General Transcription Guidelines

Casing

All words and names, other than spelled-out letters (see below), must always be transcribed in lower case. This includes the personal pronoun "I".

Examples:

- **i** need directions to **montreal**
- **maria** goes to **walmart** in **seattle**
- do you know **google chrome**

Initialisms, Acronyms and Spelled-out Letters

Only **initialisms** and **spelled-out letters** should be capitalized. When a letter is spelled-out, it must be capitalized, followed by a period, then by a space.

Examples:

- i'll be available from nine **A. M.** until five **P. M.**
- the images are stored on my **U. S. B.** drive
- i bought a new **T. shirt**
- the correct spelling is **D. A. V. E.**
- all my data is stored on several **C. D. s**
- **C. N. N. 's** coverage began at nine **P. M.**

As highlighted above, you should use an apostrophe to indicate possession (**C. N. N. 's**), but not plurality (**C. D. s**).

When the dictionary presents the word with a capital letter and a hyphen (T-shirt, X-ray, etc), remove the hyphen and format in accordance with the spelled-out letter rule (**T. shirt**, **X. ray**, etc).

If the single letter forms part of the official spelling and is not separate by a hyphen (eBay, Xbox, etc), the spelled-out letter rule does not apply and you should transcribe as a complete word (**ebay**, **xbox**, etc).

This rule DOES apply to **O. K.** Do not use **okay**. It is acceptable to transcribe just **K.** if that is what is heard.

For both pronunciations, **zee** and **zed** should be transcribed as **Z.**

Acronyms that are conventionally written in all caps, but are pronounced like **regular words**, should be transcribed like regular words.

Examples:

- the U. S. A. has been part of **nato** since nineteen forty-nine
- he is without a doubt the best **fifa** player

Punctuation

Punctuation must **NEVER** be used, other than **apostrophes** and **hyphens**, where appropriate. You are not required to separate sentences with periods (full-stops) or to punctuate with commas, semi-colons or colons, exclamation, or question marks.

Examples:

- **Incorrect** - If I can get time off work**,** I plan to drive to Brighton on Thursday**.** The weather is expected to be good**!**
- **Correct** - if I can get time off work i plan to drive to brighton on thursday the weather is expected to be good

Spoken Punctuation, such as "**google dot com**" should always be transcribed as-is, and not with punctuation symbols.

Examples:

- **Incorrect** - **www.**google**.**com
- **Correct** - **W. W. W. dot** google **dot** com
- **Incorrect** - john**.**smith**@**email**.**com
- **Correct** - john **dot** smith **at** email **dot** com

Special attention should be paid to correct transcription of **hyphens**. Use hyphens for words where the hyphen provides **semantic value**. Written-together forms like "openfaced", which are not recognized by the MS-Word speller, should not be used. In these cases, a hyphen should be used.

Examples:

- an **open-faced** sandwich
- a **drive-by** shooting
- a **part-time** worker

Numbers

All **numbers** must always be spelled out. It is best practice to hyphenate compound numbers between twenty-one and ninety-nine (except 30, 40, 50, 60, 70, 80 and 90).

Examples:

- i was born on january seventh of **nineteen eighty-four**
- i purchased **one hundred and twenty-five** plants
- sarah celebrated her **twenty-fifth birthday** today
- call **nine one one**
- these ruins are over **one thousand and fifty** years old

O. vs. Oh vs. Zero

- The letter 'o' should be transcribed as **O.**
- If used as an exclamation, 'oh' should be transcribed as **ohh**
- The **number zero** when spoken as 'o' should be transcribed as **oh**.

Special Characters

Special characters, sometimes used for short-hand notation, should not be used in transcription.

The following are examples of **incorrect** transcriptions:

- coffee & tea
- i'll arrive @ five P. M.
- you are the #1

The following are examples of **correct** transcriptions:

- coffee **and** tea
- i'll arrive **at** five P. M.
- you are the **number** one

Named Entities

Names of people, companies, songs, countries, apps, etc. should all be transcribed in accordance with previous rules (e.g., use lower case). Copyrighted terms (e.g., trademarks, brand names, registered names) are also to be treated the same way.

Well-known people, companies, songs, apps, etc. should all be transcribed with the trademarks or official spellings that they are known by.

If you are unsure how to accurately spell one of these named entities, then perform a web search to identify the correct spelling.

Examples:

- OneDrive → **onedrive**

- McDonald's → **mcdonald's**
- PowerPoint → **powerpoint**
- Spider-Man → **spider-man**
- Coca-Cola → **coca-cola**
- Inglourious Basterds → **inglourious basterds**
- O Canada → **O. canada**
- Ice-T → **ice T.**

If the correct spelling is still not clear, then make your best judgment based on the context that you have.

