<u>On transcribing sound track in Cantonese</u>

By Sam Wong

0. Format of transcription

    A sample transcription is shown as follows:-

聽講	上頭	而家	派	咗
teng1gong2	*soeng6tau4*	*ji4gaa1*	*paai3*	*ʒo2*
hear.say	superior	now	dispatch	PFV

嗰個	姓	陳	嘅	特派員。
go2go3	*sing3*	*can4*	*ge3*	*dak6paai3jyun4*
that.one	surname	*person.name*	ATTR	special.investigator

    'It is said that the superior has currently dispatched the special investigator, Chan.'

    There are four lines:-
    Line 1: Chinese character and punctuation
    Line 2: romanization
    Line 3: gloss
    Line 4: translation

    To efficiently maintain a tidy text alignment, a table without gridlines is recommended to store the contents of the first three lines. Tabs and spaces, however, also serve the purpose but work with less efficiency. It should be noted that all words are left aligned within the same column. If there is insufficient space to put a long sentence on a single line, the split should be made between phrases that are more loosely related. For example, in the above example, a split made between *ʒo2* and *go2go3* is better than one made between *paai3* and *ʒo2* since the relation between the latter pair is more closely related than the former pair.

    In addition to the four lines of transcription, background information of the sound track like the date, venue, speaker(s) and short description of the event should be clearly stated at the beginning of the file.

    The purpose of each line is described in detail in the subsequent sections:

1. Chinese character and punctuation

    Line one is a rendering of the utterances in the sound track in Chinese characters and punctuations. In this assignment, fine transcription is required, so please jot down exactly what was uttered by the speaker, including the sentence final particles, like gaa3

喫; laa3 嘑, and bo3 嚩, as well as the exclamatives, like *waa3* 嘩, and *ai1jaa3* 哎吔, in addition to the content words.

1.1. Words and phrases

In common practice, the text in Chinese characters is always written without natural delimiters. On the other hand, it should be noted that in the course of transcription, in general, **word**, instead of character, is treated as a unit while a space is used **between words** as a delimiter.

In linguistics, *word* is usually defined as the biggest element that may be uttered in isolation with semantic or pragmatic content. For instance, for the word *bo1lei1* 玻璃 'glass', the component characters, *bo1* 玻 and *lei1* 璃, are meaningless by itself in modern Cantonese, *bo1lei1* is thus treated as a word and is written without space in between.

It should be noted that sometimes there is ambiguity in that a linguistic expression can be interpreted as either a phrase or a word. For example, the term *sik6 faan6* 食飯 can be interpreted as both 'to eat rice' or 'to have meal'. In the former case, the meaning of the expression is closely related to its components, the verb *sik6* 'to eat' and the object *faan6* 'rice'. In this case, *sik6 faan6* is considered as a verb phrase and a space is used between the two components. In the latter case, when *sik6faan6* is used with the meaning 'to have a meal,' in the sense that we not only 'eat rice', but also 'drink soup and eat noodle, dessert, et cetera,' *sik6faan6* is considered as a word and no delimiter is required.

It should also be noted that this kind of ambiguity is always found in natural language. There is no rigid rules on how to resolve such ambiguities so one will need to decide on whether a linguistic expression is a phrase or a word using his/her linguistic sense based on his/her understanding of the context as a native speaker. Most of the rules in the following guideline, which is designed for Mandarin Chinese, also apply to Cantonese and can serve as a general reference:

http://www.pinyin.info/rules/pinyinrules.html

For details, one can also refer to Yin and Felly (1990). Like any rules, the above rules, however, should not be followed blindly. The reader can adjust according to his/her linguistic sense as a native speaker.

1.2. Written form

Mandarin cognates exist for most of the Cantonese morphemes so in most cases, it is not hard to locate suitable characters for transcription purpose. Sometimes the etymology of a Cantonese morpheme is not clear and we are thus not aware of its Mandarin cognate. In the event of this, one can follow the common practice in daily life

experience like those found in newspapers, magazines, blogs, and other internet resources. For instance, *bin1dou6* 'where' is usually written by using the homophones 邊度. Upon necessary, non-Chinese characters can also be used in case an expression is commonly written in that way, especially the loanwords. Some examples are shown in Table 1.

