

N.B. | indicates a space

Spoken Language Transcription Symbols

Meaning	Symbols	Example of glosses	Example of utterance	Points to note
Non-vocalized words (including whispers) (To indicate words that are not vocalized or in whispers)	@nv	海豚@nv	媽媽有 海豚@nv . 媽媽有@nv 海豚@nv . 媽媽@nv 有@nv 海豚@nv .	(1) ‘@nv’ is marked at word level. If the entire sentence is without vocalization, mark ‘@nv’ after each word (2) The symbol also applied to the case that <u>entire or part</u> of the word without vocalization or in whispers e.g.1 only 海 in 海豚 is without vocalization, type 海豚@nv e.g.2 only 豚 in 海豚 is without vocalization, type 海豚@nv
Unintelligible words (Use ‘xxx’ when the transcriber cannot hear or understand what the speaker is saying)	xxx	xxx	妹妹 xxx aa3 ! xxx 妹妹 xxx 要 aa3 !	*Note the use of space Unintelligible words are marked with ‘xxx’. No phonetic transcriptions.
Unintelligible mouthing (To indicate any unintelligible words without vocalizations or in whispers)	xxx@nv	xxx@nv	妹妹 呢度 xxx@nv. 妹妹@nv 呢度@nv xxx@nv.	
Best guess (When the transcriber is uncertain about a word, s/he makes the best guess)	___ [?]	伯伯 [?]	呢度有 伯伯[?]. 呢度<有 伯伯>[?]. <呢度有 伯伯>[?].	*Note the use of space (1) Some old transcriptions use @w. Please change @w to [?]. e.g. 伯伯@w → 伯伯[?] (2) If two or more words are guessed by the transcriber, the group of words can be enclosed by angle brackets followed by the code [?].
Simple events (To indicate a variety of sounds produced by speakers, vocalizations that are not speech might be categorized into this type)	&=___	&=sings &=cries &=laughs &=coughs &=sneezes &=yawns &=sighs &=whistles &=yells (大叫) &=pants (喘氣) &=whines (扭計聲) &=belches (喝完汽水打嗝) &=gasps (倒抽一口氣) &=groans (呻吟聲) &=growls (咆哮) &=squeals (長而尖的叫聲) &=whimpers (啜泣, 嗚咽)	&=laughs 好得意! &=whines 俾我! 眼瞓 aa3 &=yawns. 走 aa3, &=yells.	(1) Simple events include laughing, crying, yelling, etc. (2) &=laughs can be used to represent all types of laughs, chuckles, and giggles (3) Refers to CHAT manual for more details (MacWhinney 2000, p. 59)
Imitations of sounds (To indicate speakers’ imitation of the sound of an	&=imit:___	&=imit:motor &=imit:pours_water &=imit:cooks	妹妹 煮 &=imit:cooks.	When two or more words are required to describe the imitations, use an underscore ‘_’ to link up the words

object or an action)				
Actions without speech (To indicate speakers' action that is not accompanied by speech.)	0 [=! meaning]	0 [=! raspberry] (伸脰) 0 [=! feed the horse with water]	*MOT: 你餵馬馬飲水. *CHI: 0 [=! feed the horse with water]	*Note the use of space (1) When the child responses with actions without any speech, mark the actions by the codes '0 [=! ...]' (2) Do not use the zero, if there is any speech on the tier.
Fillers	aa3 eh em uh hm mm um		aa3, 妹妹你睇吓呢度!	All the fillers should be typed
Shortenings (Use brackets to include the syllable which is shortened and not pronounced)	()	bi4(bi1) (bi4)bi1 (洗)白白	馬馬 bi4(bi1).	e.g. '洗' in '洗白白' is not pronounced, type '(洗)白白'
Interjections	@i	haa1 @i haa2 @i ei3 @i	haa1 @i 佢又記得 wo3. hei3jaa4 @i 坐翻好先.	All the interjections are marked with a postcode '@i'.
Onomatopoeia	@o	汪 @o 汪汪 @o 喵喵 @o 咯咯 @o ap @o	佢汪汪 @o 咁叫 gaa3. 有隻汪汪 @o.	All onomatopoeia are transcribed with the Chinese characters or Jyutping with a postcode '@o'
Motherese / Child-directed speech (When the speaker produces a lengthened word in order to let the child understand the pronunciation or the speaker intends to teach the child to speak, '@m' is used to mark the word)	@m	史迪仔 @m jyu5 @m	佢叫做史迪仔 @m. *CHI: 鴨鴨. *MOT: 唔係, jyu5 @m. *CHI: 羊.	(1) Sometimes the speaker intends to teach the child how to pronounce the correct syllable (including onset/coda). Mark the syllable with Jyutping with '@m'. (2) Reduplication of syllable in child's speech or some child-directed forms are not applicable to this symbol, e.g. 波波, 花花.
Family specific form (To mark lexicalized individual or family specific forms not used by any other people)	@f	bookbook @f	睇 bookbook @f.	e.g. bookbook @f refers to syu1 (書) in YC's speech
Cantonese words without Chinese characters (To mark words without a Chinese character with Jyutping and '@p')	@p	caau3 @p	佢自己識去 caau3 @p 玩具玩.	
English words (To mark English words in small letters and '@s:en')	@s:en	fish @s:en no @s:en	呢條叫 fish @s:en. no @s:en 唔要!	
Incomplete words (Mark '&' in front of an incomplete word)	&	&史	*MOT: &史. *CHI: &史. *MOT: &迪. *CHI: &迪. *MOT: &仔.	Sometimes the speaker would teach the child to pronounce some words/proper names. Use '&' and different lines to transcribe the data.

