

Transcription Notation

	TYPE	CONVENTION	EXAMPLE
1.	Numbers	full form	today is first jan twenty eighteen <s/> total
	Symbols	full form	is eleven dollars eighteen cents <s/> cool <s/>
2.	Titles	full form (unless abbreviated)	doctor li-ann met mister muhamad-nor-hisham <c/> miss (um) madam aarti and professor ang at this year's meet up <s/>
3.	Acronyms	underscore '_'	he wanna study in n_a_f_a <c/> but couldn't
	Contractions	as is (unless unabbreviated)	get in <s/> so he's now gonna join n_t_u to study l_m_s <c/> linguistics i mean <s/>
4.	Multi-word nouns	hyphen '-'	you just came back from hong-kong <c/> right <s/> are you now staying in ang-mo-kio <c/> near s_t-engineering there <s/>
5.	Discourse particles	square bracket '[...]'	[oh] he still wanna eat [ah] <s/> [wah] why he like that [one] <s/> damn greedy [lah] he <s/>
6.	Fillers	round bracket '(...)'	(uh) yah <c/> it's near h_s_s <s/> to get there <c/> (um) i think you can turn left and (er) actually turn right and walk straight <s/> you should reach in (erm) five minutes <s/>
7.	Interjections	Exclamation mark '!...!'	!walao! why the hell do we need to study English [right]
8.	Paralinguistic Phenomena	(ppb) breath (ppc) cough (ppl) laugh (ppo) others	(ppc) (um) there was once i confidently strut down the catwalk <c/> as though i'm (ppo) naomi-campbell <s/> then i (ppb) trip over a stone and (ppl) fell flat on my face <s/> (ppl)
9.	Other languages	hashtag '#...#'	she went #pasar malam# to buy #roti-john# but #tak sedap# <s/> lucky #muah-chee# #shiok# <c/> else she'll be #pek chek# <s/>
10.	Unclear words	<UNK>	i went to <UNK> <c/> (er) or was it (erm) <UNK> <s/> i can't remember the name <s/>
11.	Incomplete words	tilde '~'	[oh] i don't understand the abbre~ abbrev~ (uh) abbreviation that they use nowadays <s/>
12.	Short pause (longer than 1000 ms)	<S>	
13.	Invalid	<Z>	

14.	Long-running Non-English utterances	<NEN>	[eh] speaking of which right okay <NEN> we'll just talk okay
-----	---	-------	---

Additional notation not yet addressed:

	TYPE	CONVENTION	REMARKS
1.	Fillers	<FIL/>	equivalent to [xxx]
2.	Speaker Noise	<SPK/>	equivalent to ppb, ppc, ppl and ppo
3.	Unknown	**	equivalent to <UNK>
4.	Non-primary speaker sound	<NON/>	background sounds, including sounds made by other speaker, background noise etc.

Transcription Conventions

Punctuations

- Boundaries must be marked at sentence level, down to the last phoneme.
- Conventions:
 1. End of sentence <s/> e.g. '.', '!', '?'
 2. Comma <c/> e.g. ','
- Example:

i'm fine <c/> thanks <s/> how are you <s/> i haven't seen you in ages <s/> if you drop by jurong again <c/> don't forget to ring me up <s/> we can chill <c/> go shopping <c/> and catch up over a cup of coffee <s/>

Numbers and Symbols

- Numbers and symbols must be represented in words.
- Conventions:
 1. Numbers **full form** e.g. thirtieth, twenty eighteen
 2. Symbols **full form** e.g. dollars, percent, plus
- Example:

their **tenth** anniversary concert will be held on **fourth** july **two zero one eight** <s/> around **three hundred fifty** students bought the tickets with a **twenty percent** discount <s/> [woah] that's like **minus ninety dollars** from the original price <s/>

Titles

- Titles must be represented in their full form (unless abbreviated in the audio).