Romanization	Common written form	English equivalent
cok3joeng2	chok 樣	the look when you act cool
ou1kei1	OK	okay
kaa6waai1ji4	可愛い	cute; lovely
	kawaii	

Table 1 Cantonese expressions commonly written with non-Chinese characters

For expression with more than one common written forms, like *kaa6waai1ji4* is commonly written as either '可愛い' or 'kawaii', one can select according to his/her own habit but consistency should be maintained through the whole work of transcription.

Sometimes cognate in Mandarin does not exist for a Cantonese morpheme but cognate attested in classical Chinese text does, like 擢樣 for *cok3joeng2*. In this case, character with etymological relation with classical Chinese, which is also known as the 'correct character', exists although many of them are hard characters. You are encouraged to check these characters from dictionaries but this should be done only when time is permitted and in any case should **not** be the focus of the transcription work. Some references are listed in the reference section.

It is always easy to locate a suitable written form for lexical word but for function word, it is not the case. The *utterance particles* are the hardest among all since one of the greatest contrast between Cantonese and Mandarin Chinese is the use of this category. There are only 27 utterance particles in Modern Standard Chinese (Chao 1982: 394−403) but as many as 95 in modern Cantonese (Leung 2005). For this reason, Mandarin cognates often not exist for most of these particles. Suitable characters are thus hard to found to represent these particles. Appendix 1 can be served as a general reference but the readers are strongly suggested to follow his/her own habit to avoid inconsistency.

Some operating systems may lack the necessary font or input method for inputting the special characters, one can download these tools by following this link:
http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/download_area/

Last but not the least, in the case that there is really no suitable written form in your mind at all, or you even do not understand the meaning of the expression; you can just directly put the romanization in place of.

2.   Romanization

The second line is a rendering of the utterances in the sound track in romanization. Like what we have mentioned in section 1, word is also treated as a unit when romanizing the utterances. In other words, space is not used between syllables within a word. Italic font-face is often used.

2.1. The LSHK system

In this assignment, the *Linguistic Society of Hong Kong Cantonese Romanization Scheme*, *a.k.a.* the Jyutping 粵拼 system, developed in 1993 is adopted. To understand this scheme, the readers are referred to LSHK (2002) and the following web-sites:

The Jyutping Scheme: http://www.lshk.org/node/47
Tutorials on Jyutping:
   http://www.cantonese.asia/viewnews-229.html
   http://www.iso10646hk.net/jp/learning/index.jsp
   http://www.senseasy.net/leeyuiwah/CHS/Jyutping-tutorial.latest.ppt

In addition to *Guide to LSHK Cantonese Romanization of Chinese Characters* (LSHK 2002), one can also check the romanization of a Chinese character via the following databases:
http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/ (for BIG5 characters only)
http://www.iso10646hk.net/jp/database/index.jsp

For those who are familiar with other schemes, they can refer to the charts comparing the schemes in LSHK (2002: 17−20) and the following web-sites:
http://input.foruto.com/ccc/jyt/ap01b.htm
http://en.wikipedia.org/wiki/Hong_Kong_Government_Cantonese_Romanisation

For those who have never undergone training in phonetics and phonology, nor have not learnt any romanization scheme in the past, it may be difficult to get familiarized with the scheme within a short time. The following web-sites that convert string of characters to romanization are useful resources:

Chinese Word Parser:
http://www.cantonese.sheik.co.uk/scripts/parse_chinese.php?action=parse
Jyutping Database: http://www.iso10646hk.net/jp/database/index.jsp
JyutPingEasy.Net: http://www.jyutpingeasy.net/scgi-bin/toJyutPing.cgi
HKTV Cantonese to Jyutping: http://hktv.cc/hp/cantonesetojyutping/

Although the above are helpful resources, one should always bear in mind that since it is common for a Chinese character to possess more than one pronunciation, the users are strongly advised to carefully check the computer outputs before utilizing.