Math Symbols

Math symbols must **NEVER** be used. Instead, the word associated with the symbol must be used.

Examples:

- 2 + 2 = 4 → two **plus** two equals four
- 6 / 3 = 2 → six divided **by three** equals two

Informal Words

If a speaker clearly uses an **informal** pronunciation that is **formally recognized** as a word, then use the informal word.

Examples:

- you **wanna** do that instead
- are they **gonna** come at the usual time
- i **kinda wanna** tell you something

If the informal word is not formally recognized as a word, then use the formal (official) word. Do not create new words.

Examples:

- **Full of** → **fulla**
- **Like a** → **lika**

The conjunction **'cause** is a common contraction of **because** and is the preferred spelling.

If the speaker explicitly says "**ha ha ha**" (which is different from legitimate laughing, which the guidelines consider as noise), "**boom**", or "**bang**", etc. transcribe them if, and only if, the dictionary considers it is a **formally recognized** word.

Use **yup** and **yeah** if this is what the speaker says. [The transcription yea (rhymes with day) is only used for the exclamation meaning woohoo].

Ungrammatical Words

For words which are **ungrammatical** in the given context, but are clearly articulated, transcribe them as spoken.

The following are examples of **correct** transcriptions:

- there you will find all my **document**
- **where** was the last time you called her
- **is** you completely sure about it

Please note the sentences above are **grammatically incorrect** given the context. Yet, if these words were spoken this way, these would still be considered **correct** transcriptions.

Ambiguity

Ambiguity can come in many forms:

- Homophones (words with same sound but different spelling)

Examples:

- hear / here
- bear / bare
- bee / be

- Variant Spellings (words with more than one acceptable spelling)

Examples:

- ax / axe
- doughnut / donut
- barbecue / barbeque

- Proper Nouns

Examples:

- vanessa / vanesa
- anthony / antony
- john / jon

For homophones, variant spellings and inflections, simply choose the form that you feel is most likely given the context.

For proper nouns, you may use the most commonly used variation.

Regional Spelling Variations

Be careful to observe the correct spelling conventions for the locale that you are working in. The exception would be when referring to a proper noun. For example, if working in the British locale, you would still transcribe **the kennedy center** because that is the official spelling.

Commonly Used Foreign Words

If the foreign word/language is part of the language's regular lexicon (i.e., would be understood by most speakers), write it in the foreign language using foreign language script, eg. it would be acceptable to transcribe "ciao" or "bonjour".

Accents or Diacritics in English

Use accents or diacritics in English for people, places, organisations/institutions, trademarks, and when commonly recognized by the dictionary.

Examples:

- Beyoncé → beyoncé
- Sebastián → san sebastián
- Dáil Éireann → dáil éireann
- Citroën → citroën
- Soufflé → soufflé

Side Searching / Contextual Audio

Use side searching to help with understanding context or identifying names of places, people, trademarks, etc. The contextual audio that is present in the webapp can also be very useful in improving understanding of the audio that you are transcribing. Surrounding audio is particularly useful for establishing whether or not a word is truncated at the beginning or end of the audio. If you are unsure, check surrounding audio.

Active Tags

<UNKNOWN/>

This tag is to be used in the following scenarios:

- ## Unintelligible Speech

Use this tag when there is obvious human speech, but one or more of the actual words cannot be determined. **Unintelligible speech** may be a result of a strong accent, quick speech, poor pronunciation, poor sound quality or any other reason.

For cases where part of the speech is intelligible and part is unintelligible, use the **<UNKNOWN/>** for the unintelligible portion and continue to transcribe the intelligible portions as usual.

- ## Foreign Speech

  - ### Partial-Utterance Foreign Speech

If part of the utterance is in a foreign language, then mark that part of the utterance as **<UNKNOWN/>** but transcribe the remaining words as usual.

  - ### Full-Utterance Foreign words

If the entire utterance is in a foreign language, use the **<UNKNOWN/>** tag to represent the entire utterance. But if the speaker mentions a foreign **named entity**, transcribe it.

- ## Truncated Speech

Use the **<UNKNOWN/>** tag when either or both the first word and/or the last word in an utterance is cut off (**truncated**) by the start or the ending of the recording.

Note: For cases where the entire utterance is **<UNKNOWN/>** (either unintelligible or in a foreign language), don't add any additional tags. Simply tag the entire utterance with a single **<UNKNOWN/>** tag.

Note: Do not use this tag when the speech is clear but ambiguous due to various reasons, such as homophones, variant spelling, etc.

Note: Do not use consecutive (uninterrupted) **<UNKNOWN/>** tags. If the speaker says several unintelligible words in a row, only one **<UNKNOWN/>** tag is required.

- **Mispronunciations**

For mispronunciations where the intent is <u>not</u> clear, the **<UNKNOWN**/> tag should be used.

