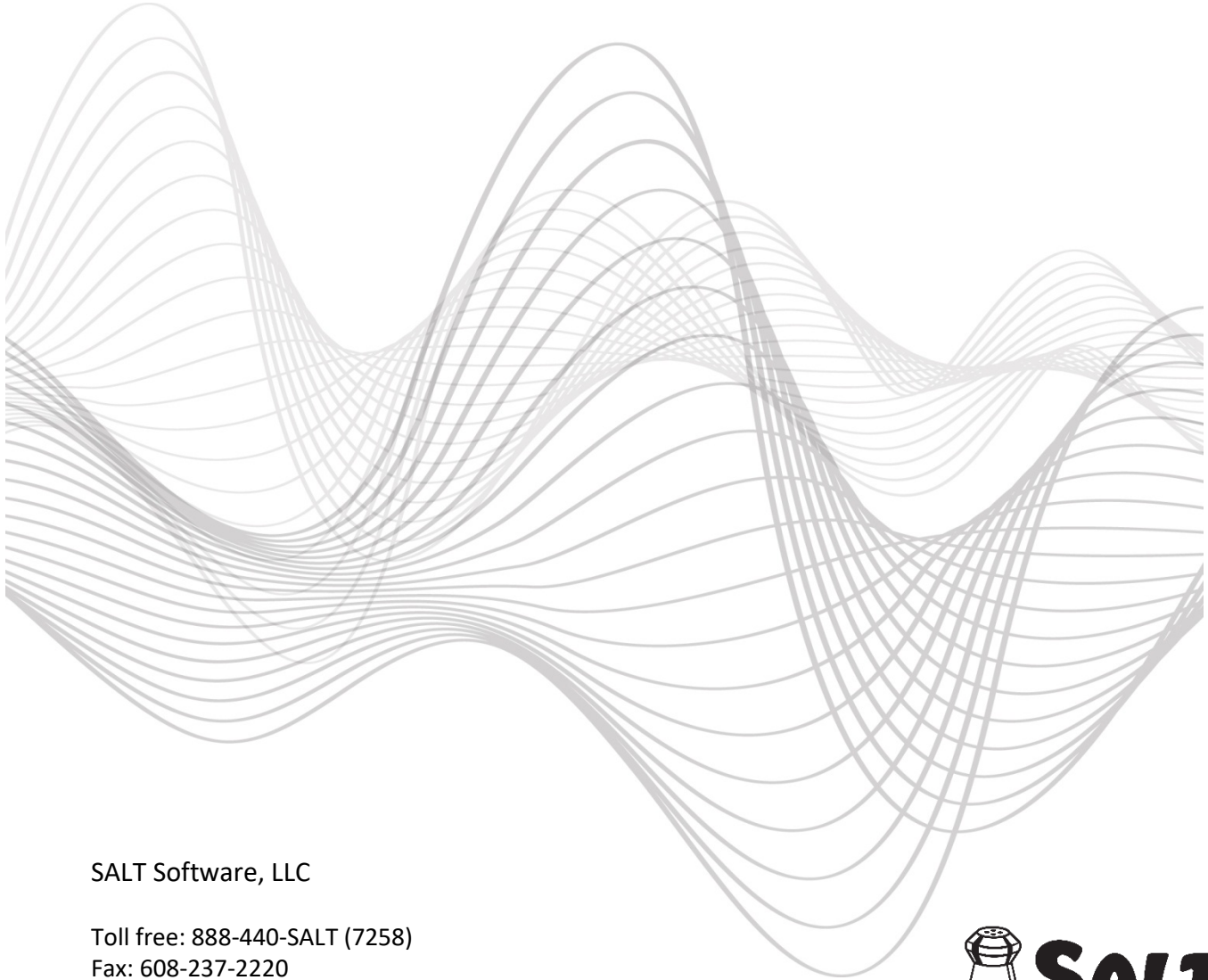# TRANSCRIPTION CONVENTIONS
Standard SALT Conventions

SALT Software, LLC

Toll free: 888-440-SALT (7258)
Fax: 608-237-2220
support@saltsoftware.com
www.saltsoftware.com

Table of Contents

Standard SALT Conventions

1. Resources

There is a lot of flexibility in the format of SALT transcripts. If you plan to compare your language samples to samples selected from the SALT reference databases, however, it is important that you use the same transcription conventions as were used with these database samples. This user guide is provided as a reference for English transcription. The SALT web site, however, is the best place to learn the SALT transcription conventions. There are courses designed to help you learn the conventions with clear explanations, lots of examples, and opportunities for you to apply what you learn. Select **Help → Web Training**. This takes you to training section on the SALT web site. The transcription conventions are also available in the help system built into the software. When you are in the SALT editor, just select the context help (press the function key F1) for a list of the transcription conventions. Another resource is the **Help → Resources → Transcription Conventions → Summary of Transcription Conventions** documentation which contains a summary of the conventions and Appendix A which gives the rationale behind which bound morphemes are marked.

What skills are needed to transcribe the samples?

1. Transcribers do not need to be experienced clinicians.
2. Transcribers should be familiar with the elicitation protocol.
3. Transcribers should be fluent speakers in the language of the sample in order to pick up subtleties in the recordings that non-fluent speakers may miss.
4. Transcribers should be familiar with the SALT transcript-entry conventions.

Note: You may find it expedient to divide the transcription task into two components: 1) a transcriber, fluent in the language of the sample, glosses the speech verbatim; and 2) a transcriber, familiar with the SALT transcript-entry conventions, edits the glossed text into a SALT transcript.

What materials are needed to transcribe the samples?

1. A transcription device for controlled audio playback is optimal. There are a growing number of options available for controlling digitized sound files. Visit saltsoftware.com/resources/linkstoequipment for suggestions.

2. The SALT editor for typing in the language sample and checking it for transcription errors.

2. Begin a New Transcript

To begin a new transcript, select **File → New** and you are presented with a header dialog box like the one shown here. You enter the appropriate information and it is used to format the speaker line, the information lines, and the initial timing line at the beginning of the transcript. You must identify the speakers in the transcript but the rest of the information is optional. *Refer to the Help Button in this dialogue box for detailed descriptions.*

*What parts of the language sample should be transcribed?*

Often the beginning of the language sample contains instructions from the examiner or a short conversation with the target speaker designed to put the speaker at ease. Unless this is part of the elicitation protocol, do not transcribe this text. Begin transcription at the point where the target speaker begins the elicitation task and then transcribe everything which was said. This includes all mazes (filled pauses, repetitions or reformulations), requests for vocabulary, and examiner prompts.

3. Transcript Format

a. *Identify the Speakers*

- General Information

  The speakers in the transcript are defined on the $ speaker line at the beginning of the transcript. The first character of each name must be a letter or number and is used to identify that speaker. This is called the "speaker id". Each speaker id must be unique since it is used to differentiate between the speakers. The speaker names are used to label the analyses. If the speaker name exceeds nine characters, only the first nine characters are used for labeling. In the example shown here, "C" is the speaker id for the child and "E" is the speaker id for the examiner.
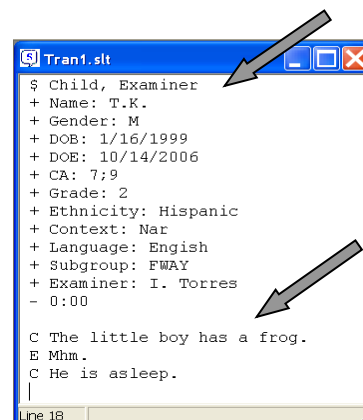
  There can be any number of speakers in the transcript but only two speaker groups can be analyzed at one time. Commas are used to separate the speaker groups.

- Transcripts with Two Speakers

  Place a comma between the two names. You may leave a space between the comma and the second speaker's name. The utterances of the first and second speaker are differentiated by a unique letter entered at the beginning of each utterance which corresponds to the first letter of each entry on the $ label line.

  Example of two speaker transcript (analyses will be labeled as BILLY and JANE):
  $ Billy, Jane
  J Kittycat.
  B I see you.
  B You big kitty.

  The first name entered on the $ label line is referred to as "1st Speaker" and the second name is the "2nd Speaker". In the above example, Billy is the 1st Speaker and Jane is the 2nd Speaker. Note that the first speaker does not refer to the first person to speak in the transcript. It is usually assigned to the speaker who is the primary focus of the transcript.

- Transcripts with One Speaker

  If you are entering the utterances of a single speaker, enter a $ followed by the single speaker's name, before the first utterance of your transcript. If you just enter child utterances, for example,

the SALT analyses will provide all data for the first speaker and print out "zeros" for the absent speaker.

Example of single speaker line:
> $ Billy

- Transcripts with More Than Two Speakers

  When entering the utterances of more than two speakers, you need to decide how you would like to have the SALT program group them for analysis because only produces analyses for the "1st" and "2nd" speaker groups during any given analyses run. The $ label line may contain from one to three speaker groups, but each group may consist of an unlimited number of speakers occurring within the transcript. Each and every speaker must be represented in one and only one group in the transcript. Utterances comprising the 1st speaker group are summarized under the "1st Speaker" in the analyses; utterances comprising the 2nd speaker group are summarized under the "2nd Speaker in the analyses; and utterances comprising the 3rd speaker group are not included in the analyses. (To obtain an analysis of the "3rd" speaker, you would simply reposition the groups on the $ label line and run an additional set of analyses.) Within a speaker group, each speaker group identifier (the first alphabetic character) must be separated by a space. The first alphabet letter of each speaker group, therefore, must be unique as it is used to differentiate that speaker (or group of speakers) from all others. Speaker groups are then separated with commas. Following are examples of various ways of specifying the multiple speaker groups. The participants in this conversation are defined as follows:
  > Achild = first child
  > Bchild = second child
  > Mother = mother
  > Father = father
  > Examr = examiner

  Examples of possible combinations:
  > $ Achild Bchild, Mother Father, Examr
  > In this first example, the 1st speaker is the two children (Achild and Bchild) combined; 2nd speaker is the Mother and Father combined; and the 3rd speaker is the Examiner. The utterances of the Examiner will be ignored in the analyses with the exception of the turn taking counts.

  > $ Achild, Bchild, Mother Father Examr
  > With this speaker line, the 1st speaker is Achild and the 2nd speaker is Bchild. The utterances of the other three speakers (Mother, Father, and Examiner) will be ignored in the analyses with the exception of speaker turn counts..