2.2. Actual or standard pronunciation?

Living language is always a dynamic system. There are sometimes variations in pronunciation from person to person within a community. For instance, in modern Hong Kong Cantonese, virtually /n/ is missed among the initials in the phonological system while a number of young speakers have lost the /ng/ and /k/ coda in their speech. These two phonological developments are the so-called *sloppy speech*. Specifically, there also exists free variation for the same word in identical context. The same person may even choose different pronunciations in two consecutive utterances, *e.g. hung4dau2bing1 ~ hung4dau6bing1* 'shaven ice with red bean.'

For this reason, the actual rendering of the same word may be different within a passage of transcription. To provide more information about the variety of the speaker, actual rendering of the pronunciation is preferred. However, transcribing in the standard variety or the transcriber's own idiolect is also acceptable.

Most of the syllables in modern standard Cantonese can be transcribed with the LSHK system. In case you find some escaped from the net, the IPA system can be used for these syllables.

3. Gloss

To help non-native speakers understand better the literal meaning and syntactic properties like word order of the utterance, an English word-by-word rendering of the utterances in the sound track is provided in line three. For translation with more than one English words, a period is inserted between the words, like 'have.meal' for *sik6faan6* and 'special.investigator' for *dak6paai3jyun4*.

For content words, it is always easy to find an English equivalent of the Cantonese words. For function words, however, it is often hard to find equivalent because different syntactic system are often found in different languages while polysemy is frequently observed for function words.

For the former problem, for example, in Cantonese, classifier must be used between numeral and noun but in English, there is no such category while measure words are optional. It is often not easy to find an English equivalent of the classifier while the translation of the classifier always does not help the reader understand the text better since the classifier system only reflects how we classify the objects in the world, which is somewhat similar to *gender* in the European languages.

For the latter problem, for instance, the word *you* in English is both a *second person singular pronoun* and a *second person plural pronoun*. If we put *you* as the gloss of both *nei5* 'the second person singular pronoun' and *nei5dei6* 'the second person plural pronoun,' those

who have inferior knowledge of Cantonese will have no way to tell the different meanings between the two.

For these reasons, for function words in some special categories, special abbreviations in small capitals are usually adopted in place of the English equivalent. The option of *small capital* is available in Format > Font (Figure 1). Table 2 shows some abbreviations commonly adopted among linguists.
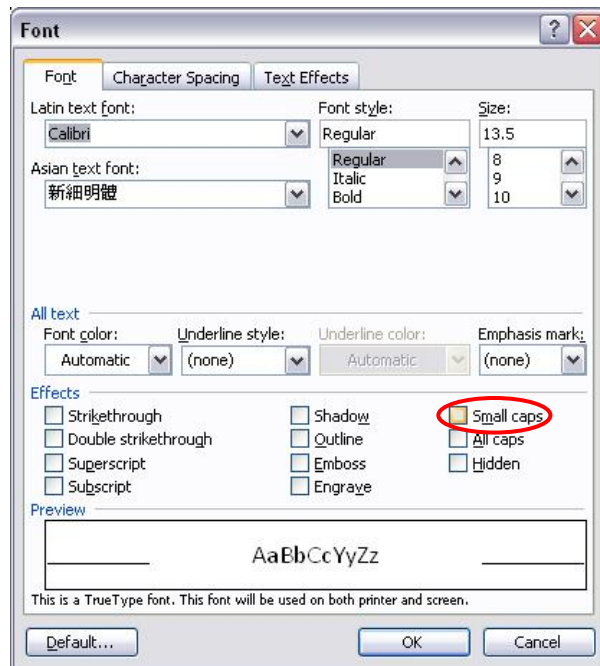

Figure 1 The option of *small capital* is available in Format > Font

1.	1SG	first person singular pronoun
E.g. 我 *ngo5*		
2.	1PL	first person plural pronoun
E.g. 我哋 *ngo5dei6*		
3.	2SG	second person singular pronoun
E.g. 你 *nei5*		
4.	2PL	second person plural pronoun
E.g. 你哋 *nei5dei6*		
5.	3SG	third person singular pronoun
E.g. 佢 *keoi5*		
6.	3PL	third person plural pronoun
E.g. 佢哋 *keoi5dei6*		
7.	ASP	aspect marker
Can be further divided into different aspects: (a) CONT: Continuous aspect *E.g.* 佢一邊食住 *zyu6* 個包，一邊等你。 (b) EXP: Experiential aspect *E.g.* 我試過 *gwo3* 搵佢啦，但搵唔到咋嘛。 (c)PFV: Perfective aspect		

