinie	buttocks
weewee	urinate, urine, penis
wazis	“what’s this”

IMPORTANT: The only approved words are those in *Webster’s Third New International Dictionary* (<http://www.m-w.com/dictionary.htm>), those in MacWhinney’s *The Childes Project*, and those listed above. If you absolutely must introduce a new spelling that is not found in any of those, please submit it for approval and addition to this list right away.

5 Transcribing Morphemes on the Main Line

5.1 Codes for Morphemicization

We will use the suffix marker only. In particular, we will not use the fusion marker, so irregular inflected forms like *sang* will be written in the standard spelling.

For diminutives, we will use “ie” rather than “y”, as in *sweetie*, *doggie*, and *chickie*, not *sweety*, *doggy*, and *chicky*. However, for now we not hyphenate these as morphemes. Later we will go back and add hyphens.

IMPORTANT: Note that the stem, the part before the suffix marker, is always spelled the same way as the uninflected word: *like+ing*, not *lik+ing*.

5.2 Standard Forms for Affixes and Clitics

We will use only the suffixes listed in the table titled *Inflectional Suffixes for English*. We distinguish all suffixes in that table except that will use *-ed* for both past tense and past participle, so *-en* will not be used.

Compounding +

Use for compound names like Peter+rabbit and Mickey+mouse and Star+wars, to make clear that these are not kinds of rabbit, mouse, or war, but rather individual entities that happen to have words as part of their names. For compound names, capitalize only the first letter of the first part. Also for made up compounds, such as when the mother refers to an object as a pretty+pretty in part of a sentence, like

*MOT: Find the pretty+pretty !

Do not use this for rapid repetitions of a word when there is no syntactic evidence that they are creating a single compound word. For example:

*MOT: Piggy piggy piggy piggy piggy .

As another example, use + in, “That’s a no+no”, but not the utterance “No no ” where there is no evidence that the two repetitions are acting as a single word. Similarly, if the entire utterance consists of “up up” then there is no evidence that it is a single word so a + should not be used.

Do not use + in nonwords like uh+oh or um+hmm, since these have standard spellings. Similarly, do not use a + for words that are listed in the dictionary as a single word, like *goodbye*.

Do not use + in nicknames unless they contain the original name within them, and there is a clear sense that the additional syllable or syllables have an independent role. For example, do not use + for the nicknames *Boboon* and *Boneen* and *Bobine*, but one could use + for *Maggie*, *Maggie+moo*, and *Maggie+mooskie+moo*.

Our spell-checker accepts compounds connected by plus only if (a) it would accept all the component words individually and it would not accepted the component words concatenated together without a plus, or (b) the compound is on one of the lists of explicitly permitted words.

6.3 Basic Utterance Terminators

The idea is that a single line, or “turn record”, should contain a whole syntactic unit. That is, if the speaker says a sentence, noun phrase, or other coherent unit, it should all go on one line. Very rarely, there may be a silent pause in the middle of a sentence, for example if the speaker stops to think in a sentence like “I don’t...know.” In that case, the sentence should still go on one line, but the pause should be marked with a #, as described below.

If a speech sequence is not a sentence, for example it is a list like “Red, blue, green...” but it is spoken with the intonation of a single utterance (you don’t have to be too technical about this — you will know) and there are no noticeable silent pauses between the words, then it should still be put on one line. If there are silent pauses of at least 100msec between words and the words do not form a syntactic unit like a sentence or noun phrase, they should be put on separate lines.

NOTE: There must be exactly one space between the utterance terminator and the previous character.

Exclamation !

Use the exclamation point only for sentences that start with a verb. In other words, we will use them for syntactic commands, not for intonation. If a command is preceded by an address, such

as “Johnny, sit down” then it should still have an exclamation point. For quoted material, as in “I said, `sit down””, see section 6.7

6.5 The comma

Use commas for syntactic juncture only. E.g., appositives, restrictive clauses. Do not use commas for pauses: see #. Do not use double commas for tag questions.

6.6 Pauses

Use # to mark silent pauses of 100msec or longer with a single syntactic unit on a single line. Use a single # for 100-200msec, ## for 200-300msec, ### for 300 msec or longer. These should be rare; if you are using them a lot, you may be putting things on one line when they should be on two. In that case, consult with someone else.

6.7 Special Utterance Terminators

Quotation on next line +”/.

Indicates that one or more of the following lines is a direct quotation. From the book, it looks like this should be used when the direct quotation is part of the same sentence as the line ending with +”/.

Quotation precedes +”.

Same as #”/ when except the quote precedes and is part of the same sentence. Normally, I think only one of the two should be used for any given quote.

6.8 Utterance linkers

Quoted sentence +”

Indicates that the current line is a direct quote, regardless of whether it is part of some larger sentence or not.

8 Scoped Symbols

The default scope of scoped symbols is the previous word, so you do not have to include angle brackets when the scoped symbol applies to exactly the previous word. However, we are departing from this standard for the tags described under **Paralinguistic Material**.

Paralinguistic Material [=! text]

We use this extensively, especially to denote

1. Nonlinguistic vocalizations, including crying, humming, whistling. If there is no speech, do not include a “0”, simply create a line whose only content is, e.g., [=! hums]. There are two subcases:

Voluntary. Approved text tags are: **cries**, **babbles**, **hums**, **whistles**, and for all kissing, lip-smacking, or other labial noises, **labial**. For any voluntary nonspeech vocalization not covered under these tags, use **nonspeech**.

Prior to version 1.5, adult speech play was sometimes transcribed with “babbles” and sometimes with “nonspeech”. As of 1.5, all adult speech play should be either marked with its own tag, **speechplay**, or transcribed and follow by an @ (see Section 4.1). Speech play includes any utterance of syllables that is purely in imitation of a child or to get the child’s attention or entertain him or her, and is not covered by any standardized spelling or by any of the other tags listed above. For example, cooing.

When the child vocalizes but does not produce intelligible words, the options are **babbles** (intended for genuine babbling in repeated syllables), attempting to spell out the utterance followed by an @, or a new, more general paralinguistic tag, **vocalizes**, intended to cover all cases where the child vocalizes without producing any comprehensible words. When the child is clearly attempting to use words but the words are incomprehensible due to noise, bad recording, or bad pronunciation, xxx can be used instead of the word (not as a paralinguistic tag).

Involuntary. Approved tags are: **laughs**, **coughs**, **sneezes**, and **burps**. For any other involuntary nonspeech noise made by the human vocal tract, use **involuntary**. This may include various breathing noises.

These tags should always occur on a line that contains no transcribed speech. However, the line can include markers for special vocal characteristics, such as singing, and the line must end in a period.

2. Speech with special vocal characteristics. Approved tags are **whispered**, **sung**, **laughed** and **read**. Use **read** even if the reading is in a fairly normal tone of voice, since we need to know and there is no other obvious way to mark reading. Note that “laughed” is a manner of speech while “laughs” records a nonspeech event. For other special vocal characteristics, including gruff voice, use **voice**. These tags should always occur at the end of a line a line that also includes some transcribed speech or a nonlinguistic vocalization tag.

If any of the above occurs during an speech but continues beyond the end of the speech (e.g., laughing) time-align the entire vocalization, including speech and nonspeech, as a single

%gls I changed my mind. We are NOT using this.

Additional notes

Transcribing the child

Transcribe the child whenever she says an intelligible word. It is not necessary to transcribe crying and unintelligible material for this standard, though there is no harm in doing it.

Accuracy of time alignment

Try to align to 20msec accuracy, the accuracy of the *adjust small* buttons. However, we may revise this later, if a lot of time is being spent on high accuracy time alignment.