English proper names	(First capital letter)	Ruby Hello_kitty	*CHI: [&仔 . Ruby 姐姐 俾 . Hello_kitty wo3 !	(1) The first letter of the English proper name must be in capital letter. The rest of the letters are in small letters. (2) If more than one word is used to refer to the same entity, use an underscore to link up the words.
Proper name	(No space in between the characters /Jyutping)	踢躉小企鵝	佢最近先睇完套戲叫踢躉小企鵝 aa1maa3 .	Note the use of proper name which can be the name of person, movies, etc. No space should be in between the characters of the proper name.
Errors and substitution of words/syllables (To mark errors or substitutions by [*] together with the correct words or intended meaning)	[*]	皮皮 [* 叮噹] 叮噹皮皮 [* 叮噹]	睇 皮皮 [* 叮噹] . 有 叮噹皮皮 [* 叮噹] aa3 .	*Note the use of space The child may substitute some syllables with specific word. e.g. YC uses 叮噹皮皮 or 皮皮 to refer to the cartoon character 叮噹. Mark the intended meaning in brackets ‘[* ...]’.
Descriptions of words/syllables (To provide brief explanations on the text tier)	[= <u>meaning</u>]	sir@s:en [= police] <yes@s:en sir@s:en> [= police]		(1) The child may overgeneralized some words/phrases to mean different things. e.g. YC uses ‘yes sir’ to refer to “police”. A pair of square brackets ‘[= ...]’ are used to mark the meaning of those words/phrases. (2) Use angle brackets to indicate the phrase if more than one word represent the meaning, e.g. the entire phrase ‘yes sir’ represents “police”.
Repetition (To mark the number of repetition of words or syllables)	[x N]	糖 [x 7] 花 [x 3] <man1 走佢先> [x 2]	&=whine 糖 [x 7] ! 花 [x 3] 有好多 aa3 . 快啲 <man1 走佢先> [x 2] .	*Note the use of space (1) When the speaker repeats the entire phrase, use angle brackets to bracket the phrase and mark the number of repetition at the end. (2) Reduplication of syllable in child’s speech or some child-directed forms are not applicable to this symbol, e.g. 波波, 花花.
Delimiters	. ? ! +/. +... ,	係. 咩? 係,	妹妹食完 laa3 . 媽媽喺邊度 aa3 ? 唔制 aa3 ! 佢話 +/. 我想 +... 係, 佢講咗我知. &=whine 俾阿妹!	*Note the use of space ‘+/.’ is used to mark incomplete utterances that are being interrupted by another speaker whereas ‘+...’ is used to mark incomplete but not interrupted utterances. ‘,’ is used to mark pauses in the same utterance. *The symbol ‘#’ or ‘(.)’ are not used to indicate pauses.
SFP and SFP cluster	(Jyutping with tone)	laa1 gaa3 aa3 gaa3laa3bo3	你食埋先玩 laa1 . 媽媽會凍親 gaa3 .	(1) All the sentence final particles (hereafter, SFP) are transcribed by Jyutping. No Chinese characters