  > $ Achild Bchild, Mother Father Examr
  > In this last example, the first speaker is Achild and Bchild combined; and the second speaker is Mother, Father, and Examiner combined.

b. *Identification Information*

  Follow the $ speaker line with the transcript identification lines. These lines begin with a plus sign (+) and are used to provide identification information at the beginning of your transcript. This information may include the target language, the speaker's name and age, participant number, date of transcript, description of codes, etc. Enter such information on lines that begin with a plus sign (+) in the first

column. You may enter as many information (+) lines as you like at any point in your transcript. The SALT program ignores these lines for analysis. They are included for identification or commenting purposes only. However, the Database menu looks at the contents of the plus lines to extract specific information for automatically matching database records. These lines are all optional, but you are encouraged to include them for purposes of accuracy and convenience.

c. *Utterance Format*

Begin each utterance with the speaker id, followed by a blank space, followed by the exact words of that speaker. Each utterance is entered on a separate line and ended with an ending punctuation mark such as a period or question mark.

d. *Long Utterances*

Notice that when any speaker has two or more adjacent utterances each utterance is entered on a separate line. Long utterances that require more than one line will be automatically wrapped by the editor. Do not press the <Enter> key until the end of the utterance.

e. *Legal Word Characters*

SALT does not have a dictionary to define words. Instead a word is considered to be any combination of legal word characters. In general, words consist of letters, numbers, hyphen, and a special set of characters, such as the *asterisk* to mark omissions and the *slash* to mark bound morphemes, which have specific meaning in SALT. Words, which may be entered in either upper or lower case characters, are separated by blank spaces or commas. Words may be "quoted". Other characters, such as the *period* used to mark the end of an utterance or the *colon* used to mark pauses, are not legal word characters.

4. Ending Punctuation & Other Punctuation Marks

SALT uses the ending punctuation mark to determine the type of utterance. Every utterance must end with one of these end-of-utterance punctuation marks. They include a period, exclamation mark, question mark, tilde, greater-than sign, and caret. No other characters should follow the ending punctuation, not even quotation marks or special codes.

a. Ending Punctuation

- Statements end with a period or an exclamation mark.

- Questions end with a question mark. Use a question mark for interrogatives and for tag questions. Tag questions are brief questions which are attached to statements, and that seek agreement. For example:
    E What happen/ed next**?**
    C The animal bit him, right**?**

  Use a question mark even if the question is in the form of a quote. For example:
    C He said, "Is that you, frog"?

- Intonation prompts end with a tilde. An intonation prompt is an open-ended utterance used to prompt another speaker for a response using rising intonation. It is not presented in the form of a question but rather as a "fill in the blank". For example:

E And then the boy~

Even though these prompts are intended to elicit more elaborate language, they may result in shorter, non-spontaneous responses because they also take much of the burden of sentence construction off the other speaker. Because of this, it is important to distinguish the utterances which are the result of intonation prompts.

- Abandoned utterances end with a greater-than sign. If a speaker voluntarily stops in mid-utterance, end that utterance fragment with a greater-than sign and do not use a period. Consider this example where the speaker abandons the first utterance:
  C And then the boy look/ed in the**>**
  C The animal bit him.

- Interrupted utterances end with a caret. If a speaker is interrupted before completing an utterance, end the utterance fragment with a caret and no period, as in this example where the speaker interrupts the examiner's utterance.
  E What did**^**
  C He took the frog home.

b. *Other Punctuation Marks*

Commas and double quotes may be used freely to punctuate an utterance. But make sure that the last character of the utterance is the ending punctuation mark. Notice in the following examples that the ending punctuation mark, the period, is placed after the final quote.
  C Then the boy said, "I hear something".

All other punctuation marks have special meaning and should not be used when entering an utterance. Do not use a period for abbreviations because the period is reserved for use as ending punctuation. Either spell out the word or just leave the period off as in the following example:
  C And he saw the Mrs frog.

5. Utterance Segmentation

The analysis of oral language samples requires recorded speech to be segmented or divided into utterances. Utterance segmentation is relatively easy when the speaker produces short, simple utterances preceded and followed by utterances of another speaker. Utterance segmentation becomes more complicated when the speaker produces multiple utterances per speaking turn or when the speaker produces complex utterances.

Although you can define your own rules for segmenting utterances, the SALT reference databases follow specific rules when speakers produce successive utterances. The SALT reference databases are segmented into **communication units (C-units)**. The only exceptions are the *Bilingual Spanish/English Story Retell* and *Bilingual Spanish/English Unique Story* databases which are segmented into **modified communication units (MC-units)**. The following sections describe the rules for segmenting utterances into C-units (Section 5-a) and MC-units (Section 5-b).

a. *Communication Units (C-units)*

Disclaimer: There is variation in the literature on how to segment utterances into C-units. If you intend to compare your sample with samples selected from the SALT reference databases, you should segment utterances following the same rules.

The rules for C-unit segmentation were summarized from Hughes, D., McGillivray, L., & Schmidek, M. (1997), Loban, W., (1976), Strong, C.J., (1998), and from Jon Miller's class notes from Fall, 1999.

C -units are defined as "an independent clause and its modifiers" (Loban, 1976). A clause, whether it is the main clause or a subordinate clause, is a statement containing both a subject and a predicate. Grammatically, a subject is a noun phrase and a predicate is a verb phrase. Main clauses can stand by themselves and can be segmented into one C-unit. Subordinate clauses DEPEND on the main clause to make sense. They cannot stand alone or be separated from the main clause. So a C-unit will either consist of a main clause or a main clause with its subordinating clause(s). The following examples are broken down into main and subordinate clauses. The **main clause** is bolded and the subordinate clauses are underlined.

The canary was perched on a branch when the man approached him.
Anastasia was angry with her mother because she didn't get to buy a toy.
When the boy looked in the jar **he saw** that the frog was missing.

Notice the subordinate clauses cannot stand alone or are incomplete without the main clause. Thus, they are not separated (segmented further) from the main clause. Each of the above utterances consists of one C-unit and would be transcribed as:

    C The canary was perch/ed on a branch when the man approach/ed him.
    C Anastasia was angry with her mother because she did/n't get to buy a toy.
    C When the boy look/ed in the jar, he saw that the frog was missing.

Coordinating and Subordinating Conjunctions

When segmenting into C-units it is important to understand the different types of conjunctions which are used to link clauses. There are coordinating conjunctions and subordinating conjunctions.

- Coordinating Conjunctions

  The segmenting rule is simple when utterances contain coordinating conjunctions. These conjunctions link two **main** clauses which should be separated/segmented into two utterances (or two C-units) that can each stand alone. Common coordinating conjunctions include: and, but, so (but not "so that").

  Example 1:
      C The frog was sit/ing on a lily pad.
      C And then it jump/ed in.
  Example 2:
      C He had to catch the frog.
      C Or the waiter would make them leave.
  Example 3:
      C He climb/ed up on the branch/s.
      C But they were/n't branch/s.
  Example 4:
      C My aunt gave me money for my birthday.
      C So I use/ed it to buy some new jeans.

- Subordinating Conjunctions

  Subordinating conjunctions **link** a main clause and a subordinate clause. A C-unit includes the main clause with all subordinate clauses attached to it. The following are subordinating conjunctions:
  Early Development: because, that, when, who
  Later Development: after, before, so (that), which, although, if, unless, while, as, how, until, as__as, like, where

  Examples:
  C He went to the store because he was out of milk.
  C When the boy saw it, the frog jump/ed.
  C The man, who usually come/3s to my exercise class, was/n't there today.
  C We can/'t find my cat who always run/3s away.

- "because" and "so"

  Always consider "because" as a subordinating conjunction (it will never begin an utterance) unless it is preceded by the utterance of another speaker as in this example:
  C I like/ed the movie alot.
  E Why did you like it?
  C Because it was really funny.

  The word "so" can be either a coordinating conjunction or a subordinating conjunction. If its usage means "so that", it is a subordinating conjunction. Otherwise it is a coordinating conjunction.

  Example 1 ("so" used as a coordinating conjunction):
  C He had to go home.
  C So we could/n't go to the game.

  Example 2 ("so" used as a subordinating conjunction):
  C He had to go home so his mom could take him to the dentist.

Other rules for segmenting C-units

- Sentence fragments

  Sentence fragments are counted as separate C-units when the final intonation contour of the utterance indicates that a complete thought has been spoken. For example:
  C The boy, the dog, and the frog, they were friend/s.
  versus
  C The boy, the dog, and the frog. { fragment based on intonation }
  C They were friend/s.

- Elliptical responses

  Elliptical responses (sentence fragments) to questions or prompts from the examiner are counted as separate C-units. For example:
  E What did you do next?
  C Shop/ed.

- Yes/No responses

  If a question or intonation prompt is posed, segment the yes/no response from the subsequent utterance when succeeded by a complete utterance/c-unit. Examples:

  > E Is that the Spanish teacher?
  > C No.
  > C That/'s my Science teacher.

  > E Do you want to read your book now?
  > C No.
  > C I don't.

  > E Do you have any pet/s?
  > C Yeah.
  > C I have a dog.

  If a question or intonation prompt is posed, do not segment the yes/no response to stand alone when followed by an incomplete utterance/c unit. For example:

  > E Do you have any pet/s?
  > C Yeah, a dog.

  If an utterance begins with an affirmation or starter, and does not follow a question or intonation prompt, do not segment the affirmation/starter from the subsequent words. Examples:

  > E I like dog/s.
  > C Yeah, I do too.

  > E That sound/3s interesting.
  > C Yeah it was.
  > C It was really fun.
  > C Yeah we had such a great time.

- Tags

  Do not segment phrases such as "you know", "I guess", and "I mean" when they are used as tags. For example:

  > C He/'s gonna live with his dad, I guess.
  > C And then, you know, they were go/ing to this town.

- Questions as Tags

  Do not segment questions when they are used as tags. For example:

  > C They got in trouble, right?
  > C He miss/ed the bus, did/n't he?

- Dialogue Quotes

  Dialogue quotes which are embedded in, or as part of, an utterance are counted as one C-unit as in this example:

  > C And the boy said, "That/'s my frog".

  Successive main clauses that occur in direct quotes are counted as separate C-units. For example:

C And he said, "I/'m ready".
C "I want to go to the store now".

- Grammatical errors

   Ignore grammatical errors when segmenting utterances. For example,
   C They is[EW:are] go/ing now.   { child said, "They is going now." }
   C We *are go/ing too.   { child said, "We going too." }

- Pauses and intonation

   Do not ignore pauses and intonation when segmenting utterances but, whenever reasonable, segment utterances based on grammar rules. When listening to speech, for example, there is sometimes a significant pause (with or without ending intonation) between a main clause and a subordinate clause. This inclines one to segment the utterance. With C-unit segmentation, however, the utterance would not be segmented as in this following example where the speaker paused for two seconds between the main clause and the subordinate clause:
   C I like/ed the movie alot :02 because it was really funny.

   In the following segment, however, you have to consider pause time and intonation:
   C I like/ed the movie alot.
   : 0:02
   E Mhm.
   C Because it was really funny.

