

Guidelines for Transcription and Tagging for the KIT Speaking Test Corpus Version 1

Yusuke Tanaka* and Katsunori Kanzawa

1 Guidelines for Transcribing Words

1.1 Permissible Symbols

- Comma, period, question mark: [,] [.] [?]
 - Commas, periods, and question marks are to be inserted by the transcriber at appropriate places. Even if a sentence does not form a proper question, insert a question mark if it seeks confirmation or asks something, and the intonation rises at the end.
- Apostrophe: [']
 - Use for possessives and contractions where apostrophes are typically required.
- Hyphen: [-]
 - Use appropriately as needed, such as in compound nouns.
- Double quotation marks: [“ ”]
 - Use for titles of movies or books, or for quoting someone’s speech.

1.2 Symbols Not Permitted

- Exclamation mark: [!]
 - Do not use the exclamation mark even in emphatic or exclamatory sentences.
- Single quotation marks: [‘ ’]
 - Avoid using single quotation marks to prevent confusion with apostrophes.
- Colon, semicolon: [:] [;]
 - Do not use colons and semicolons in transcriptions, even in contexts where standard written language would typically include them.
- All brackets: [()] [[]] [<>]
 - Avoid all forms of brackets except for angle brackets (<>) used for tagging.
- Numbers
 - Spell out all numbers.
- Abbreviations
 - Abbreviations such as No. (number), vol. (volume), and est. (established) are strictly prohibited. All such terms must be spelled out in full using the alphabet.

*Faculty of Economics, Kumamoto Gakuen University; Email: yu-tanaka@kumagaku.ac.jp

1.3 Insertion of Spaces

- Spaces must be added:
 - After punctuation marks (comma, period, question mark).
 - Adjacent to tags.
 - * e.g., <F>ah</F> <F>mm</F> <F>mhm</F>
 - When spelling out alphabets letter by letter.
 - * Do not insert spaces for acronyms like TOEIC or SNS.
- Spaces must not be inserted:
 - Between a word and the tags that enclose it.
 - * e.g., <F>ah</F> <F>mm</F> <F>mhm</F>

1.4 Spelling

Use American English spelling as a general rule. Transcribe English as closely as possible to how it is heard. Even if the pronunciation is influenced by Japanese English, spell out the words if it is clear from the context which words are being spoken. Pronunciation errors such as pronouncing [r] as [l] can be ignored.

2 Tagging Methods

The tags used in this transcription are as listed in Table 1. The sequence of applying these tags is as follows:

- <?><CO, TO, RE><laughter><R, SC><JP><F>want</F> ...

Table 1: List of tags

Tag	Description
<F> </F>	Fillers, backchannels, and exclamatory words
<R> </R>	Repetition (confident in transcription)
<R?> </R?>	Repetition (not confident in transcription)
<SC> </SC>	Self-correction (confident in transcription)
<SC?> </SC?>	Self-correction (not confident in transcription)
<TO> </TO>	Timeout
<RE> </RE>	Recording error
<nvs> </nvs>	Non-verbal sound
<CO> </CO>	Interrupted speech
<?> </?>	Uncertain words
<??> </??>	Completely unintelligible words
<H pn="X"> </H>	Proper nouns, discriminatory terms, etc.
<JP> </JP>	Japanese words or expressions
<.> </.>	Pause of 2–3 seconds
<..> </..>	Pause of longer than 3 seconds
<laughter> </laughter>	Laughter during speech

2.1 Fillers

- Fillers are sounds made when thinking about what to say next. Common fillers include:
 - (1) ah, eh, er, em, erm, mm, mhm, uhu, ah-huh, oh, hum, wow, you know, well, let me see
- If a sound does not match any of the previously mentioned examples yet is considered a filler, it should be transcribed as accurately as possible. All fillers must be enclosed within <F></F> tags. When several fillers occur in sequence, each filler must be individually tagged with its own <F></F>.
 - (2) <F>ah</F> <F>mm</F> <F>mhm</F>
- When a filler has a distinctly Japanese sound, both the <F></F> and <JP></JP> tags should be applied. The tags must be nested in the order of <JP><F></F></JP> to maintain consistency.
 - (3) <JP><F>etto</F></JP>
- If a filler is repeated, apply a repetition tag to each occurrence.