		aa1maa3	你食咗好多 gaa3laa3bo3. 婆婆眼瞓 aa1maa3.	should be used. (2) SFP clusters are formed by two to three SFP. No space should be used in between the SFP.
Imitation (To mark completely imitated utterances of the child)	[+imit]	睇書. [+imit]	*MOT: 睇書. *CHI: 睇書. [+imit]	[+imit] is used only when the whole utterance is imitated. This symbol should not be used to mark partial imitations.
Notes to other transcriber (Make notes with the code ‘zzz’ at the beginning of an utterance for any uncertain transcriptions)	zzz	zzz	*CHI: [zzz.ap4ap4@o].	

Notes to spoken language transcriptions

(1) Template

- A template of transcription ‘YC_template.cha’ is available in the folder:
 ACQ:\YC_Elan\templates\YC_template.cha

(2) Input system

- Chinese/Cantonese characters
 - Before working on the transcription, please make sure the computer has installed the ‘GRF Chinese font’ which contains the largest lexicon of Cantonese characters.
 - Use any Chinese/Cantonese characters input system, such as Cangjie, Simple, Pinyin, Jyutping, etc, to input the Chinese/Cantonese characters.
- Jyutping
 - If there are no available characters to be written for certain Cantonese words, use Jyutping romanization system to input the words.
 - Further information can be found in the website in the ‘bookmarks.html’. Open any browsers, such as IE or Mozilla Firefox. The homepage shows the ‘bookmarks.html’ which contains several url links about Jyutping input system and dictionaries for Jyutping.
- Other languages
 - No HKSL signs should be transcribed in CLAN.
 - All the spoken languages other than Cantonese, such as English and Mandarin, should be typed with the corresponding written system and marked with postcodes accordingly.
 e.g. English words should be marked with ‘@s:en’: yes@s:en sir@s:en
 Mandarin words should be marked with ‘@s:zh’: 你@s:zh 好@s:zh
 - Check the reference for coding in CHAT manual (MacWhinney 2000, p.25, 36-38)

(3) Utterance

- Each line starts with an asterisk and the ID of the speaker, followed by a colon and a tab (indicated by ␣).
 e.g. *RUB: 媽媽 喺 邊度 aa3 ?
- You can use the shortcuts by either (i) Select from menu ‘Tier’; or (ii) press ‘Ctrl’+‘1’, ‘Ctrl’+‘2’, etc. to insert the standard format.

(4) Spacing

- Every word are separated by a space .
 e.g. 媽媽 喺 邊度 aa3 ?
- Note the aspect markers, negators, classifiers, A-not-A question markers, etc. are all separated by a space.
 e.g. 你頭先食咗糖 laa3 .
 你唔玩呢個 aa4 ?
 有一隻貓 .
 你愛唔愛個玩具 aa3 ?
- No space is required in between the characters in the words such as ‘A 仔’, ‘AA 咗’, etc. (‘-仔’ is a nominal suffix. ‘AA 咗’ forms an adjective.)
 e.g. 隻鴨仔去咗邊 aa3 ?

支筆俾你寫到 gwat6gwat6 咁 la3wo3.

- No space is required in between the form of motherese or any addressing.
 e.g. 波波, 花花, 杯杯, etc.
 叔叔, 伯伯, 哥哥, etc.

(5) Sentence final particles (SFP) and clusters of sentence final particles

- All the sentence final particles (hereafter, SFP) are transcribed by Jyutping. No Chinese characters should be used
 e.g. *RUB: 媽媽 喺 邊度 aa3?
- SFP clusters are formed by two or more SFPs. No space should be used in between the SFP.
 e.g. *RUB: 妹妹 識得 邊度 aa1maa3?
 *RUB: 你 食咗 好多 gaa3laa3bo3.