   If there is a significant pause and ending intonation (falling for statements, rising for questions) between the speaker's first utterance and the examiner's "Mhm", segment the utterances as shown above. Otherwise, give the speaker credit for subordination and transcribe these "prompt sounds" as interjections as follows:
   C I like/ed the movie alot :02 < > because it was really funny.
   E <Mhm>.

b. *Modified Communication Units (MC-units) – used with bilingual Spanish/English samples*

   Because Spanish is a pronoun-drop language (Bedore, 1999, 2001; Rojas & Iglesias, 2006), the English and Spanish transcripts in the Bilingual Spanish/English Story Retell databases were segmented into MC-units which were developed specifically for these samples to account for these omitted pronouns and to provide consistency in the transcription. Utterances containing successions of verbs without subjects are segmented and a fragment [F] code is placed at the end of each utterance lacking a stated subject as a result of this segmentation.

   Examples:
   1) C The gopher look/ed out of the hole.
      C and bit the boy [F].

   2) C The frog jump/ed.
      C and land/ed in the water [F].
      C and hit his head on a rock [F].

   3) C He ran.
      C jump/ed [F].

    C then fell down [F].

In the next example, the speaker repeats the verb "ran" for emphasis (*once in the first utterance and twice in the second utterance*). The underscore character is used in the second utterance to avoid over-inflating values such as mean length of utterance (MLU) and number of total words (NTW). *See Section 16 on "Words Repeated for Emphasis" for details.* Because of this repetition, the utterances in this example are not segmented.

    1)   C He ran ran.
    2)   C He ran ran_ran.

Contrast the previous example with the next example where the speaker again repeats the verb "ran" for emphasis. But because the verbs are listed with the conjunction "and", the utterance is segmented. The underscore character is again used to link words repeated more than once. Note that words which are linked using the underscore character are not segmented.

    1)   C He ran.
          C and ran [F].

    2)   C He ran.
          C and ran_and_ran [F].

6. Transcriber Comments

Comments improve the readability of the transcript and clarify events that have taken place. Comments may be inserted anywhere in the transcript without affecting the analyses.

a.   *Comments within Utterances*

Comments which are inserted in utterances must be enclosed in braces and placed before the end-of-utterance punctuation mark. The following example shows a comment inserted at the end of an utterance at the point where the speaker laughs.

    C Then the animal bit him on the nose **{C laughs}**.

b.   *Comment Lines*

Comments, not related to specific utterances, may be included for clarification. Enter your comment on a line that begins with an equal sign. You may enter as many comment lines as you want at any point in the transcript, stating each new comment line with an equal sign. You are not required to include final punctuation marks at the end of comment lines. In this example, a comment is included on a line beginning with an equal sign to explain why there are so many unintelligible segments.

    = **child talks softly and there's a lot of background noise**
    C Then he X for XX.
    C Then XX.

7. Unintelligible Segments

Unintelligibility can be the result of many things including the speaker turning away from the microphone or speaking too softly, equipment failure, or background noise. It can also be caused by the speaker's phonological difficulties or due to an unfamiliar listener. As a general rule, if you cannot understand the speaker's utterance after listening to it three times, it should be considered partly or completely unintelligible. Use an X to mark an unintelligible word. Use XXX when the entire utterance is unintelligible.

And use XX to mark a segment which is anywhere between a single word and the entire utterance. Note that it is often difficult to distinguish whether the unintelligible segment consists of a single word, multiple words, or the entire utterance (*after all, it's unintelligible*).

a. *Unintelligible Words and Utterances*

Consider the following examples:
(1) C He X for the class.
(2) C He XX.
(3) C XXX.

The first example uses X to mark an unintelligible word. The second example uses XX to mark an unintelligible segment, more than one word but less than the entire utterance. The third example uses XXX to mark the entire utterance as unintelligible. SALT considers an utterance to be partly intelligible if the utterance contains both intelligible and unintelligible segments as in the first two examples. An utterance is unintelligible if it consists entirely of unintelligible words or syllables as in the third example. When analyzing your transcript, SALT gives you the option of removing partly intelligible and unintelligible utterances from calculations such as MLU or number of different words.

b. *Partially Intelligible Words*

Partially intelligible words are not recognized as unintelligible for SALT analyses purposes, but are treated the same as any other word. This is because SALT considers any word to be intelligible unless it consists entirely of Xs. For instance, SALT treats the word "Xing" the same as any other intelligible word. If you want this word to be considered unintelligible, you should transcribe it as "X".

8. Spelling Conventions

Consistency is very important for a reliable and representative sample. Since each word with even a slightly different spelling appears as a different word in the analysis, it is important to follow some standard spelling conventions to ensure consistency within and between transcripts. Be aware that the SALT editor's error check does not monitor spelling consistencies.

a. *Abbreviations.* Periods are used to mark the end of utterances and are not considered legal word characters. Abbreviated words should either be spelled out or left as an abbreviation but without the period, e.g., Mr, Mister, Mrs, Dr, Doctor

b. *Yes words*: OK, AHA, MHM, UHHUH, YEAH, YEP, YES

a. *No words*: NAH, NO, NOPE, AHAH, MHMH, UHUH

d. *Filled pause words*: AH, EH, ER, HM, HMM, UH, UM, MM, and any word with the code [FP].

e. *Hyphenated words.* Follow standard spelling conventions, e.g., mother-in-law, twenty-five

f.   *Numbers.* Enter numbers in written form or as digits, e.g., twenty-one, 21

g.   *Clock time.* Do not use colons when typing clock time because it will be interpreted as a pause. Type out the words connected with an underscore character, e.g., eight_thirty, nine_o'clock

h.   *Counting and "spelled" words.* Use the underscore character to connect the numbers or letters, e.g., 1_2_3_4_5, C_A_T

i.   *Proper Names and Titles.* When proper names or titles are used, you should give the speaker credit for just one word. In the following example, the title "Frog Where Are You" is transcribed as a single word using the underscore character (*not the hyphen*).

    C I have Mrs_Nelson for math class.
    C The book is Frog_Where_Are_You.

j.   *Routine Phrases.* Routine phrases are phrases which are learned and used as a single unit. In the following examples, the routine phrases are transcribed as single words using the underscore character.
    C Once_upon_a_time there was a boy and a dog.
    C The next morning he woke up to say, "Good_morning" to the frog.

To be consistent, make sure you use the underscore character, not the hyphen, when you link words together.

k.   *Shortened words.* There are two ways to transcribe shortened words such as "cuz" for "because" and "ya" for "you".

1.   Ignore the shortened word and type out the full word. So if the speaker says "cuz", you enter it as "because". If "ya" is used instead of "you", type "you".

2. Type the word which was spoken and then use the root identification convention to identify the full word using the vertical bar, e.g., cuz|because, ya|you.

No matter which format you use for these examples, the word roots are analyzed as "because" and "you". *Refer to Section 17 for information on using the vertical bar to identify words.*

l.   *Sound Effects and Idiosyncratic Forms*

•   Sound Effects "%" are non-word vocalizations which represent specific sounds such as those made by an animal, e.g., "meow", or an object, e.g., "vroom". Words like "shh" and "psst" are not coded as sound effects because they have word status, that is, they are common substitutes for specific words, i.e., "shh" for "be quiet" and "psst" for "hey" or "look at me".

When a sound effect is essential to the meaning or structure of the utterance, it should be entered as any other word except it should begin with a percent sign (%).There aren't recommended spellings for these sound-effect words. Just represent them as accurately as possible and try to be consistent. If the sound effect is not essential to the meaning or structure of the utterance do not enter it as a word; instead note its occurrence with a comment.

Consider the following examples:
(1) C The dog went %grr.
(2) C Then the boy heard %ribbit_ribbit.

(3) C The dog growl/ed {child makes growling sound} at them.
(4) C Then the boy heard the frog {ribbit ribbit}.
(5) C They fell down.
      = child makes falling down sounds

In the first two examples, the sound effects are essential to the meaning and structure of the utterance. Therefore, these sound effects are given word status and are preceded with a percent sign. Notice that the sound effect in the second example consists of two sounds, "ribbit ribbit". These sounds are linked together so they count as a single word to avoid overly inflating measures such as mean length of utterance (MLU) and number of total words (NTW).

In the last three examples, the sound effects are not essential to either the meaning or the structure of the utterance. They do, however, contribute to the quality of the language sample and should be included as comments. In examples (3) and (4), the comment is placed within braces as part of the utterance. In example (5), the comment is entered as a separate entry on a line beginning with an equal sign. This comment could also have been placed in braces at the end of the utterance. *Refer to Section 6 for additional information on marking transcriber comments.*

- Idiosyncratic Forms "%": In the process of mastering the phonological system, young children sometimes produce speech that differs from the adult version. Enter a percentage **(%)** symbol before an idiosyncratic form used by the child. Although idiosyncratic forms are not adult-like productions, they are stable productions by the child. They are consistent in reference to an object, person, or situation. It is not uncommon for these forms to remain in a child's vocabulary for many years.

    For example:
        %vroom   ("car")
        %coopa   ("cookie")

    If such forms are interpretable by family members or other familiar persons, you can be assured that they are idiosyncratic forms rather than word errors.

m. *Concatenatives*

BETCHA (*bet you*)	LIKETA (*like to*)	OUTTA (*out of*)	USETA (*used to*)
COULDA (*could have*)	LOOKIT (*look at it*)	SHOULDA (*should have*)	WANNA (*want to*)
GONNA (*going to*)	MUSTA (*must have*)	SPOSTA (*supposed to*)	WHATCHA (*what are you*
GOTTA (*got to*)	OUGHTA (*ought to*)	TRYNTA (*trying to*)	WOULDA (*would have*)
HAFTA (*have to*)			

n. Sounds with Specific Meanings
   HMM, HUH (question or affirmation)
   IDK (intones *I don't know*)
   UHOH (something in wrong)
   SHH or %SHH (be quiet)
   PSST or %PSST (to get someone's attention)

o. *Other Spellings*. AIN'T, A LOT, ATTA, NOONE, OH, OOH, OOP, OOPS, OOPSY

9. Bound Morphemes

If you want SALT to make calculations in morphemes as well as words, you must separate each bound morpheme from the free morpheme. Use a slash (/) for bound morphemes which follow the free morpheme (suffixes) and use a backslash (\) for bound morphemes which precede the free morpheme (prefixes). There should be no spaces between the bound morpheme(s) and the free morpheme.
Grammatical morphemes have been a useful index for tracking growth and complexity of young children's language. **Refer to Appendix A for the rationale behind which bound morphemes are marked.**

a. *Regular Plural Inflections /S*
Use the bound morpheme /S to indicate regular plural inflections, e.g., frog/s, tree/s. The /S is added to the noun stem without changing the spelling of the stem word, e.g., "babies" becomes "baby/s". The plural bound morpheme is not used for those words which do not have singular forms, e.g., "in the woods", "tore his pants", "congratulations".