For mispronunciations where the intent is clear, the standard word should be used. For example:

- los angeles is in **californa** (doesn't pronounce the 'i') → los angeles is in **california**
- open **misocroft** office (due to the context, we can infer the intent) → open **microsoft** office

<FILL/>

A **<FILL/>** tag must be applied for both words which are not fully spoken and uttered sounds which lack any semantic value.

- **Incomplete Words**

Use the **<FILL/>** tag when a word is damaged and not fully uttered by the speaker.

Examples:

- i would like a burger and a **sod**… actually just a burger   → I would like a burger and a **<FILL/>** actually just a burger
- please call me tomo… tomorrow → please call me **<FILL/>** tomorrow

- **Filler**

Use **<FILL/>** tag if the speech includes sounds without semantic value, such as **umm**, **amm**, **ehh**, etc.

Examples:

- **amm** what would you like for dinner → **<FILL/>** what would you like for dinner
- **umm** perhaps we should **umm** change our plans → **<FILL/>** perhaps we should **<FILL/>** change our plans

- ## Exclamations and Interjections

Exclamations and interjections such as **uh-huh** and **mm-hmm** are treated as words, not fillers, and should be transcribed. They do have semantic value and are therefore distinct from utterances such as **umm**, **amm**, **ehh**, etc.

Example:

- Speaker 1 - are you going to the beach → are you going to the beach
- Speaker 2 - **uh-huh** i'm **umm** leaving soon → **uh-huh** i'm **<FILL/>** leaving soon

In that example, **uh-huh** is an exclamation that is used to express assent. The same would be true if the speaker had said **mm-hmm**. In contrast, the **umm** is merely a vocal filler and should be tagged with the **<FILL/>** tag.

- ## Repetitions

Repetitions must not be tagged if the words are not damaged; simply transcribe the words as uttered by the speaker, regardless of the number of repetitions.

Examples:

- i i do not know what to to to say → **i i** do not know what **to to to** say
- are are are you sure about that → **are are are** you sure about that

Note: If the speaker uses official words as a filler (such as "like", "I mean"), do transcribe those words.

Note: Consecutive fillers should be transcribed with a single **<FILL/>** tag.

Example:

- i would like a **umm err** burger → i would like a **<FILL/>** burger

<CNOISE/>

If a **Continuous Noise** is present through the majority of the recording, at a volume around or above the speech's volume level, the **<CNOISE/>** tag must be placed at the **beginning** of the transcription.

Examples:

- Continuous human noise (laughing, singing, crying, etc.)
- Continuous ambient noise (rain, traffic, etc.)
- Continuous electronic noise

- Continuous background music
- Continuous mechanical sounds
- Continuous, unintelligible crowd noise.

Note: If a **Continuous Noise** is present in the recording but stays well below the speech's volume level through the majority of the recording, it must not be tagged.

- Singing

The following rules apply when **singing** is present in the recording:

Singing-only recordings: Recording includes only singing (or singing + music) = tag **<UNKNOWN/>**.

- A single **<UNKNOWN/>** tag may be applied to discard recordings entirely comprised of singing or unintelligible/truncated/foreign speech.

Background Singing: Recording includes speech + background singing = Transcribe speech + tag **<CNOISE/>** for the singing.

- **Background singing** must be indicated with a **<CNOISE/>** tag.

Intermittent Singing: Speaker changes from speech to singing (or vice versa) = Transcribe speech + tag **UNKNOWN** for the singing parts.

- **Intermittent singing** must be indicated by an **<UNKNOWN/>** tag, while the surrounding speech segments transcribed normally.

<SN/>

The **<SN/>** tag is required for a short duration noise which occurs at a volume that is around or above the speech's volume level.

Examples:

- Short duration human noise (cough, sneeze, loud breath, etc.)
- Short duration electronic noise (loud crackling, popping, etc.)
- Short duration mechanical sounds (doorbell, car horn, etc.)
- Short duration sound effects

- Isolated Sudden Noise

If an **isolated sudden noise** occurs independently of uttered words, then it should be tagged independently (**with spaces**) of the transcribed words:

Examples:

- he's about to ring the bell **<SN/>** did you hear it

- **<SN/>** there are too many **<SN/>**

- <u>Overlapped Sudden Noise</u>

If the sudden noise overlaps with a word, append the tag (with **no spaces**) to the end of the word.

Example:

- i will definitely**<SN/>** call you tomorrow

Note: If a **Sudden Noise** is present in the recording, but occurs well below the speech's volume level, it must not be tagged.

Note: Interjections and exclamations, such as "mmm", "mm-hmm" and "hmm", are not regarded as noise and should be transcribed.