- Conventions:

1. Titles	full form	e.g. mister, missus, professor
-----------	------------------	--------------------------------

- Example:

yesterday <c/> i watched **doctor**-doolittle again on netflix <c/> and then met up **professor** james-tan and **missus** siti-maisarah at coffee-bean around noon <s/> what a coincidence <c/> **mister** low and **miss** chan happened to be there as well <s/> they were celebrating **mister** low's promotion to an **associate prof** <s/>

(*note: in this case, *associate prof* is transcribed as such, since *professor* is being abbreviated to *prof* in the audio)

Abbreviations

- Abbreviations must be represented in their abbreviated form (unless unabbreviated in the audio).

- Conventions:

1. Acronyms	underscore '_'	e.g. m_o_e, t_v, u_k, h_and_m
2. Contractions	abbreviated form	e.g. you're, wanna, cause

- Example:

besides **n_a_p_f_a** <c/> i'm okay with national-service [lah] <s/> after **n_s** <c/> i literally just watch **t_v** all day <c/> every day <c/> while waiting to enter **n_t_u** <s/> [oh] i did travel to **u_s** before that <s/> i went new-york <c/> **n_y_c** <s/> i visited times-square <s/> but **didn't** manage to go central-park **cause** too rushing <s/> i also went to los-angeles <c/> my cousins are staying in **l_a** <c/> so they showed me around their **u_c_l_a** campus <s/> **they're gonna** come to singapore next week <s/>

(*note: in this case, *national-service*, *new-york* and *los-angeles* are transcribed as such, since they are unabbreviated in the audio)

Multi-Word Nouns

- Multi-word nouns must be connected by hyphen.
- Conventions:
 1. Multi-word nouns **hyphen '-'** e.g. south-korea, one-m_d_b, changi-airport, coffee-bean
- Example:

to get to **suntec-city** <c/> take the m_r_t on the **east-west-line** heading towards **pasir-ris** and alight at **city-hall** <s/> after exiting the station <c/> don't take the **raffles-city-shopping-centre** exit <s/> instead <c/> walk via the underpass at **citylink-mall** and you should reach there in like ten minutes <s/> you can also take the **circle-line** or **downtown-line** and alight at promenade or esplanade <s/> [oh] you should check out the **o_s_g-bar-kitchen** <s/> they've good food there <s/>

Discourse Particles

- Discourse particles, or pragmatic markers which directs the flow of conversation without adding significant paraphrasable meaning to the discourse, must be written between square brackets.
- Conventions:
 1. Discourse particles **square bracket** e.g. [oh], [oo], [wah], [lah],
 '[...]' [eh], [liao], [lor], [sia], [one]
- Example:

[woah] the place was freaking huge **[lah]** <s/> #siao# **[liao]** **[lor]** <c/> can #mati# **[sia]** <s/> w_t_h <c/> there's no way anyone can finish in one day **[man]** <c/> unless he some kind of superman **[ah]** <s/> **[eh]** wait <c/> i think my brother confirm can **[one]** <c/> cause he likes this kinda arty farty things <s/>

Fillers

- Fillers, or meaningless sounds that marks a pause or hesitation in speech, must be written between round brackets.

- Conventions:

1. Fillers	round bracket '(...)	e.g. (um), (uh), (erm), (er), (hm)
------------	--------------------------------	---------------------------------------

- Example:

(um) last year [ah] <s/> where did i go last year <s/> **(er) (hm)** so hard to remember <c/> so old already [lah] me <s/> (ppl) **(erm)** if i'm not wrong <c/> i went to **(er)** south-korea with my mum to celebrate my (ppb) twenty sixth birthday <s/> [eh] wait <c/> or was it twenty fifth <s/> **(er)** nineteen eighty nine <c/> now is twenty eighteen <s/> **(uh)** sorry [man] <c/> my maths damn #rabak# <s/> (ppl)

Interjections

- Interjections, or words added to a sentence to convey an emotion or a sentiment such as surprise, disgust, joy, excitement, or enthusiasm, must be written between 2 exclamation marks.