Do not mark irregular plurals, e.g., geese, deer.

b. *Possessives /Z*
Use the bound morpheme /Z to mark possession, e.g., dad/z, Mary/z. Do not mark possessive pronouns, i.e., mine, his, hers, ours, yours, its, theirs.

c. *Plural Possessives /S/Z*
Use the bound morpheme /S/Z to indicate both plurality and possession. The plural inflection comes first followed by the possessive inflection, e.g., "the frog/s/z baby/s".

d. *Third Person Singular Verb Inflections /3S*
Use the bound morpheme /3S to indicate third person singular verb inflections, i.e., "he look/3s in the tree", "the frog jump/3s in the water". Do not mark irregular verbs (e.g., has, was) or when the sound of the root changes (e.g., do → does, say → says).

e. *Progressive verb form /ING*
Use the bound morpheme /ING for present progressive inflections that indicate ongoing action, e.g., call/ing, walk/ing. The /ING is added to the verb stem without changing the spelling of the stem word, e.g., swim/ing. Do not mark the gerund use of the verb, e.g., "reading is fun", "I like dancing".

f. *Regular Past Tense /ED*
Use the bound morpheme /ED to mark regular past tense inflections, e.g., look/ed, call/ed, like/ed. The /ED is added to the word stem with no change in the spelling of the stem, e.g., cry/ed, like/ed.
Do not mark irregular past tense verbs such as "went", "had" or "made". Do not mark predicate adjective, e.g., "was tired", "is bored". *Transcription hint*: regular past tense /ED verbs never follow a BE or GET verb.

g. *Past Participle /EN*
Use the bound morpheme /EN to mark past participle inflections, e.g., take/en, eat/en, prove/en. The /EN is added to the word stem with no change in the spelling of the stem, e.g., "prove/en.
Do not mark irregular forms (e.g., gotten, spoken, seen, been) or when the sound of the root changes (e.g., write → written). *Transcription hint*: regular form is present tense + EN as a separate syllable. /EN

verbs always follow HAVE, HAS, or HAD.

h. *Negative contractions /'T, /N'T*
    Use the slash to separate the verb stem from the contraction (e.g., can/'t, did/n't). However, do not use the slash to mark contractions where the <u>sound</u> of the root word is different in the contracted form (e.g., won't, don't). Thus, the word "didn't" is slashed because the sound of the root word is the same (did → did/n't). But the words "won't" and "don't" are not slashed because the sound of the root words changed (will → won't, do → don't). Do not slash the contraction "ain't" because this word does not exist as an uncontracted form.

i. *Contracted verbs*
    /'S, /'RE, /'M, /'LL, /'D, /'VE →   IS, ARE, AM, WILL, WOULD, HAVE
    Examples: I/'ll, I/'m, I/'d, we/'re, he/'s, we/'ve.

    /H'S, /H'D, /D'S, /D'D, /'US →   HAS, HAD, DOES, DID, US
    Examples: "HE/H'S been sick.", "THEY/H'D better go now.", "WHAT/D'S he do for a living?", "Why/D'D the boy look for the frog?", "LET/'US go".

k. *Prefixes* (**not marked in the samples stored in the SALT reference databases**)
    Use the backslash, "\", to mark prefixes, e.g., UN\HAPPY, DIS\LIKE. Note, however, that the prefix convention was added for prefix-based languages, such as Persian, but may be used to mark prefixes in any language.

Automating Bound Morpheme Identification

To help with identifying bound morphemes, a lookup file containing approximately 5,500 inflected words with their corresponding bound morpheme coding is available. This root identification file (RIF), <u>English Inflections.RIF</u>, is used to automatically identify inflected words. Because this is not a complete list, you should identify the bound morphemes as you are typing your transcript. This file may then be used to catch those you miss.

Select **Edit → Identify Roots** to look at each word in the transcript. If the word is not found in the active RIFs, that word is ignored. If only one choice is found, the word is automatically identified. If a word contains more than one root option, the user is presented with a list of choices to select from. Note: the active RIFs are selected using the *Setup menu → Language Settings* option. *See Lesson 2 for directions on using this utility.*

10. Mazes

Marking mazes provides an opportunity to document problems associated with utterance formulation and word finding that would otherwise go undetected. Mazes refer to repetitions, revisions, false starts and filled pauses, and are marked by enclosure in parentheses. When mazed words are removed from the utterance, the remaining words can stand by themselves. It is important to mark mazes so that they are not counted as part of the utterance. This excludes them from Mean Length of Utterance counts and other similar values.

When you have a choice of what words or phrases to mark, parenthesize the earliest occurrence as the maze. Consider the last occurrence of the word or phrase as the successful production. Incorrect maze analysis of repetitions and revisions will result if the first occurrence is not marked as the maze.

a. *Repeated Words and Phrases*

Mark any repeated word or phrase as a maze and place it in parentheses. Consider the last occurrence of the word or phrase as the successful production. Consider these examples:
1) C And (the) the boy was sad.
2) C He saw (the the) the frog.
3) C (The frog) the frog was gone.

In the first example, the speaker says, "And the, the boy was sad." The first instance of the word "the" is parenthesized. In the second example, the boy repeats the word "the" three times; the first two instances are enclosed within the same set of parentheses. In the last example, the phrase "the frog" is repeated; the first instance is parenthesized.

Repetitions for emphasis: if the speaker repeats one or more words for emphasis, do not mark the repeated words as mazes because mazes are used to indicate formulation problems. Consider the following example where the speaker repeats the word "everywhere" for emphasis, not as the result of a formulation problem.
1) C And they looked (everywhere) everywhere.   { *not correct* }
2) C And they looked everywhere, everywhere.   { *correct* }

Refer to Section 16 for more information on marking words repeated for emphasis.

b. *Revisions*

Mark any false start or reformulation as a maze and place it in parentheses. Remember that when maze words are removed from the utterance, the remaining words can stand by themselves. When you have a choice of words to parenthesize, select the earliest occurrence as the maze. Several examples follow:
C And then (the boy) the dog bark/ed at the tree.
C He saw (his own frog) his frog (and) with a lady frog.
C (He want/ed to) he took the baby home.

c. *Filled Pauses*

Words or vocalizations that fill in pauses should be placed in parentheses. The default list of filled pause words include AH, EH, ER, HM, UH, UM, and any other word coded as [FP]. Consider the following example where the speaker uses the filler "um" twice:
C It was a big (um um) bird.

In this next example, the speaker uses the word "like" in the middle of the utterance. Several transcription options are given.
1) C He was like really angry.   { *correct if "like" contributes to the content* }
2) C He was (like) really angry.   { *not correct: "like" isn't recognized as filled pause* }
3) C He was (like[FP]) really angry.   { *correct: like is recognized as a filled pause* }

In utterance (1), the word "like" is considered to contribute to the content of the utterance and is not mazed. This is a reasonable interpretation if the speaker does not use the word "like" in this manner very often. Perhaps it is used here to provide emphasis. Some speakers, however, use words such as "like" as fillers throughout their narrative. If this is the case, the word "like" should be mazed. In utterance (2), the word "like" is mazed incorrectly. Since "like" is not one of the default filled pause words, the SALT program would not recognize this maze as containing a filled pause. Instead it would categorize the maze as a revision (from "like" to "really"). In utterance (3), the word "like" is mazed and

is coded as [FP]. SALT interprets all mazed words with the code [FP] as filled pause words. The *Setup menu → Lists → Standard Word Lists* option may be used to change the default list of filled pause words. *Refer to Section 19 for more information on inserting bracketed codes at the end of words and utterances.*

d. *Adjacent Mazes*

If different types of mazes are adjacent, you should combine them in a single maze. This will allow consistency between your transcript and the reference database transcripts. In this example, a repetition, a filled pause and a revision are all included in one set of parentheses.
> C And so (it it um then he) then it bit him.

11. Part Words and Stuttering

Part words occur when a speaker fails to complete a word. Use an asterisk to replace the portion of the word which is missing. Part words are usually treated as maze components and are parenthesized.

a. *Part Word Revisions*

Consider this example:
> C He saw (hi*) them.

This example illustrates a part-word revision where the part word ends with an asterisk and is parenthesized.

b. *Part Word Repetitions*

In the following example, the speaker stuttered twice on the "b" in "boy". The stuttered parts of the word end with an asterisk and are parenthesized.
> C The (b* b*) boy woke up.

To mark stuttering in the middle of a word, separate the two parts of the word, before and after the stuttering, with underscore characters. Word linking is typically used to represent titles and proper names as single words. It is used here to join two or more segments of the same word. Consider these examples:
1) C The rab_ (b* b*) _bit got away.
2) C (The rab_ b* b* _bit um) the frog got away.
3) C Frog__ (wh*) _Where_Are_You.
   (*there are two underscore characters after the word "Frog"*)

In the first example, the speaker stuttered on the "b" in the middle of the word "rabbit". Notice that the stuttered sounds are asterisked and parenthesized. The second example illustrates how to mark stuttering in the middle of a mazed word. Since the entire word is parenthesized, do not parenthesize the stuttered sounds. When these split words are stored for analysis, the underscore characters are removed. Thus the word "rabbit" is stored without the underscore characters. The third example illustrates how to keep the underscore character for titles or names. You double up on one of the underscore characters so, in the third example, the title is stored as "Frog_Where_Are_You".

c. *Part Words at the End of an Incomplete Utterance*

Part words are usually treated as maze components and are parenthesized. There is, however, an exception to this rule. When the last word of an abandoned or interrupted utterance is a part word, use the asterisk to mark the part word but do not parenthesize it as it does not meet the definition of a maze. In the following example, the child abandons the first utterance mid-word.

C Then the f*>
= child turns to the next page
C The boy look/ed in the hole.

Alternate coding:

If you prefer to count a string of repeated words or part-words as one dysfluency, consider coding it as follows:

C The (b_b|b*) boy woke up.
C (The rab_ b_b|b* _bit um) the frog got away.
C And (the_the_the|the) the frog was gone.

The Maze Summary section of the Verbal Facility report identifies "(b* b*)" as two part-word repetitions and identifies "(b_b|b*)" as one part-word repetition. It identifies "(the the the)" as three word-level repetitions and identifies "(the_the_the|the)" as one word-level repetition.

Refer to Section 17 for more information on using the vertical bar to identify the root form of the text which precedes it.

