Transcription Guidelines

There are various tools available for transcribing recordings, and LaBB-CAT supports the transcription file formats used by the most commonly used tools. Each of these tools has its own capabilities for specifying speakers and meta data, and adding annotations.

Beyond the specific tool and file format used for transcription, there are some general principles that can facilitate subsequent processing of speech data in LaBB-CAT.

Spelling

Many automatic annotation tasks involve looking up standard dictionaries, and words that are not found are not annotated, so it's important to use standard spelling where possible.

- Use conventional spelling, and if you are unsure of how to spell something, look it up in a dictionary, or on a map.
- Write all numbers out in full, with spaces instead of hyphens - e.g.
  - × "1984"
  - × "nineteen-eighty-four"
  - ✓ "nineteen eighty four"
- When abbreviations are used, use capital letters with spaces in between each letter if each letter is said separately, otherwise use capitals with no spaces - e.g.
  - × "NSA"
  - × "N A S A"
  - ✓ "N S A"
  - ✓ "NASA"
- All words should be spelt out in full, like "and" and "suppose". Final gs and ds should not be dropped from words even if that's what the speaker says - e.g.
  - × "skippin' an' jumpin' an ol' rope"
  - ✓ "skipping and jumping an old rope"
- A single word should always be spelt as an entire word, even if there is a pause between syllables.
- Don't tidy up the speech. Leave in the repetitions, fillers and errors.

There may be a limited set of shortened words and contractions defined, which is fine as long as they're consistently used - e.g. if you use "cos" as a shortened version of "because", always spell it "cos", and never "cause", nor "'cause", nor "coz". For example:

- gonna
- sorta

- cos
- kinda
- gotta
- dunno
- wanna
- yip
- yeah
- okay
- uh huh
- gee
- jeez

Disfluencies

It's important to be consistent with the spelling of filled pauses:

- ah
- er
- um
- mmm

The spelling of the last of these, with three m's, is recommended because if it's spelled with one m - "m" - this can match the name of the letter "M" in the dictionary, so the pronunciation can be tagged as `/ m/`, and if it's spelled with two m's - "mm" - this sometimes matches an alternative spelling of the word "millimeter", so the pronunciation can be tagged `/'m -l -"mi-tə/`.

Unfilled pauses can be transcribed with a hyphen surrounded by whitespace; some modules use such pause information to help with automatic annotation (e.g. force-alginment with HTK benefits with pause annotations like this) - e.g.

- × "stop-before you begin"
- ✓ "stop - before you begin"

Incomplete words should marked with a tilde ~ (not a hyphen which may be interpreted as a pause) at the end of the word e.g.:

- × "hesi-"
- ✓ "hesi~"

For very short hesitations - "b~ bu~ but" - some pronunciation module can infer a pronunciation for such words, without the need for a manual pronunciation tag.

Word Tags and other in-situ Annotations

Some transcription tools allow tagging individual words with extra information, but others do not. For these, the only way to, for example, tag a word with its pronunciation, is to use transcription conventions.

LaBB-CAT optionally supports the following transcription conventions, if you are using ELAN transcripts, Praat TextGrids, or plain text files for transcripts:

- The pronunciation of an invented word or a hesitation can be marked by using square brackets immediately after the word (i.e. with no space between the word and the annotation) - e.g.
  "stut~[stVt]"
- The full form standard form of a hesitation (or other word with non-standard spelling) can be marked by using parentheses immediately after the word (i.e. with no space between the word and the annotation) - e.g.
  "stut~[stVt](stutter)"
- Noises can be annotated using square brackets surrounded by white space, e.g.
  "now [tongue click] where were we"
- Comments can be added by using curly braces surrounded by white space, e.g.
  "It hit me about here {points to temple}"

Utterances/Lines

Some processes, like forced alignment, involve processing individual utterances in the recording, which correspond to lines of text in many transcription systems. Very long or very short utterances can be difficult to process.

Ideally, each line in a transcript should be 5 to 15 words long, and line breaks should be made where there are pauses in speech.

Some annotation tools allow for marking periods of simultaneous speech - i.e. periods during which there's more than one person speaking. These periods should be aligned as accurately as possible, because some automatic processing (e.g. forced alignment) ignore simultaneous speech; short simultaneous-speech utterances ensure that as little speech as possible is ignored.