E.g. 你做**咗 *zo2*** 功課未啊?		
(d) PROG: Progressive aspect		
E.g. 佢食**緊 *gan2*** 飯啊，你一陣再打嚟啦。		
(e) HAB: Habitual aspect		
E.g. 佢睇**開 *hoi1*** 中醫嘅。		
For other aspects, ASP can be used.		
8.	ATTR	attributive
E.g. 我**的 *dik1*** 書、小明**嘅 *ge3*** 書		
9.	BA	pre-transitive construction
E.g. 你**將 *zoeng1*** 本書擺喺枱度。		
10.	CL	classifier
E.g. 一**件 *gin6*** 衫、一**隻 *zek3*** 牛		
11.	COP	copular
E.g. 小明**係 *hai6*** 男仔嚟㗎。		
12.	EXCL	exclamative
E.g. 呢件衫真係勁靚**呀 *aa3***！		
13.	INTJ	interjection
E.g. **嘩 *waa3***！但係點解唔可以係法國呢？		
Others like: 嗱 naa4、哎吔 aai1jaa2		
14.	NEG	negation
E.g. 呢個**唔 *m4*** 係紅色係黃色嚟㗎！		
15.	NMZ	nominalization
E.g. 食**嘅 *ge3***、用**嘅 *ge3***，呢度咩都有。		
16.	PASS	passive marker
E.g. 條魚**畀 *bei2*** 隻貓食咗呀！		
17.	ACCU	accusative marker
E.g. 你**令 *ling6*** 佢好傷心呀！		
18.	FUT	future marker
E.g. 我**會 *wui5*** 做功課㗎嘞！		
19.	PRT	particle
For particles appearing in non-clausal-final position *E.g.* 你**呀 *aa3***，唔好再等喇！		
20.	Q	question particle
E.g. 呢條數到底點計**呢 *ne1***？		
21.	SFP	sentence final particle
* For all other particles appearing in clausal-final position *E.g.* 你啊，唔好再等**嘞 *laa3***！		
Others like: 喎 wo3, 噃 bo3, 呀 aa3, 吖 aa1, 喇 laa3, 嘅 ge3, 呃 a3, 嘞 lak3		

Table 2 List of common abbreviations

For function words falling out of the above categories, the English equivalent is used. For proper nouns, translation is not necessary but italic *'person.name'* and *'geographic.name'* are used.

7

4. Translation

Line four is a translation of the utterance in plain English. It should be noted that colloquial English should be used to match the genre of the sound track. Upon necessary, if there is a huge distance between the literal meaning and the translation; in other words, if the words used in the translation are very different from those in the word-by-word gloss, the literal translation should be attached before the translation. The following shows an example:

嗰	個	人	曉得	官話	唔	呢	？
go2	*go3*	*jan4*	*hiu2dak1*	*gun1waa2*	*m4*	*ne1*	
that	CL	person	know	Mandarin	NEG	Q	

Lit. 'Does that person know Mandarin?'
'Does he understand Mandarin?'

For more fine details on glossing rules, you are referred to the following resources:

Leipzig Glossing Rules: http://www.eva.mpg.de/lingua/resources/glossing-rules.php
Interlinear morphemic glosses:
http://www.ling.hawaii.edu/ldtc/website/syllabus/sp06/LehmannGlossing.pdf

Comments and suggestions are welcome! For other questions concerning transcription, please direct to Mr Sam Wong Tak-sum at egwts@polyu.edu.hk or Miss Ariel Chan Shuk-ling at ariel.sl.chan@polyu.edu.hk .

Reference

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California.

Guan Jiecai. 1990. *A Dictionary of Cantonese Colloquialisms in English*. Hong Kong: Commercial Press.