---

12. Omissions

You should mark omissions because they may be an indication of utterance formulation problems. An omission occurs when a word, bound morpheme, or bound clitic which is obligatory for grammatical correctness is absent. Omissions are not included in calculations such as "mean length of utterance", "number of different words", or "words per minute". They are, however, counted in the "Word and Morpheme Summary" and listed in the "Word List Tables", the "Bound Morpheme Tables", and the "Clitic Tables".

a. *Omitted Words*

The asterisk symbol is used to indicate an omitted word. At the point in the transcript where the word was omitted type an asterisk followed by the omitted word. There should be no blank spaces between the asterisk and the omitted word. Consider the following examples in which the omitted word in each utterance begins with an asterisk:

C The boy ***is** call/ing for the frog.
C The little boy was look/ing ***for** him.
C The dog fell ***out** of the window.

b. *Omitted Bound Morphemes*

A slash followed by an asterisk is used to indicate the omission of a bound morpheme in obligatory context. Type the slash, the asterisk, and the missing bound morpheme at the point in the transcript

where it was omitted, as in the following examples.
    C The dog was follow**/*ing** him.
    C And then he see**/*3s** the little mommy and the baby/s.
    C The mom frog said the baby frog could go to the boy**/*z** house.

c. *Omitted Word or Omitted Morpheme?*

Omitted contractions may be transcribed in two ways, as an omission of a word or as an omission of a bound morpheme. Suppose the speaker says, "He not in the jar". This could be transcribed as either,
    C He ***is** not in the jar.
or,
    C He**/*s** not in the jar.

In the first transcription the omission is treated as an omitted word. In the second transcription, the same omission is treated as an omitted bound morpheme. It doesn't make much difference which way you mark these types of omissions since omissions are not included in any of the calculations based on words or morphemes. The important thing is to mark the occurrence of the omission.

d. *Omitted Part of Speech*

Consider the next example where it is not obvious what the omitted word should be:
    C In the night the frog ***VERB** out of the jar.

The verb has clearly been omitted but it is not obvious which verb. Some possibilities include "got", "climbed", "crawled" or "escaped". In this case, the omission is marked as "*VERB" to indicate that the verb was omitted. An article has been omitted from the following utterance.
    C They look/ed ***ARTICLE** the jar.

If there are a lot of omissions, it may be informative to mark these omissions with the part of speech. If you are not interested in the part of speech which was omitted, use something generic, like *WORD.

e. *Multiple Omissions and Errors*

> General Rule: Do not mark more than two omissions and/or word errors in an utterance. Instead, mark the entire utterance as having a problem by inserting the utterance code [EU] at the end of the utterance.

Suppose the speaker said "The dog fell the window". You could "fix" this utterance by adding the following two omissions:
    1) C The dog fell ***out *of** the window.

What if the speaker had said, "Dog fell the window"? This could be fixed as:
    2) C ***The** dog fell ***out *of** the window.

What if the speaker had said, "Dog window"? You could fix this because you are familiar with the story and know that, at one point in the story, the dog falls out of the window. But this is not recommended.

A common transcription mistake is to "fix" utterances so much that it becomes more of a guessing game than an accurate transcription. A good rule of thumb is to restrict your fixes to no more than two omissions. Following this rule, example 1) above would be allowed but example (2) above would not.

To mark a problem with an utterance containing more than two omissions, insert the utterance code [EU] at the end of the utterance, immediately preceding the ending punctuation as shown here.

2) C Dog fell the window **[EU]**.

Coded utterances can be counted and called up later for further analysis. *Refer to Section 19 for more information on [EU] and other error codes.*

13. Pauses

Information about pauses, together with other analytic information, may help to identify a number of problems such as word-finding difficulties or poor comprehension. Generally, pauses indicate a possible turn-change or topic-change. Most people notice moments of silence that last longer than two seconds. Marking pause time is optional and there are no minimum or maximum pause times. Pauses may be marked whether they occur in the middle of an utterance or between two utterances. When transcribing samples for the reference databases, all pauses two seconds or longer were marked.

a. *Pauses Within Utterances*

If you are interested in the frequency and duration of pauses occurring within utterances, mark this information by entering the length of pauses in ":seconds" format. The pause-time information should be entered at the point in the utterance where the pause occurs. Pause-time should be separated with a blank space from any adjacent words as in this example showing a three-second pause:

 C So he **:03** got up on the rock.

Untimed pauses: if you want to mark the occurrence of significant pauses but do not want to time them, you may enter a colon without time as in the following example:

 C And he held some (um **:** um) stick/s.

Pauses adjacent to mazes: if a pause occurs either immediately before or after a maze, include that pause with the maze as in the following example:

 C And then the cat (um **:03**) scratch/ed me.

b. *Pauses Between Utterances*

If you are interested in the frequency and duration of pauses which occur between utterances, code this information by entering the length of the pause in "minutes:seconds" or ":seconds" format on its own line. This "pause line" must start with either a colon or a semicolon. If you are interested in marking speaker's turns, you should understand the significance of beginning the pause line with a colon or with a semicolon. This is because the pause line can be used to force the end of one turn and the beginning of another. If the pause line occurs between utterances of two different speakers, then it doesn't matter whether you use the colon or semicolon - the first speaker's turn ends and the other speaker's turn begins. If the pause occurs between consecutive utterances of the same speaker, however, use a semicolon (;) if the speaker's turn does not end (*the more common case*) or use a colon (:) to force the end of the speaker's current turn and the beginning of that same speaker's next turn.

The following example shows a four-second pause between the utterances without a change in speaker turn.

 C The boy fall/3s down.
 ; :04
 C And he land/3s in the water.

New format introduced with SALT 18 – the pause from the previous example can be entered more

simply as:

    C The boy fall/3s down.

    ;04

    C And he land/3s in the water.

Where the pause time begins with a colon or semicolon.

Enter an empty pause line if you want to mark the occurrence of significant pauses but do not want to time them. As you see in the following example, empty pause lines need only begin with a colon or semicolon.

    C The boy fall/3s down.

    ;

    C And he land/3s in the water.

14. Overlapping Speech

When two or more speakers talk at the same time, their utterances are entered separately because transcription is linear. Use angle brackets <like these> to enclose the sections of each utterance which are spoken concurrently. In conversational samples, overlapping speech may be an indicator of discourse issues. In narrative samples, it documents the amount of within-utterance prompting that occurs (*see part d in this section*).

a.   *One Speaker Begins Before the Other Speaker Ends*

Notice in the following example that the beginning of the child's utterance overlaps the end of the examiner's question. Also notice that the examiner's utterance still needs ending punctuation (the question mark) and that the closing angle bracket is placed before the question mark.

    E Then what **<happen/ed>**?

    C **<The dog>** ran away.

Sometimes the first speaker will stop talking as soon as the other speaker begins. In the following example, the examiner's utterance ends with a caret indicating that it was interrupted.

    E Then what **<hap*>^**

    C **<The>** dog ran away.

b.   *Overlapping Speech in the Middle of an Utterance*

The overlapping speech may occur in the middle of an utterance. Just mark the overlapping segments and continue transcribing the utterance as in this example.

    C The boy and **<the dog>** look for the frog.

    E **<Good>**.

c.   *Overlapping Speech and Mazes*

The overlapping speech may occur within or around mazes as in the following example.

    C **<(The animal)>** the animal pick/ed him up.

    E **<Keep go/ing>**.

The order of the angle brackets and the parentheses is not important. The child's utterance could also have been transcribed as:

    C (<The animal>) the animal pick/ed him up.

The following example illustrates overlapping speech occurring partly within the maze and partly after the maze.

    C The (um the the **<um) the>** animal pick/ed him up.
    E **<Keep go/ing>**.

d. *Examiner Prompts in the Middle of the Target Speaker's Utterance*

The examiner may use words such as "uhhuh" or "mhm" as feedback to encourage the target speaker to continue with the story. If these words overlap the speaker's words, then mark them as you would any other overlap. However, if such feedback is given in the middle of the speaker's utterance but does not overlap the speaker's words, use empty brackets to indicate the location of the prompt. In this example, the examiner said "Mhm" in the middle of the speaker's utterance, between the words "boy" and "picked up".

    C And then the boy **< >** pick/ed up the little frog.
    E **<Mhm>**.

15. Parenthetical Remarks

A parenthetical remark is a word or clause, occurring within an utterance, which has been added by the speaker as an explanation, comment or question. When a parenthetical remark interferes with the rest of the utterance, enclose that part of the utterance in double parentheses. Consider the following examples:

    1) C The boy **((I don't know his name))** call/3s for the frog.

    2) C Then the **((what/'s that call/ed))** < > gopher bite/3s him on the nose.
        E <Gopher>.

In the first example, the speaker says, "The boy, I don't know his name, calls for the frog." In the second example, the speaker says "Then the, what's that called?". The examiner supplies the label "gopher" and the speaker continues, "gopher bites him on the nose". (*Notice, in the second example, that there is no question mark after the parenthetical.*) These parenthetical remarks, like mazes, interfere with the rest of the utterance. Parenthetical remarks, however, should not be confused with mazes since mazes, unlike parenthetical remarks, may be an indication of formulation difficulties.

Consider the next two examples where the parenthetical remark does not interfere with the rest of the utterance:

    1) C And then, **I don't know why**, he start/ed yell/ing at me.
    2) C And then, **I don't know why**, but he start/ed yell/ing at me.

In these examples, the parenthetical remark contributes to the content of the utterance and is not enclosed in double parentheses. In fact, the parenthetical remark in the second example is integral to the utterance; without it, the utterance is not grammatically correct.

Unless parenthetical remarks occur frequently within a language sample, they may not be of much analytical interest. Parenthetical remarks do have an impact, however, on language measures such as the "mean length of utterance" and "number of different words". The "main body" of the utterance contains the words used to calculate these measures. By default, the main body of the utterance excludes maze words and parenthetical remarks. Maze words are always excluded from the main body of the utterance. The *Setup menu* → *Word Base* option can be used to specify whether or not to include parenthetical remarks.

16. Words Repeated for Emphasis

Speakers sometimes use repetition to provide emphasis. To prevent these repetitions from overly inflating measures such as mean length of utterance (MLU) and number of total words (NTW), the underscore is used to link segments which are repeated more than once. The first use of each word in the repetition is transcribed normally. Words repeated more than once are linked together and treated as a single word.

a. Examples
    1) C The dog ran ran.
    2) C The dog ran ran_ran.
    3) C The dog ran very very_very fast.
    4) C They looked everywhere and everywhere_everywhere.
    5) C They looked everywhere and everywhere_and_everywhere.

In example 1, the repetition involves a word which is repeated once and is not linked. In examples 2 – 5, the words are repeated more than once and are linked together. Repeated words do not have to be contiguous. In example 4, the word "and" is not repeated but it occurs between the first instance of "everywhere" and the two repetitions.

b. Identifying Linked Words as Instances of the Word Being Repeated

In examples 2 – 5 above, the linked words result in new words, e.g., "ran_ran" and "very_very". To avoid overly inflating the number of different words (NDW), these words should be identified as instances of the word being repeated, e.g., "ran" and "very". To do this, follow each linked segment with a vertical bar "|" and the repeated word as shown in these examples:
    2) C The dog ran ran_ran|ran.
    3) C The dog ran very very_very|very fast.
    4) C They looked everywhere and everywhere_everywhere|everywhere.
    5) C They looked everywhere and everywhere_and_everywhere|everywhere.

