

3. Transcription Conventions

3.1. Characters and Special Symbols

Transcription of speech in Chinese (including Taiwan Mandarin and other dialects of Chinese) should include only traditional Chinese characters and a limited set of full-width Chinese punctuation marks (，。？！、， see [Section 3.9](#) for details on punctuations), in UTF-8 encoding.

Use the same limited set of special mark-up symbols outlined in the [\[en_US\] Transcription Guidelines version 3.0](#) document. Follow the conventions there also for transcribing speech in English and other non-Chinese languages. **IMPORTANT:** Don't use the full-width versions of the mark-up symbols (i.e. 《》，：，（（），【】).

A few examples relevant to Taiwan Mandarin:

- "我有¥100。" = "我有一百塊日幣。"
- "我有\$10." = "我有十元美金。"
- "6 + 6 = 12." = "六加六等於十二。"
- "我的電子郵件是 m-golden@amazon.com" = "我的電子郵件是<lang:English>M dash golden</lang:English>小老鼠 <lang:English> Amazon </lang:English> 點 <lang:English>com</lang:English>。"
 - If you heard @ pronounced as "小老鼠", then type in 小老鼠。
 - If you heard @ pronounced as "at", then type in <lang:English>at</lang:English>

3.2. Spelling and Grammar

Use standard traditional Chinese orthography to transcribe what the speaker says. Don't use phonetic spelling such as Bopomofo or Pinyin

3.2.2. Mispronounced Words

An example relevant to Taiwan Mandarin:

- "我覺得這個計畫很荒妙。" = "我覺得這個計畫很荒謬。"
- "他很羅唆。" = "他很囉唆。"

3.2.3. Non-Standard Usage

Examples relevant to Taiwan Mandarin:

- "他在家吃飯跟我。" = "他在家吃飯跟我。"
- "哎哟，我兩個書把掉下去了！" = "哎哟，我兩個書把掉下去了！"
- "想對喜歡的人告白說出愛老虎油？" = "想對喜歡的人告白說出愛老虎油？"

When in doubt, consult 教育部重編國語辭典: <http://dict.revised.moe.edu.tw/cbdic/> as well as Wiktionary: <https://en.wiktionary.org/>. To reference the names of song titles, movies, TV shows, brands, etc. <https://www.amazon.com/> and <http://google.com/>.

3.5. Contractions

In Taiwanese Mandarin, two or more syllables are commonly contracted into one in causal speech. Transcribe these contracted syllables using the full citation forms.

- "那樣(*pronounced 驕*)好嗎?" = "那樣好嗎?"
- "這樣(*pronounced 降*)子你就可以看電視了!" = "這樣子你就可以看電視了!"

3.6. Interjections

Transcribe the following interjections (嘆詞) in Taiwan Mandarin:

- | | | | | | | |
|-----|-----|------|-----|------|-----|-----|
| • 啊 | • 哈 | • 唉 | • 呸 | • 嗯 | • 噢 | • 吼 |
| • 咦 | • 嘻 | • 哎呀 | • 啐 | • 嗯哼 | • 喔 | • 蛤 |
| • 嘿 | • 呵 | • 哎喲 | • 哼 | | • 欸 | |
| • 嗨 | | | | | • 囉 | |
| • 哟 | | | | | | |

Note: The Mandarin interjection "嗯" is not to be confused with the interjection "#嗯". Use context to disambiguate the two different uses. See [Section 3.11.2](#) for the list of filler words.

3.9. Punctuation

Transcribe Chinese utterances using the following limited set of Chinese full-width (全型) punctuation marks as required by the grammar rules:

End Punctuations			
	Symbol	Unicode	Descriptions
Periods 句號	。	U+3002	Use to end a sentence or a thought. <ul style="list-style-type: none">• 那個城市很安全。
Exclamation Points 驚嘆號	！	U+FF01	Similar use to the English exclamation point. <ul style="list-style-type: none">• 那是我看過最大的南瓜！

			<ul style="list-style-type: none"> 我到底什麼時候才學得到教訓！
Question Marks 問號	？	U+FF1F	<p>Similar use to the English question mark.</p> <ul style="list-style-type: none"> 那不是很簡單嗎？ 你最懂了，對不對？
Sentence-Internal Punctuation			
Commas 逗號	，	U+FF0C	<p>Use to break up short sentences or phrases. Somewhat similar to the English comma (except that it cannot be used to list things).</p> <ul style="list-style-type: none"> To set off a direct address: <ul style="list-style-type: none"> 瑪麗安，仔細聽我說的話。 我打電話給你們，我的朋友們，只是要抱怨一下我老婆。 To break up compound and complex sentences: <ul style="list-style-type: none"> 我想加入你們，但是很可惜我那時段有課。 我和馬可斯不能去那個爵士音樂會，所以我們看電視上的。 To set off introductory words and phrases: <ul style="list-style-type: none"> 因此，他們取消了他們的旅行。 休息過後，小組繼續他們的會議。 Around parenthetical phrases: <ul style="list-style-type: none"> 紐約時報上的那篇報導，至少可以這麼說，是個震撼彈。 可以住到靠海的飯店，就像我們去年住的那間，就太好了。
Listing commas 頓號	、	U+3001	<p>Use to separate things on when enumerating. Similar to the English serial comma.</p> <ul style="list-style-type: none"> 這是我的狗、貓、魚和烏龜。 我喜歡跳傘、滑滑雪板和騎越野單車。
Interpunct 間隔號	·	U+2027	<p>Use to mark divisions in transliterated foreign names.</p>