- Conventions:

1. Interventions	Exclamation marks '!...!'	e.g !wow!, !aiyo!, !wah!
------------------	-------------------------------------	--------------------------

Paralinguistic Phenomena

- Paralinguistic phenomena, or non-speech sounds like breathing, coughing and laughing, must be written between round brackets.

- Conventions:

1. Paralinguistic phenomena	(ppb) (ppc) (ppl) (ppo)	breathing coughing laughing other PP noises
-----------------------------	--	--

- Example:
(ppo) so he went home <c/> thinking that <c/> **(ppc)** sorry <c/> thinking that the long day was over <s/> then he was just gonna step into the house <c/> **(ppb)** his phone rang <s/> the hospital asked him to come back **(ppc)** cause of an emergency <s/> so poor thing he <s/> **(ppl)** (er) no choice [lor] <s/> life of a doctor <s/> **(ppb)**

Other languages

- Words from other languages must be written between hashtags.
- Conventions:
 1. Other languages **hashtag '#...#'** e.g. #bo pakai#, #kiasu#, #kueh-lapis#, #roti-prata#
- Example:
 singapore slang <s/> (hm) let me think <s/> first <c/> you confirm hear this before <c/> singaporeans so **#kiasu#** <s/> yah <c/> sometimes join queue also don't know for what <s/> (ppl) then there's **#shiok#** <s/> you can use it when you **#makan#** something delicious <s/> now even got **#shioklicious#** [lah] <s/> (ppl) then if you cannot **#tahan#** <c/> i mean cannot tolerate someone <c/> cause he so **#bodoh#** or something <c/> then you say (uh) **#buay tahan#** <s/> (erm) what else <s/> (ppb) [oh] for food and drink <c> there's chicken-rice <c/> **#wanton-noodle#** <c/> **#nasi-lemak#** <c/> **#roti-prata#** <c/> **#ice-kacang#** and (um) **#teh-tarik#** <s/>

Unclear Words

- Unclear words, or words that transcribers are unsure of, must be included in the transcription.
- Conventions:
 1. Unclear words **<unk>**
- Example:
 she said when she was there <c/> she visited the **<unk>** <c/> or was it **<unk>** <s/> (hm) i'm not sure what the name is <c/> but anyway <c/> i heard it's (er) a really pretty place <c/> it's like (uh) a go to place when you visit that city <c/> yah <s/>

Incomplete Words

- Incomplete words must be written as verbatim.
- Conventions:
 1. Incomplete words **tilde ‘~’** e.g. idiosync~, disambig~
- Example:

mary is **ambix~ ambidextr~** (um) ambidextrous is it <s/> yah <c/>
ambidextrous <c/> can use both right and left hands to write <s/> [oh] [oh] but
right <c/> i heard that this type of people more likely to be **schiz~** (er) **~phrenic**
<c/> what’s that word <c/> schizophrenic [ah] <s/> yah <s/> o_m_g <c/> that’s
scary [siot] <s/>

Short Pause

- Short pauses, or periods of silence that are longer than 1000 ms must be marked.
Always an entire interval.
 1. Short pause **<S>**

Invalid part

- Invalid parts of the speech must be marked. Always an entire interval. Invalid parts include:
 - Invalid speech
 - Noise
 - Non-primary speaker’s voice
 - Sounds from the monitor/speaker
 - Continuous noise eg. long laughs, music, singing etc.
- 1. Invalid part **<Z>**

Run-on Non-English Utterances

- Long utterances that are not in English can be tagged as <NEN>.
- **Do not mistake for #...#**, this notation is purely for singular words or short phrases of other languages that can be transcribed immediately.
- Will be transcribed at a later phase.
- Example: ya <NEN> but basic like <NEN> basic <NEN>