In example 2, the words "ran" and "ran_ran" would be counted as two uses of the same word. Note in example 5 that the linked phrase "everywhere_and_everywhere" is identified as a form of the word "everywhere". This identification is not strictly accurate since the word "and" was also repeated but the vertical bar may only be used to identify one word.

Refer to Section 17 for more information on using the vertical bar to identify the root form of the text which precedes it.

17. Root Identification

Use the vertical bar when you want to identify a different word root than the one which was spoken. Be sure that the vertical bar and the root word directly follow the word used with no spaces between. Root identification instructs SALT to consider the word immediately preceding the "|" symbol as the word which was actually said, and the word immediately following to be the word root. This convention was originally created to identify Spanish verb forms (See Part 3) but has been expanded to identify words repeated multiple times for emphasis, and overgeneralization errors. Routines have been built into the SALT software to automate this identification process.

a. Identifying Words Which Are Repeated Multiple Times for Emphasis

Speakers sometimes use repetition to provide emphasis. To prevent these repetitions from overly inflating measures such as mean length of utterance (MLU) and number of total words (NTW), the underscore is used to link segments which are repeated more than once so that they are counted as one word. For example:
    C The dog ran ran_ran.

These linked segments result in new words, e.g., "ran_ran", which should be identified as instances of the word being repeated, e.g., "ran". To do this, follow the linked segment with a vertical bar "|" and the repeated word as shown in this example:
    C The dog ran ran_ran|ran.

Refer to Section 16 for details on linking words repeated for emphasis.

b. Identifying Words Due to Overgeneralization Errors

Overgeneralization errors result in non-words which should be identified as instances of the attempted word using the vertical bar. Consider the following examples:
    1) C He **falled|fall**[EO:fell].
    2) C Then the dog **droppeded|drop/ed**[EO:dropped] the bee_hive.

Notice in example (2) that the bound morpheme is marked in the word following the vertical bar. Also notice that overgeneralized words are marked as errors using the [EO] code.

Refer to Section 19 for more information on coding overgeneralization errors.

c. Automating Root Identification

To simplify transcription, lookup files containing words with their corresponding root forms are available. These root identification files (RIFs) are used to automatically identify a different word root than the one which was produced. The "Identify Roots" command in the Edit menu looks up all words in the transcript that have not been previously identified with the vertical bar. If the word is not found in the active RIFs, that word is ignored. If only one choice is found, the word is automatically identified. If a word contains more than one root option, the user is presented with a list of choices to select from. Note: the active RIFs are selected using the *Setup menu* → *Language Settings* option. There are two Spanish RIFs (see Part 3) , one French RIF (see Part 4) , and one English RIF.

The RIF file, <u>English Inflections</u>, contains approximately 5,500 inflected words. Because this is not a complete list, you should identify the bound morphemes as you are typing your transcript. This file may then be used to catch those you miss. *Refer to Section 9 for a discussion on marking bound morphemes.*

18. Elapsed Time

Timing your language sample is optional. If you insert timing markers at the beginning and end of your transcript, SALT will compute elapsed time and rate of speaking in terms of utterances/minute and words/minute. This information may be important in assessing a speaker's productivity or fluency as we have found rate of speaking to be highly correlated with age.

If you are working with a digitized language sample, the timing information is displayed while the sample is

playing. Just note when the sample begins and when it ends. If the timing information is not provided for you, then replay the entire sample using a stop watch or clock to measure the length of the sample.

Two timing lines are needed, one at the beginning of the transcript to initialize the clock, and one at the end of the transcript to stop the clock. You have the option of inserting other timing lines, perhaps minute markers, in the transcript.

a. *Format of the Timing Markers*

Timing markers are inserted on lines beginning with a hyphen. Type the hyphen, a blank space, and the time in one of the following formats:

Format	*Examples*
hours:minutes:seconds	01:04:23 or 1:04:23 or 1:4:23
minutes:seconds	1:00 or 5:13 or 05:13
:seconds	:05 or :5 or :75
hours:minutes:seconds.hundredths	00:00:35.78
hours:minutes:seconds:frames	00:06:15:08 *calculated at 30 frames/second*

b. *Starting the Clock*

When a new transcript is created using the header dialog box, an initial timing entry of 0:00 is automatically supplied on a line beginning with a hyphen. This is the default initial time and may be edited if you wish to start the clock at some other time. The initial clock time must precede the first utterance. If the initial clock time is missing, the default time of zero is assumed.

c. *Stopping the Clock*

To stop the clock, insert a timing line at the end of your language sample that provides the final clock time. If the initial time is 0:00, then the final time is the elapsed time. If the initial time is not 0:00, then the elapsed time is calculated by subtracting the initial time from the final time. In the following example the elapsed time is 3 minutes and 35 seconds.
> $ Child, Examiner
>   **- 0:08**
>   C Once there was a boy who had a frog.
>   ….
>   C And the boy took the baby frog home.
>   C The end.
>   **- 3:43**

d. *Sequential Timing Lines*

Sequential timing lines are defined as timing lines which don't have any entries between them, not even timing lines. The *Setup menu → Analyze Settings → Timing Lines* setting determines whether or not to subtract the elapsed time between sequential timing lines from the total transcript time (by default, the elapsed time is subtracted). Note that pause times are never subtracted from elapsed time.

- *Example of when you would want the elapsed time to be subtracted*:

   If there is an interruption during the language sample, you may need to stop and restart the clock. To do this, insert two timing lines with no utterances between them at the point of the interruption.

The elapsed time between these two timing lines is subtracted from the overall elapsed time. The following example illustrates how the clock is "paused" when someone came into the room. All the timing lines are entered as hours:minutes:seconds.

$ Child, Examiner
- 00:07:15
C Once there was a boy who had a frog.
...
= interruption – someone came in the room to get some papers
- 00:08:50
- 00:09:15
...
C And the boy took the baby frog home.
C The end.
- 00:11:15

Based on the beginning and ending timing markers, the total elapsed time is 4 minutes (from 7:15 to 11:15). But there are no utterances or pause lines between the middle two timing markers. So the time from 8:50 to 9:15 (25 seconds) is subtracted from the total elapsed time resulting in an elapsed time for this transcript of 3 minutes and 35 seconds.

- *Example of when you would NOT want the elapsed time to be subtracted*:

  Suppose you are including minute markers throughout your transcript and you are transcribing the sample where the client is mostly nonverbal, e.g., very young, uses a lot of non-transcribed sign language, relies heavily on communication boards. It is possible that there are no transcribed utterances between timing markers but you would want to include all elapsed time. You could either change the setting so that the time is not subtracted or you could insert a comment line between the timing markers.

19. Codes

Codes are the most flexible part of the SALT transcription conventions. You can create codes to mark anything that you are interested in quantifying for which there *is no* transcription convention. The power of this feature comes from the ability to directly call up the coded words and utterances and to obtain frequency summaries.

A code consists of characters enclosed within square brackets [like_these]. Codes cannot contain blank spaces and cannot be split between lines. Avoid using symbols which have special meaning in SALT such as transcript-entry and search symbols. In particular, avoid using the "@" sign and the "=" sign as their usage will generate a warning message in SALT.

There are two types of codes: word codes and utterance codes. A word code is any code which is attached to a word, and an utterance code is any code which is not attached to a word. Although you can devise your own codes to analyze any feature of the language sample that you are interested in, this discussion will focus on the codes used within the reference database samples.

a. *Word Codes*

    A word code is any code which is attached to the end of a word without any space between the word and the code. There's no limit to the number of words that may be coded or to the number of codes

that may be attached to each word. Consider the following examples:
1) C The big frog were**[EW:was]** mad.
2) C Then he**[NoRef]** pick/ed up the little frog.
3) C The dog fell from la**[CS]** ventana**[CS]**.

In these examples, word codes are used to mark problems with specific words: 1) the speaker used the word "were" but should have used "was", 2) the speaker used the pronoun "he" but failed to establish the referent, and 3) the speaker code-switched into Spanish. The [EW] code is a standard code for all the reference database transcripts. The [CS] code is a standard code for the Bilingual Spanish/English Story Retell database transcripts (*see Part 3*). The [NoRef] code is not a standard code. It is included here to illustrate that you may develop your own set of codes to mark any feature of the speaker's language you are interested in quantifying.

b. *Utterance Codes*

An utterance code is any code which is not attached to a word. This code may occur anywhere in the utterance before the end-of-utterance punctuation mark. There's no limit to the number of utterance codes that may be inserted. Consider the following examples:
1) C And the boy walk all up to the rock **[EU]**.
2) C And the frog not was there **[WO]**.
3) C While he was asleep, the frog got out of the jar **[SI-2]**.

In these examples, the utterance codes are used for a variety of purposes: 1) to mark an utterance-level error, 2) to mark non-standard word order, and 3) to code for subordination index, i.e., this utterance has two clauses. Remember, you can create codes to mark anything that you are interested in quantifying.

c. *Error Codes*

There is a special category of codes called "error codes". The five codes listed in this section, [EO:=], [EP:=], [EW:=], [EW], and [EU], are the default error codes (*note that the equal sign in these codes is a placeholder matching any 1 or more characters*). This list of codes, which SALT recognizes as error codes, can be changed using the *Setup menu* → *Lists* → *Current Code Lists* option. Although all codes are included in most code summaries, there are a few places, such as the *Standard Measures Report*, where only the error codes are summarized. The following error codes are used consistently when transcribing samples for the reference databases – four for word-level errors and one for utterance-level errors.

- [EO:word]

  This is a word-level error code used to mark overgeneralization errors. The correct word is put inside the brackets following a colon. There is no space between the word and the code. Consider the following utterances which all contain overgeneralization errors with the root form of the word identified using the vertical bar and the [EO] code attached to the end of the word:
  1) C He falled|fall**[EO:fell]**.
  2) C There were deers|deer**[EO:deer]** in the woods.
  3) C That/'s hises|his**[EO:his]** wife.
  4) C The next day the boy woked|wake**[EO:woke]** up.
  5) C Then the dog droppeded|drop/ed**[EO:dropped]** the bee_hive.

  In examples 1 – 4 the overgeneralized words are identified with the root form of the word. Notice that the bound morphemes in these words are not marked because we don't want to give the child

credit for an incorrectly-used bound morpheme. In example 5, however, the child is given credit for the first correctly-used bound morpheme but not for the second incorrectly-used bound morpheme (*notice that the bound morpheme is slashed in the word following the vertical bar rather than in the overgeneralized form*). *Refer to Section 17 for information on using the vertical bar to identify words.*

- [EP:word]

  This is a word-level error code used to mark pronoun errors. The correct word is put inside the brackets following a colon. There is no space between the word and the code. The samples in the SALT reference databases used the [EP] code to mark errors with personal pronouns, possessive pronouns, and reflexive pronouns. Consider the following utterances which all contain pronoun errors with the root form of the word identified using the vertical bar and the [EP] code attached to the end of the word:
    1) C Dr_De_Soto worked with her**[EP:his]** wife.
    2) C Him**[EW:he]** look/ed everywhere.
    3) C That/'s ours**[EP:our]** car.
    4) C She did it by herselves**[EP:herself]**.