Hutton, Christopher M. and Kingsley Bolton. 2005. *A Dictionary of Cantonese Slang: The Language of Hong Kong Movies, Street Gangs and City Life*. Honolulu: University of Hawai'i Press.

Lehmann, Christian. 1983. Directions for interlinear morphemic translations. *Folia Linguistica* 16: 193−224.

Lo, Tam Fee-yin. 2007. *Cantonese Colloquial Expressions*. Hong Kong: The Chinese University Press.

Matthews, Stephen and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.

So, Siu-hing Simon. 2002. *A Glossary of Common Cantonese Colloquial Expressions*. Hong Kong: The Chinese University Press.

Swofford, Mark. Basic rules of Hanyu Pinyin orthography (Summary). Pīnyīn.info: a guide to the writing of Mandarin Chinese in romanization. Modified 2010. Accessed 2nd October, 2012. http://www.pinyin.info/readings/zyg/rules.html

Yin, Binyong, and Mary Felly. 1990. *Chinese Romanization: Pronunciation and Orthography.* Peking: Sinolingua.

Bai, Wanru 白宛如 1998：《廣州方言詞典》（《現代漢語方言大詞典・分卷》，李榮 主編），南京：江蘇教育出版社。

Chao, Yuen Ren 趙元任 著，丁邦新 譯 1982：《中國話的文法》，香港：香港中文大學出版社。

Cheung, Hung-nin Samuel 張洪年 1972：《香港粵語語法的研究》，香港：香港中文大學出版社。

Kong, Zhongnan 孔仲南 1933：《廣東俗語攷》，南方扶輪社。1992 年經上海文藝出版社重新影印出版。

Leung, Chung-sum 梁仲森 2005：《當代香港粵語語助詞的研究》，香港：香港城市大學語言資訊科學研究中心。

LSHK 香港語言學學會粵語拼音字表編寫小組 2002：《粵語拼音字表》，第二版，香港：香港語言學學會。

Li, Xinkui 李新魁、黃家敎、施其生、麥耘、陳定方 1995：《廣州方言研究》，廣州：廣東人民出版社。

Mai, Yun 麥耘、譚步雲 1997：《實用廣州話分類詞典》，廣州：廣東人民出版社。

Rao, Bingcai 饒秉才、歐陽覺亞、周無忌 1997：《廣州話詞典》，廣州：廣東人民出版社。

Wong, Shek Ling 黃錫凌 1941：《粵音韻彙》，香港：中華書局。

Yu, Xuepu 虞學圃、溫岐石 1915：《新輯寫信必讀分韻撮要合璧》，原著於 1782 年成書，近年經香港陳湘記書局重印。

Zheng, Ding'ou 鄭定歐 1997：《香港粵語詞典》，南京：江蘇教育出版社。

Appendix 1: A list of Chinese characters for transcribing Cantonese expressions

Chinese Character	Pronunciation
喔	[ɔʔ²²]
囃	[hai³⁵³]
喀	[hɐʔ]
嘞	[lɐʔ]
吖	a1 / aa1
呀	a3 / aa3
啊	aa2
呀話？	aa6waa5
哎	ai1
呝	ak3 / aak3 (*utterance particle*)
罷	baa2
嘪	be3
誒	e6
欸	ei3
咖嘛	ga1ma3
㗎	ga3 / gaa3
㗎嘑	ga3la3 / gaa3laa3
噉	gam2
咁	gam3
嘅	ge3
吓	haa2
下話？	haa6waa5
唏	hei1
愾	hei3
噷	hm
嘑	la3 / laa3
喇	laa1 / la1
嗱	laa4
呢	le1 (*utterance particle*), li1 (*pronoun*)
唎	le3
咧	le4
耒	le5
囉	lo1
囉喎	lo3wo3
囉	lo4
咪	mai6
緡	man1
咪嘢	mi1je5
哦？	o2
哦	o4

卅	saa1aa6
聽日	ting1jat6
哇	waa1
喎	wo3
啝	wo4
搞	wo5
咋嘛	za1ma3
咋	zaa3
咋？	zaa4
唧	zek1

Last Updated: 6 January 2015 3:32 PM