2.2 Repetition

- If the same word or expression is repeated, the first instance should be tagged with <R></R>.
 - (4) When <R>he</R> <R>he</R> he was a child ...
 - (5) <F>Oh</F> <R>there is a</R> there is a station and this town located between the rivers.
- Fragments of subsequent words should also be tagged with <R></R>.
 - (6) When he <R>wa</R> <R>wa</R> was a child.
 - (7) <R>I wan</R> I want to build a kind of rotary in the station.
- The spelling of fragments should match the subsequent word, if known.
 - (8) This book is <R>inter</R> interesting.
- If a sound break causes a change in transcription, adjust the spelling to closely match the sound. Even if the spelling differs from the subsequent word, apply the <R></R> tag if the portion can be identified as a fragment.
 - (9) I think he is a very <R>ka</R> kind person.
 - (10) My <R>pe</R> parents don't allow me to live in Tokyo.
- Fillers may also appear within <R></R>.
 - (11) <F>Oh</F> <R>there <F>mm</F> is a</R> there is a station and this town located between the rivers.
- If the transcriber is not confident about the segment to which <R></R> applies, use <R?></R?> instead.

2.3 Self-correction

- The <SC></SC> tag is applied to segments of speech where the speaker hesitates before settling on a final expression. It is crucial that the tagged segment and the following section are not identical.

(12) He <SC>don't</SC> doesn't know anything about this.

(13) <SC>He passed the exami</SC> he will pass the examination <SC>last</SC> next year.

- Even if the correction occurs in the order of a correct expression followed by an incorrect one, the latter is considered the expression that the speaker ultimately deemed appropriate. Therefore, the earlier part is treated as hesitation and tagged with <SC></SC> as usual.

(14) He <SC>doesn't</SC> don't know anything about this.

- When there are multiple instances of hesitation or self-correction, reapply the <SC></SC> tags at points judged to be appropriate boundaries.

(15) It would be a kind of trash so <SC>it's a</SC> <SC>it's waste</SC> it's a kind of waste.

- If the correction follows the pattern where Expression A is followed by Expression B and then returns to Expression A, <SC></SC> should still be applied to both the initial Expression A and Expression B. This is because, even though the first Expression A matches the final expression that the speaker deemed appropriate, another hesitation occurs in between.

(16) He <SC>doesn't</SC> <SC>don't</SC> doesn't know anything about this.

- If it is difficult to determine whether a segment is a simple repetition (<R></R>) or a reformulation/self-correction (<SC></SC>), follow the guidelines below.
- Cases where <R></R> should be prioritized: In (17), it is unclear whether “vivid” (an adjective) was uttered as a hesitation before “vividly” (an adverb), or whether “vivid” is simply a sound fragment of “vividly”. In such cases, as in (17), <R></R> should generally be prioritized.

(17) It's difficult to keep the plants <R>vivid</R> vividly.

- Cases where <SC></SC> should be prioritized: For **verb person, verb tense, and noun number**, prioritize <SC></SC>. Also, since “I'm” and “I am” differ only in pronunciation and do not differ in basic meaning, apply <SC></SC>.

(18) I <SC>close</SC> closed the door when I left school.

(19) <SC>Many line</SC> many lines are passing the station so this must be a big town.

(20) <SC>I'm</SC> I am a high school student.

- In some cases, <R></R> may be embedded within <SC></SC>, as shown below.

(21) <SC>It's <R>pla</R> planning</SC> it's planned by my teacher.

- However, in cases such as the following, <R></R> should not be embedded within <SC></SC>.

(22) It's <R>pla</R> <SC>planning</SC> planned by my teacher.

- In some cases, <SC></SC> may be embedded within another <SC></SC>, as shown below. In (23), “it’s planning” is a hesitation before “it’s planned”, and within “it’s planning” there is a self-correction from “planned” to “planning”.

(23) <SC>It’s <SC>planned</SC> planning</SC> it’s planned by my teacher.