			<ul style="list-style-type: none"> • 馬克・吐溫 (from English) • 努爾・白克力 (from Uyghur)
--	--	--	--

Don't use any other Chinese punctuations (e.g., 分號、引號等). Don't use half-width English punctuation marks in Chinese context and vice versa. If you are unsure of the punctuation rules for Taiwan Mandarin, consult 臺灣教育部《重訂標點符號手冊》修訂版 (https://language.moe.gov.tw/001/Upload/FILES/SITE_CONTENT/M0001/HAU/c2.htm).

When transcribing a language other than Mandarin, use punctuation symbols and rules that are appropriate for that language.

- 一杯<lang:English>iced latte? Okay, no problem</lang:English>! 沒問題。

Note: Some punctuation use is stylistic/subjective. Differences of opinion are not necessarily errors.

3.11. Disfluent Speech

3.11.2. Filler Words

For Taiwan Mandarin, transcribe only the following fillers, preceded by the hashtag:

- #呃
- #啊
- #嗯

Note: The Mandarin filler word "#嗯" is not to be confused with the interjection "嗯". Use context to disambiguate the two different uses. See [Section 3.6](#) for guidelines on interjections.

3.12. Overlapping Speech

3.12.1. Conversational Telephony

If the overlapping speech begins in the middle of the character, place the [overlap] tag before the character.

- "看起來是在煮什(*speech from an interferer*)錦飯。" = "看起來是在煮什[overlap]錦飯。"

3.14. Non-Target Languages

Speakers in Taiwan, especially older speakers, may code-switch between Mandarin, Taiwanese, and other dialects of Chinese (e.g. Hakka, Cantonese), sometimes in the same sentence. Switches to other Chinese dialects should be indicated with the foreign language tags (<lang:Foreign></lang:Foreign>) in the same ways as outlined elsewhere in the guidelines. Use the following language tags to bracket speech spoken in non-Taiwan Mandarin varieties of Chinese:

Tags	Varieties of Chinese
<lang:TSM></lang:TSM>	Taiwanese (i.e. Taiwan Southern Min)
<lang:Hakka></lang:Hakka>	Taiwan Hakka
<lang:OtherChinese></lang:OtherChinese>	All other varieties of Chinese (e.g. Cantonese, Shanghainese, etc)

Transcribe all varieties of Chinese using the standard orthography of Chinese, if the transcriptionist understands the variety and to the extent possible. Transcribe Otherwise, transcribe the non-target Chinese variety as ().

- "快下來，要吃飯(in Taiwanese)了喲！" = "快下來，要 <lang:TSM>吃飯</lang:TSM>了喲！"
- "那一家的蚵仔煎(in Taiwanese)特別好吃！" = "那一家的<lang:TSM>蚵仔煎</lang:TSM>特別好吃！"
 - If the speaker uses the Mandarin pronunciation of "蚵仔煎", then transcribe the utterance as "那一家的蚵仔煎特別好吃！"
- "林憶蓮唱的哪首廣東歌叫無心睡眠(in Cantonese) 。" = "林憶蓮唱了一首廣東歌叫<lang:OtherChinese>()</lang:OtherChinese> 。

Don't use the <lang:Foreign></lang:Foreign> tags around loanwords that have been grammaticalized and fully adopted into common use in Taiwan Mandarin. If it is unclear whether a word is a loanword or not, consult 教育部重編國語辭典: <http://dict.revised.moe.edu.tw/cbdic/> or <https://en.wiktionary.org/>. A word that is listed in these dictionaries is a strong ground to consider it an established loanword, even if it is of foreign origin.

- "她好像自己開了一個部落格。" = "她好像自己開了一個部落格。"
 - If the speaker uses the English pronunciation of "blog", then transcribe the utterance as "她好像自己開了一個<lang:English>blog</lang:English>。"

Examples involving using the <lang:Foreign></lang:Foreign> tags in conjunctions with other markup tags:

- "我在<lang:English><initial>IBM</initial></lang:English>工作。"

- <lang:Japanese>おひさ [overlap] </lang:Japanese> 。 "

3.15. Non-Speech

3.15.1. Non-Speech Noises

Don't insert a non-speech tag in the middle of a character. If a non-speech sound occurs in the middle of a character, add the tag exactly **before** the character in which it occurred.

- "我每天一定ㄟ-(ring)-ㄛ兩公升的水。" = "我每天一定[ring]喝兩公升的水。"