Note: **<CNOISE/>** and **<SN/>** tags are not exclusive from each other and both may be applied in the same transcription as long as both of them are distinctively present and required.

Multiple Speakers and Speech Overlap Tag

Some recordings include two or more speakers interacting with each other or the recording device. In these scenarios, all speech must be transcribed within the same line, regardless of the number of speakers, and according to their position in time.

Example:

- **Speaker I : word1 word2 word3   word4 word5**
- **Speaker II** :   word1 word2
- **Speaker III**:   word1 word2 word3

Expected transcription:

word1 word2 word3 word1 word2 word4 word5 word1 word2 word3

Background Speech

Background Speech is defined as speech which was not intended as part of the recording but is still audible. All background utterances that are audible should be transcribed in accordance with the rules above. This includes speech captured from background speakers who are in the room, and background speech heard from a radio, television, etc.

Example:

- **Speaker I** :   **word1 word2**
- **Speaker II** :   **word1 word2 word3**
- **Background Speech**: **word1 word2**   **word3 word4**

Expected transcription:

word1 word2 word1 word2 word1 word2 word3 word3 word4

As noted above, when the background speech is inaudible and takes the form of ambient noise, use the **<CNOISE/>** tag.

<OVERLAP> <OVERLAP/>

Use the **<OVERLAP> <OVERLAP/>** tags to highlight a section of the recording which presents two or more speakers at the same time, including audible background speech (as defined above).

The **<OVERLAP> <OVERLAP/>** tags are also required for overlapping foreign language speech.

Unlike other tags, this pair of tags must go around the overlapped speech section, as shown below.

Independent Speech **<OVERLAP> Overlapped Speech <OVERLAP/>** Independent Speech

The louder, predominant, most notable speech will be transcribed, while the speech which goes behind the main speech will be ignored and not transcribed.

If no speaker appears to be louder or more important than the other, simply use your best judgment based on the context.

Example:

In this example, Speaker I's speech is **most prominent**:

Overlap

- **Speaker I** :   **word1 word2 word3 word4 word5**
- **Speaker II**:   **word1 word2 word3 word4 word5**

Expected transcription:

word1 word2 word3 <OVERLAP/> word4 word5 <OVERLAP/> word3 word4 word5

<UNKNOWN/>, **<FILL/>** and **<SN/>** tags are still applicable inside the **<OVERLAP> <OVERLAP/>** tags, so do insert them whenever necessary.

Example:

In this example, Speaker II's speech is **most prominent**:

- **Speaker I :** word1 word2 word3 word4 word5 word6 word7
- **Speaker II**: word1 word2 **<FILL/>** word3 **<UNKNOWN/>** word4

Expected transcription:

word1 **<OVERLAP/>** word1 word2 **<FILL/>** word3 **<UNKNOWN/>** **<OVERLAP/>** word4

The **<UNKNOWN/>** and **<FILL/>** are both regarded as speech and so the **<OVERLAP> <OVERLAP/>** tags should be used even when the only overlap is a **<FILL/>** or an **<UNKNOWN/>**

Example:

- **Speaker I :** word1 word2 word3 **<UNKNOWN/>** word6 word7 word8 word9
- **Speaker II**: **<UNKNOWN/>**

Expected transcription:

word1 word2 word3 **<OVERLAP/>** **<UNKNOWN/>** **<OVERLAP/>** word6 word7 word8 word9

Note: Only use the **<OVERLAP> <OVERLAP/>** tags where there is overlapping speech. If the speaker is interrupted by noise but no speech, you would use a noise tag but not overlap tags.

Note: There should be spaces between the tags and the speech.

Note: It is a common mistake for transcribers to transcribe all overlapping speech. As noted above, please **ONLY** transcribe the most prominent of the speakers.

Note: The **<OVERLAP> <OVERLAP/>** tags are **NOT** meant to be used to indicate speaker change.

Spelling Resources

Dictionaries

Locale	Dictionary Name	Dictionary URL
en-US	Lexico	**https://www.lexico.com**
en-GB	Lexico	**https://www.lexico.com**
en-US/GB	Merriam	**https://www.merriam-webster.com/**
en-US	Cambridge Dictionary (American English)	**https://dictionary.cambridge.org/dictionary/essential-american-english/**
en-US/GB	Dictionary.com	**https://www.dictionary.com/**
en-AU	Macquarie Dictionary	**https://www.macquariedictionary.com.au/**
en-AU	Preferred Australian English Spelling Dictionary	**https://www.australian-dictionary.com.au/**

English Spelling Variants

American and British English spelling differences
https://en.wikipedia.org/wiki/American_and_British_English_spelling_differences

Canadian, British and American Spelling – Comparisons
http://www.lukemastin.com/testing/spelling/cgi-bin/database.cgi?database=spelling