- Fillers can also occur within <SC></SC>.

(24) <SC>It’s <F>er</F> planning</SC> it’s planned by my teacher.

- If the transcriber is not confident about the segment to which <SC></SC> applies, use <SC?></SC?> instead.

2.4 Interrupted Utterances

- Utterances where the sentence ends midway should be entirely enclosed within <CO></CO> tags. **Always place a period immediately after the </CO> tag.**

(25) <F>Oh</F> O K. So it’s getting dark but is it O K for you to come out? <CO>Is that</CO>.

- Learners often begin to say ‘so’ and then falter. In such instances, tag the utterance as described below. Note that any fillers following ‘so’ should not be considered part of the interrupted utterance.

(26) <CO>So</CO>. <F>Um</F>.

2.5 Timeout & Recording Error

- Speech that is cut off due to running out of response time should be enclosed within <TO></TO> tags. Similarly, speech that begins before recording starts, and is thus missing the beginning, should be enclosed within <RE></RE> tags.

2.6 Proper Nouns

- Proper noun tags should be applied to any proper nouns within the corpus that could cause issues upon release. Within the tags, select the appropriate code for proper nouns from Table 2. These tags must be used for any words that could identify test takers or others, such as personal names, school names, or company names. Apply tags appropriately even if these details are not directly mentioned but can be inferred from the context. Additionally, use these tags for the entire statement if it contains discriminatory remarks or defamation. When unsure whether a tag is necessary, err on the side of caution and apply one. Names of public figures (politicians, writers, historical figures, celebrities) and titles of books or movies generally do not require tags, except in cases of defamation. Tags should not be applied to statements that are critiques or reviews unless there is uncertainty about their appropriateness, in which case, apply a tag as a precaution.
 - Targets: Names of test takers, their families or friends, nicknames, affiliated companies, and school names that could lead to personal identification.
 - Not Targeted: Items that do not facilitate the identification of the test taker or their associates, such as pet names, titles of movies or books, and names of celebrities, writers, politicians, etc.
- Always append a serial number to each code to enable differentiation between multiple terms associated with the same type of code. If a word appears more than once within a single utterance, continue to assign a unique number to each occurrence of that word. Even if a term appears only

Table 2: Proper noun codes

Target	Usage
1. Personal names	<H pn = “name1”>...</H>
2. School names	<H pn= “school name1”>...</H>
3. Company names	<H pn= “company name1”>...</H>
4. Other	<H pn= “others1”>...</H>

once within an utterance and no other terms requiring the same type of code are present, still assign it a number.

(27) I am studying at <H pn= “school name1”>K University</H>. Before that, I used to study at <H pn= ” school name2 ” >F University</H>. Two years ago, I moved to <H pn=“school name1”>K university</H> to do more specific research.

2.7 Segments with Uncertain or Impossible Comprehension

- For segments that can be transcribed based on context guessing but where comprehension remains uncertain, apply the <?></?> tags.

(28) <?>They</?> should be very beautiful.

- If it is completely impossible to comprehend what is being said, and context guessing is not feasible, making transcription impossible, apply the <??></??> tags empty.

(29) <??></??> should be very beautiful.

2.8 Use of Japanese

- When Japanese is spoken, apply the <JP></JP> tag.

(30) <F>Mm</F> <R>I</R> I don’t like <JP>osechi</JP>.

2.9 Pauses

- If a pause occurs in speech, apply a <.></.> tag for pauses lasting 2–3 seconds and a <...></...> tag for pauses exceeding 3 seconds.

2.10 Non-verbal Sounds

- Use tags from Table 3 for non-verbal sounds such as laughter, sighs, coughs, and yawns.

Table 3: Non-verbal sound tags

Tag	Usage
<nvs>laughter</nvs>	Laughter or nervous laughter
<nvs>sigh</nvs>	Sigh
<nvs>cough</nvs>	Cough
<nvs>yawn</nvs>	Yawn

2.11 Speaking While Laughing

- When speaking occurs alongside laughter, enclose the relevant section within <laughter></laughter> tags.

(31) It's a kind of <laughter><JP>mama-chari</JP></laughter>.