(6) Motherese

- Motherese refers to the speech directed to children used by the adult speaker (usually mothers).
 e.g. 波波, 花花, 杯杯, etc.
 Transcribe the words in Chinese characters or Jyutping accordingly without any spacing in between the words/Jyutping.
No postcodes are required.
- When the speaker directs his/her speech slowly to the child in order to let the child to understand the pronunciation or the speaker intends to teach the child to speak, ‘@m’ is used to mark the word.
 e.g. 史迪仔@m
- Sometimes the speaker intends to teach the child how to pronounce the correct syllable (including onset/coda). Mark the syllable with Jyutping with ‘@m’.
 e.g. *CHI: 鳴鳴.
 *MOT: 唔係, jyu5@m.
 *CHI: 羊

Grammatical categories

Tagging at mor tier	Grammatical categories	Examples
adj	adjective	靚, gwat6gwat6 咁
adv	adverb	慢慢, 先
asp	aspectual marker	咗, 翻, 住
aux	auxiliary/modal verb	得, 可以
cl	classifier	個, 隻, gau6
conj	connective	但, 同, zi3
det	determiner	呢, 個
co	expressive utterance	ai1jaa3
gen	genitive marker	嘅
loc	localizer	上面
pro	pronoun	我, 你, 佢
n	noun	蘋果, 枱
n:prop	proper noun	情情, Ruby
neg	negative morpheme	唔, 未
on	onomatopoeic expression	汪汪@o
prep	preposition	喺, 比
quant	quantifier	每
pro	reflexive pronoun	自己
Sfp	sentence-final particle	aa3, laa1, gaa3, bo3
v	verb	俾, 食, 洗
v:dir	directional verb	去, 嚟, 返
wh	wh phrases	邊個, 喺邊, 邊度, 做咩, 乜嘢

Common errors

- Please check the table ‘lexicon_YC_Cantonese_2009may05.xls’ for common errors and choice of glosses:
 ACQ:\Data Checking\Gloss\Cantonese\lexicon_YC_Cantonese_2009may05.xls
- Please check the reference ‘YC_reference.doc’ for some remarks on YC’s family
 ACQ:\Data Transcription\Reference\YC_materials_log\YC_reference.doc
- Please check the ‘log book’ of YC’s performance and specific words used by YC in each recording session

Common errors	Error types	Right examples	Wrong examples	Remarks
Choice of glosses	Meaning gloss instead of phonetic transcription	八達通	八 pu4 pu2 [% octopus]	
	SFP vs. SFP clusters	gaa3laa3bo3	gaa3 laa3 bo3 嫁喇噃 嫁 喇 噃	(1) Jyutping only (2) No space in between SFP clusters
	Chinese characters vs. Jyutping (Specific Chinese characters or Jyutping are chosen for specific grammatical categories)	返 翻	番 返	(1) 返 /faan1/ <v> e.g. 返學, 返埋嚟 (2) 翻 /faan1/ <asp> e.g. 開翻電視
		下 吓 haa2@i	吓 吓 haa3	(1) 下 /haa4/ <det^cl^co^loc> e.g. 掂一下佢試吓 (2) 吓 /haa5/ <asp> e.g. 俾媽媽睇吓 laa1 (3) haa2@i <co> e.g. haa2@i , 咩話 ?
Symbols	Underscore	&=imit:pour water Hello_kitty	&=imit:pour water Hello-kitty	Underscore (no hyphens) are used to link up the words which form one gloss/word.
		bookbook@f Hello_kitty	Bookbook@f hello_kitty	
	Capitalization of English letters	bookbook@f Hello_kitty	Bookbook@f hello_kitty	Capital letters are used for (i) proper names in spoken languages (ii) signs in sign languages
	Lexicalized Cantonese words	bi4bi1	bb BB bibi bb@s:en BB@s:en baby@s:en	Some English (or any other languages) words are lexicalized to Cantonese (with tone contour). Please use Chinese characters / Jyutping to transcribe the data instead of marking those words with English words (with postcode '@s:en')
	Family specific form	bookbook@f (book)book@f	book@s:en bookbook@s:en book@s:en [x 2]	
	Space	大個 大個	大個 大個	(1) 大個 <adj + cl> "big" e.g. 咁大個圈 (2) 大個 <adj> "grow up" e.g. 阿妹大個咗
		大隻 大隻	大隻 大隻	(1) 大隻 <adj + cl> "big" e.g. 咁大隻眼睛. (2) 大隻 <adj> "strong/muscular" e.g. 佢睇落有好多肌肉. 好似好大隻.
Jyutping	Onset	ai3jaa1 ai3jaa3	ai1yaa1 ai1yaa3	There are no onset 'y' in Jyutping.
	Tone	ap4@o	ap8@o	No tone '7', '8', '9' in Jyutping

Reference

MacWhinney, B. 2000. The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.