- [EW:word]

  This is a word-level error code used to mark other words which are used incorrectly for which you know the intended word or part of speech. The correct word, if known, is put inside the brackets following a colon. There is no space between the word and the code.

  Examples:
    1) C The big frog were**[EW:was]** mad.
    2) C He check/ed on**[EW:in]** the boot.
    3) C The boy was find/ing**[EW:looking]** for him.

- [EW]

  This is a word-level error code used to mark extraneous words. There is no space between the word and the code.

  Examples:
    1) C And then the boy is a**[EW]** sleep/ing.
    2) C Then the kid climb/ed a rock to see more**[EW]** better.
    3) C And they were**[EW]** have nine baby/s.
    4) C The boy was in**[EW]** mad.
    5) C And after that the boy that**[EW]** climb/ed the tree.

- [EU]

  This is an utterance-level error code used to mark errors which cannot be associated with a specific word. It alerts the user to an utterance that would need more detailed analysis later. Look at the examples and notice that the [EU] code is inserted between the last word and the ending punctuation mark.

  Examples:
    1) C And they came to stop/ed **[EU]**.
    2) C They scream for everywhere **[EU]**.

3) C He was have/ing a frog **[EU]**.
4) C He was a tree **[EU]**.

The [EU] code is also used when there are too many things to fix by marking omissions and/or word errors.

> General Rule (expanded): Do not mark more than two omissions and/or word errors in an utterance. Instead, mark the entire utterance as having a problem by inserting the utterance code [EU] at the end of the utterance. Also, do not mark any utterance containing unintelligible segments as [EU] since it is difficult to know whether or not the unintelligible segments would

d. *Other Codes used in the SALT Reference Databases*

In addition to the error codes listed in the previous section, the following codes were consistently used when transcribing the samples for some of the reference databases.

- [FP]

  The "filled pause" code is a word-level code used to mark non-standard filled pause words in all of the SALT reference databases. The standard set of filled pause words include: AH, EH, ER, HM, UH, UM. When these words occur in mazes, the SALT software automatically recognizes them as filled pause words. Other words, when they occur in mazes, may or may not be used as filled pauses. The [FP] code is attached to these words when they are used as filled pauses. In the following example, both of the words within the maze are recognized as filled pause words.
      C The dog (um like**[FP]**) fell down.

  Refer to Section 10 for more information on marking mazes.

- [SI-0], [SI-1], [SI-2], …

  Subordination Index (SI) is a measure of syntactic complexity which produces a ratio of the total number of clauses (*main and subordinate clauses*) to the number of utterances. The SI analysis counts clauses. Language samples, which have been transcribed and segmented into C-units, have one of the following SI codes inserted at the end of each qualifying utterance: [SI-0], [SI-1], [SI-2], etc. which means subordination index – 0 clauses, 1 clause, 2 clauses, etc.. The "Subordination Index" reports, selected from the Analyze and Database menus, count the individual SI codes and compute the composite SI score.

- [WO]

  This is an utterance-level code, used in the Bilingual Spanish/English Story Retell databases, to signify words or phrases within an utterance which are out of order in Standard English or Standard Spanish. The content (semantics) of the utterance is correct; however the word order is awkward. For example:
      C And then fall down the dog and the boy [WO].

- [I]

  The "Imitation" code is a word-level code, used in the Bilingual Spanish/English Story Retell databases, to identify words which were provided to the speaker by the examiner. The first use of the word provided by the examiner is coded with the [I] code. Subsequent uses of the same word do

not receive the [I] code. For example:
    C And then the ((what is that call/ed)) <> owl[I] scare/ed him.
    E <Owl>.
    C Then the owl knock/ed him down.

In the first utterance, the child stopped in the middle of the utterance to ask the examiner for the name of the owl ((parenthetical remark)). The examiner then provided the word "owl" which the child immediately used. The child's use of the provided word receives the [I] code. Subsequent uses of this word are not marked as imitations. Note that if the examiner provides a vocabulary word but the speaker does not use that word until later in the sample, that word is not an imitation and does not receive the [I] code. Marking words provided by the examiner is useful when doing a detailed analysis of the vocabulary the speaker used to retell the story. *Refer to Section 15 for more information about marking parenthetical remarks.*

- [F]

  This is an utterance-level code placed at the end of each utterance lacking a stated subject as a result of segmenting utterances using modified communication units (Bilingual Spanish/English Story Retell databases). For example:
      C The gopher look/ed out of the hole.
      C and bit the boy [F].

  Refer to Section 5 on "Utterance Segmentation" for rules on modified communication units.

- [CS]

  This is a word code attached to all code-switched words (e.g., *Spanish words in English transcripts or English words in Spanish transcripts*). For example (*English transcript*):
      C The dog fell from la**[CS]** ventana**[CS]**.

20. Beginning and Ending Markers

By default, all entries in the transcript are included for analysis. There are times, however, when you may wish to exclude entries at the beginning or end of the transcript. The easiest way to do this is to insert beginning and ending markers in your transcript.

@begin

To exclude entries at the beginning of the transcript, Insert the entry @begin at the point in the transcript you wish to begin analysis. Do not put any spaces between the "@" sign and the word "begin". Suppose, for example, you elicited a story retell from a speaker who needed a lot of prompting to begin. You could start transcribing the sample at the point where the speaker begins the retell. However, you may want to capture the initial prompting as part of the transcript but you would not want to include these utterances in the analysis of the story retell. To do this, you would insert the entry "@begin" at the point in the transcript where the speaker begins the retell.

    $ Child, Examiner
    E Now it's your turn to tell me the story.
    E Tell me the story.
    C Um.

   - :03
   E Once there was^
   - :05
   @begin
   C A boy :03 and a dog.
   ...

   The @begin and @end markers come in sets. In this case there is an implied @end marker at the end of the transcript.

@end

To exclude entries at the end of the transcript, insert the entry @end at the point in the transcript you wish to end analysis. Do not put any spaces between the "@" sign and the word "end". As an example, suppose elicited a story retell followed by a set of comprehension questions. You could score the comprehension questions without transcribing them. However, if you wish to include the speaker's responses in the transcript, you would insert the entry "@end" at the end of the story retell, before the comprehension questions.

   ...
   C So he took the frog home with him.
   @end
   E Is that it?
   C Yes.
   E Great.
   E Now I would like to ask you some question/s about the story.
   E Ok, first question, who is look/ing for the frog?
   C The boy.
   ....

   The @begin and @end markers come in sets. In this case there is an implied @begin marker at the beginning of the transcript.

Multiple @begin and @end sets

To exclude one or more sections in the middle of your transcript, insert @begin at the point in the transcript you want to begin analysis and insert @end at the point in the transcript you want to end analysis. As an example, suppose you are eliciting a conversational sample and, as part of the language sample, the client narrates a movie he or she went to see. To analyze just the conversation, you would only include the conversational sections.

   $ Child, Examiner
   ...
   E What did you do over the weekend?
   C I watch/ed The_Lion_King.
   E Was it good?
   C Yes.
   @end
   C My favorite part was when Simba had to leave.
   C And then he made some friend/s.
   ...   { C continues to tell about the movie }

```
@begin
E Wow, you really like/ed the movie.
E Did you do anything else over the weekend?
C I forgot to tell you about something else in the movie.
@end
C Simba got in a fight.
...  { C continues to tell about the movie }
E Ok.
@begin
E Tell me about what you/'re do/ing in gym class today.
...
```

The @begin and @end markers come in sets. In this case there is an implied @begin marker at the beginning of the transcript and an implied @end marker at the end of the transcript.

Adjusted elapsed time

The beginning and ending markers, if found in your transcript, mark the sections of the transcript included for analysis purposes. The elapsed time, if included, is adjusted (estimated) to reflect the sections of transcript included for analysis. Suppose the elapsed time of the entire transcript is 4 minutes and includes 200 words (average of 50 words for each minute). If only 150 of the 200 words are included for analysis, the elapsed time is adjusted to 3 minutes.

Transcript cut

Suppose that the transcript cut is used to restrict the analysis to the first 50 utterances and your transcript contains one or more @begin or @end markers. The 50 utterances would be selected from the utterances included in the analysis, i.e., those not excluded by the @begin and @end markers.

Both the transcript cut and the beginning and ending markers are used to specific the section of the transcript included for analysis. The transcript cut affects all transcripts while the beginning and ending markers only affect transcripts containing the markers.

Also see

- **Help → Resources → Transcription Conventions → Summary of Transcription Conventions**
- Appendix A for the rationale behind which bound morphemes are marked

Appendix A: Rationale Behind Marking Bound Morphemes

Originally posted as a SALT blog on December 11, 2018

Why do we mark some bound morphemes and not others?

The SALT reference databases use a very specific set of conventions for marking bound morphemes. Our protocols do not represent the only way to mark morphemes but if you want to compare your sample against our databases, it is important for you to use these same conventions.

The SALT rules for marking bound morphemes may seem obscure at first. But there is method to our madness. This blog explains the rationale behind SALT's conventions for marking bound morphemes. It goes beyond, "because it's been this way for a long time and here are the rules to memorize". Or, a little more clinically, "because these are the developmental morphemes and here are the rules to memorize". Our hope in writing this blog is that by understanding the "why" behind the conventions, the rules will make more sense and, hopefully, it will be easier to implement them.

To a large extent, SALT's convention for marking bound morphemes is patterned after the conventions used by Roger Brown (1973) in the calculation of MLUm (mean length of utterance in morphemes) back in the early days of language sample analysis.

Inflectional morphemes versus Derivational morphemes

Brown excluded derivational morphemes in the calculation of MLUm. To understand why, we first need to distinguish between the two types of bound morphemes – inflectional and derivational. *Many of the explanations and examples in this section are taken from the following websites:*
> https://semanticsmorphology.weebly.com/inflectional-and-derivational-morphemes.html
> http://www.mathcs.duq.edu/~packer/Courses/Psy598/Ling-Morphology.pdf

- **Inflectional morphemes** are used to show some aspects of the grammatical function of a word. *They are always suffixes and always result in the same part of speech.* We use inflectional morphemes to indicate if a word is singular or plural, whether it is a comparative or possessive form, and to mark tense. *Inflectional morphemes never change the grammatical category* (part of speech) of a word. For example, *shoe* and *shoes* are both nouns, *tall* and *taller* are both adjectives, and *look* and *looked* are both verbs. The inflectional morphemes simply produce different versions of the words.

