



LHCP-ASR

Manual revision guidelines

<https://www.mllp.upv.es>

12 March 2024

1 Goals

The main goals for this revision process are as follows:

1. To generate verbatim transcriptions of each talk that could be suitable for ASR system development and evaluation.

2 Transcription guidelines

Spelling:

- Use British spelling (in accordance with the [CERN style guide](#)).

Always transcribe what was said:

- We are looking for verbatim transcriptions (not subtitles).
- Transcribe as phonetically as possible, using valid words (except incorrect pronunciation that changes meaning, see “Speech disfluencies”).
- Always transcribe what is spoken, do not change or correct what the speaker says (except incorrect pronunciation that changes meaning).
 - Example: (saying) “*it’s me!*” → “*it’s me!*”. Do NOT transcribe as “*it is me!*”.
- Do not correct grammar.
- Include punctuation marks and capital letters, when needed. BUT do not capitalise at the beginning of a sentence, except when the word is to be capitalised by itself.
- If a punctuation mark is spoken, it should be transcribed as a word.
- When the speaker corrects themselves: transcribe what was said.
- Do not use abbreviations, e.g., “*Mr.*” → “*Mister*”
- Use standard (ASCII) punctuation mark symbols.

Letters and spelled out words:

- Each letter will be individually typed and uppercased, separated from the other letters by a space.
 - Example: “*my name is Smith S M I T H*”

- For acronyms (e.g., “*USA*”), the individual letters will not be separated by spaces. However, if the speaker pronounces all the words (e.g., “*United States of America*”), keep the expanded text as transcription.

Numbers, digits, ordinals, variable names, equations:

- Numbers, digits and ordinals will be transcribed as words.
 - Example: “*four hundred eighteen, forty-fourth, two thousand and two*”
- Decimals are transcribed as words.
 - Example: “*two point six*”
- Greek-letter symbols (variables, constants, functions...) are transcribed as words.
 - Example: “*epsilon, lambda, ...*”
- Equations must be written out in words as they are spoken.
 - Example: “*P equals N over V times R T*”
- Zero (“0”), when pronounced “o”, must be transcribed as “o”.
- Further examples: “*three squared plus two, seven point three minus three point eight*”.

Non-speech events:

- Do NOT transcribe them. Examples:
 - Speaker noise (unvoiced sounds): lip, smack, cough, etc.
 - Non-stationary noise: door slam, window, etc.
 - Music.
 - Chatter and noise.

Speech disfluencies:

- **Hesitations:** Label them according to language-specific instructions (see Section 2.1).
- **Incorrect pronunciation:** A speaker incorrectly utters a word from a phonetic perspective, causing a different word/sentence meaning. Transcribe with the correct word (the one they wanted to say).

- Example: (saying) “*beer* in mind” → (transcript) “*bear* in mind” (from the context, they actually wanted to say “*bear*”).
- But not: (saying) “has to be *uphold*” → (transcript) “has to be *upheld*” (as the meaning doesn’t change, it’s just a grammatical error, keep “*uphold*”)
- **Foreign words:** For words, titles, etc. in foreign languages that are pronounced differently from the main spoken language pronunciation rules, they are to be transcribed in the foreign language.
 - Example: “*there was a coup d’état in ...*”, “*Thank you, Monsieur Président*”.
- **Repetition:** The speaker, unintentionally, repeats a word or expression. It is transcribed as it is pronounced.
- **Word cut-offs:** Imagine the speaker utters only half a word. This half-word is NOT to be transcribed.
 - Example: “*From the bot- uh the bottom of my heart*” → “*From the uh the bottom of my heart*”.

Other situations:

- **Incomprehensible passages:**
 - Short ones (i.e., 1-3 words, isolated or within a comprehensible passage): use <UNK> for the unknown words.
 - Moderate-long to long ones: do not transcribe them (leave them out of any segment).
- **Overlapping speech:** Transcribe only what the main speaker says.
- **Foreign language sentences or passages:** Do not transcribe them (leave them out of any segment).

2.1 Language-specific instructions

2.1.1 English

- Label hesitation sounds as “uh” (“um” is also accepted).

3 Use of the transcriptions for development and evaluation

These guidelines are intended to be useful in order to obtain revised transcriptions that are verbatim with respect to the audio, and also that contain annotations (e.g., hesitation annotations) that might be useful not only for system development and evaluation, but maybe also for system training.

When using the revised transcriptions resulting from the use of these guidelines for development and evaluation, it might be important to consider the removal or modification of some of the annotations in order to compute more meaningful scores, including:

- Hesitations: Labels according to language-specific instructions (see Section 2.1), e.g., “uh” and “um” for English.
- Repetitions: Cases when the speaker, unintentionally, has repeated a word or expression.
- <UNK> label (short incomprehensible passage).
- Punctuation: Some or all punctuation marks could be removed.
- Capitalisation: To complete text capitalisation, every letter after a full stop could be uppercased. Alternatively, to remove capitalisation, some or all uppercase letters could be turned into lower case (proper nouns, acronyms, spelled out words...).
- Spelling: According to these guidelines, British spelling rules are to be followed. This could be taken into account for score computation (e.g., does the evaluated system produce British spelling? Otherwise, should this be penalised in the score?).
- Numbers: According to these guidelines, numbers are to be transcribed as words. This could be taken into account for score computation (e.g., does the evaluated system produce numbers written as words? Otherwise, should this be penalised in the score?).