     There are eight inflectional morphemes. They are shown in the following table:

Inflectional Morphemes		Added to	Example
-s, -es	plural	nouns	I have two black cat**s**.
-'s, -s'	possessive	nouns	My dog**'s** bark is very loud.
-er	comparative	adjectives	I have long**er** hair than you do.
-est	superlative	adjectives	He has the bigg**est** pumpkin.
-s	3rd person singular	verbs	She run**s** fast.
-ed	past tense	verbs	He play**ed** basketball.
-en	past participle	verbs	She has eat**en** everything.
-ing	progressive tense	verbs	He is play**ing** basketball.

     When children use inflectional morphemes, they are (generally) demonstrating their knowledge of the base

word *as well as* their ability to encode the plural, possessive, or tense of that root word.

- **Derivational morphemes**, in contrast, are used to create new words or to make words of a different grammatical class (part of speech) from the root form. For example, by adding the derivational morpheme -er the verb *read* becomes the noun *reader*. The addition of -ize changes the adjective *normal* to the verb *normalize*. Similarly, we can derive the adjectives *helpful* and *helpless* by adding -ful and -less to the noun *help*.

However, some derivational morphemes do not change the grammatical category of a word but they do significantly change the meaning of the word. For example, we can derive the nouns *neighborhood* and *kingdom* by adding the derivational suffixes -hood and -dom to the nouns *neighbor* and *king*. And derivational prefixes such as un- and re- generally do not change the category of the word to which they are attached. Thus, both *happy* and *unhappy* are adjectives, and both *fill* and *refill* are verbs. But each of these pairs of words, although clearly related, have very different meanings.

Derivational morphemes may be either suffixes or prefixes and usually, but not always, result in a different grammatical category. The following table lists some of the common derivational morphemes:

Common Derivational Morphemes (Suffixes)	Added to	Results in	Examples
-ize	nouns adjectives	verbs verbs	rubber**ize** normal**ize**
-ful	nouns	adjectives	play**ful**, help**ful,** beauti**ful**
-ly	nouns adjectives	adjectives adverbs	man**ly**, friend**ly** proud**ly**
-sion	verbs	nouns	discus**sion**
-hood, -dom	nouns	<same>	neighbor**hood**, king**dom**
Derivational suffixes which overlap with inflectional suffixes (*though they serve a different purpose*)			
-er	verbs nouns/verbs	nouns nouns	read**er** grad**er**
-ed	verbs	adjectives	am tir**ed**, was bor**ed**
-en	verbs	adjectives	this spot is tak**en**
-ing	nouns/verbs nouns/verbs	nouns adjectives	bik**ing** is fun interest**ing** story
Common Derivational Morphemes (Prefixes)	Added to	Results in	Examples
un-, a- dis-, re- anti-	Adjectives verbs nouns	<same>	**un**happy, **a**typical **dis**like, **re**fill, **re**evaluate, **re**view **anti**-aircraft

According to Brown, young children generally do not learn a base word and then apply a derivational morpheme to encode extra information. Instead, they usually learn these as fully-formed, independent words with their own specific meaning. Although derivational morphemes can logically be split into a root word and a prefix or suffix, these smaller parts are not meaningful to the child speaker and so they should not be considered separate morphemes in this case.

Guo, et al. (2018) gives evidence for this, including:
- There is ample psycholinguistic evidence that base words and derived words (e.g., *beauty*, *beautiful*) are stored as separate lexical entries and should be given equal weight.

- Children may learn a derived word (e.g., *beautiful, interesting*) before the base word (e.g., *beauty, interest*). It seems unlikely that children would add the derivational morphemes to these base words to form the derived words. <u>Side note</u>: to test this, we looked at samples from 355 typically-developing children under the age of 7 taken from the SALT Play and Conversation databases. The derived word *beautiful* was used five times while its base form *beauty* was only used once. And *interesting* was used twice and there were no instances of its base form *interest*.

They consider that derivation is a word-formation process, not a grammatical encoding process. Therefore, derivation reflects a speaker's lexical skills, not grammatical skills.

So what are the rules behind the SALT conventions?

RULE 1: Do not mark derivational morphemes. **Do mark (most) inflectional morphemes.**

Why?
In short, we want to mark bound morphemes when they reflect the child speaker's understanding that the prefix/suffix has a meaning separate from the root word. When using derivational morphemes - learned as fully-formed, independent words - a child speaker is only utilizing a single meaning. In contrast, when using an inflectional morpheme, the child is utilizing two meanings: the root word and the encoded plural/possessive/tense meaning.

By not marking derivational morphemes, we do not give the speaker credit for bound morphemes which change the meaning of the word (e.g., happy → unhappy) or change its grammatical category (e.g., friend → friendly).

However, when the child speaker is likely to have understood the separate meanings of the bound morphemes, we do want to mark them. So we mark most inflectional morphemes. Most, but not all...

RULE 1a: Do not mark the comparative (inflectional) morphemes -er and -est.

Why not?
Although -er and -est are inflectional morphemes, Brown did not count them because they are not obligatory. According to Guo, et al (2018), this means that it is a stylistic choice whether to use comparative and superlative form rather than the uninflected adjective. For instance, when given a choice of several balls of varying sizes, a child may select the largest one and say, "*I have the big one*" unless prompted to make a comparison.

RULE 1b: Do not mark irregular forms.

Why not?
Irregular forms are counted as single morphemes because children (generally) learn them as separate forms, rather than inflections of their base forms.

The following table lists examples of irregular words:

Category	Examples of irregular words
plural	man → men, foot → feet, cactus → cacti, deer → deer
all possessive pronouns	I → mine, he → his, she → hers, we → ours, you → yours, it → its, they → theirs
3rd person singular	have → has, is → was
past tense	begin → began, break → broke, go → went, get → got

past participle (*regular form is present tense + EN as separate syllable*)	begin → begun, break → broken, go → gone, get → gotten, see → seen, be → been
negation	will → won't

Some words are irregular because the sound of the base form changes. These words follow the standard spelling for inflected or contracted words but change its sound. Some examples follow:

Category	Examples of changed sound
plural	leaf → leaves, wolf → wolves
3rd person singular	do → does, say → says
past participle	drive → driven, write → written
negation	do → don't

RULE 1c: Do not mark plurals for words which do not have a singular form.

Why not?
Children would not have learned the singular form in order to then apply the rule for plurals. Following are some examples:

Examples of Plurals Without a Singular Form			
belongings	glasses (*spectacles*)	pants	shorts
binoculars	goggles	panties	suds
breeches	jitters	remains	tights
clothes	knickers	riches	trousers
drawers	pajamas	shenanigans	tweezers

RULE 2: Do not mark concatenatives.

Why not?
Brown counted concatenatives as single morphemes because, like irregular forms, children may have stored them as holistic chunks. Following is a list of concatenatives:

Examples of Concatenatives (*meaning*)			
betcha (*bet you*)	liketa (*like to*)	outta (*out of*)	useta (*used to*)
coulda (*could have*)	lookit (*look at it*)	shoulda (*should have*)	wanna (*want to*)
gonna (*going to*)	musta (*must have*)	sposta (*supposed to*)	whatcha (*what are you*)
gotta (*got to*)	oughta (*ought to*)	trynta (*trying to*)	woulda (*would have*)
hafta (*have to*)			

RULE 3: Mark contracted words.

Why?
Contractions combine two words into one (e.g., *we are* → *we're*). The speaker is given credit for the same number of morphemes whether using two words or the one contracted word.

Contractions		Examples
-'t -n't	negation	I can**'t** leave yet. He does**n't** know better
-'s	is has does us	It**'s** time to go. He**'s** been sick. What**'s** he do for a living? Let**'s** go.
-'re	are	You**'re** late.
-'m	am	I**'m** ready to take the test.
-'ll	will	I**'ll** wait over here.
-'d	would had did	He**'d** do it. He**'d** better leave now. Why**'d** the boy look over there?
-'ve	have	We**'ve** a lot to do.

Summary

These rules can be summarized as:

- Only mark the following inflectional morphemes and contractions.

Inflectional Morphemes		Contractions	
/s	plural	/'t, /n't	negation
/z	possessive	/'s, /'re, /'m	is, are, am
/3s	3rd person singular	/'ll, /'d	will, would
/ed	past tense	/'ve, /h's, /h'd	have, has, had
/en	past participle	/d's, /d'd	does, did
/ing	progressive tense	/'us	us

- Do not mark irregular forms, concatenatives, or plurals which do not have a singular form.

Formatting Notes

- Use a slash (/) for bound morphemes which follow the free morpheme (suffixes) and use a backslash (\) for bound morphemes which precede the free morpheme (prefixes). There should be no spaces between the free morpheme and the bound morpheme(s).
- When the spelling of a free morpheme such as CRY changes with the addition of the bound morpheme, use the root spelling of the free morpheme (as if the bound morpheme is not there). Then, simply add the slash plus the bound morpheme (i.e. CRY/ED). If this is not done, the stem CRI will be treated as a different word from CRY and thereby inflate Type-token ratio (TTR) as well as Number of Different Words (NDW).

A Final Note

While we feel that Roger Brown's research published in 1973 has stood the test of time, we understand that it is not the only way to understand the use of morphemes or to calculate MLUm. If you have a language sample from which you want to compare the transcript with samples selected from the SALT reference databases, then these conventions are the most appropriate. However, these conventions may not be appropriate for all speakers or in all cases. Future blog posts will highlight some of these alternate scenarios and how they can be accommodated using custom coding schemes in SALT.

References

Bedore, L.M. (1999). The acquisition of Spanish. In O. Taylor and L. Leonard (Eds.), *Language Acquisition Across North America: Cross-cultural and cross-linguistic perspectives* (pp. 157-208). San Diego, CA: Singular Publishing Group, Inc.

Bedore, L.M. (2001). Assessing morphosyntax in Spanish-speaking children. In A. Iglesias (Ed.), Communicative Assessment of the Hispanic Child. Seminars in Speech and Language, 22:1, 65-77.

Brown, R. (1973). *A first language: The early stages.* Cambridge. MA: Harvard University Press.

Guo, L., Eisenberg S., Bernstein Ratner, N., & MacWhinney, B. (2018). Is Putting SUGAR (Sampling Utterances of Grammatical Analysis Revised) Into Language Sample Analysis a Good Thing? A Response to Pavelko and Owens (2017). Language, Speech, and Hearing Services in Schools, doi:10.1044/2018_LSHSS-17-0084.

Hughes, D., McGillivray, L., Schmidek, M. (1997) *Guide to Narrative Language: Procedures for Assessment.* Eau Claire, WI: Thinking Publications.

Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve.* Urbana, IL: National Council of Teachers of English.

Rojas, R., & Iglesias, A. (2006). Bilingual (Spanish-English) narrative language analyses: Why and how? *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, *13:1*, 3-8.

Strong, C. (1998). *The Strong Narrative Assessment Procedure.* Eau Claire, WI: Thinking Publications.

Index