An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese

Wai Yi P. Wong, Marjorie K. M. Chan, Mary E. Beckman

10.0 INTRODUCTION

The Hong Kong Cantonese variety of Chinese (hereafter “Cantonese”) poses an interesting challenge for prosodic typology and transcription for three closely inter-related reasons. First, compared to the Mandarin varieties of Chinese, Cantonese has far fewer polysyllabic wordforms. The majority of the syllables are potentially free-standing morphemes, and there is no contrast between “stressed” syllables and reduced (“neutral-tone”) syllables.

Second, there is an extremely dense syntagmatic specification of tone. Every syllable in an utterance has a lexical tone, even if it is a grammatical morpheme or pragmatic particle, and there is a rich inventory of non-segmental pragmatic morphemes (“boundary tones”) that can be added after the final lexical tone to mark the ends of intonational phrases. Finally, while there are occasional consonant lenitions, vowel deletions and attendant resyllabifications that create “fused forms” of familiar phrases in running speech, there seem to be no other reliable categorical markings of intermediate levels of prosodic grouping between the syllable and the intonational phrase. Thus, it is difficult to define a low-level unit comparable to the “prosodic word” of Greek (Arvaniti and Balthazani, this volume), the “accentual phrase” of Korean (Jun, this volume), or the “tone sandhi group” in the Wu varieties of Chinese (Jin 1985; Selkirk and Shen 1990). The C_ToBI (Cantonese Tones and Break Indices) conventions are designed to annotate and explore these tone and
juncture phenomena in spoken language corpora. They are developed within the Autosegmental-Metrical approach of the ToBI framework, which represents the string of tones on its own autosegmental tier, independent of the different structural positions that license tones at different levels of the metrical hierarchy.

While the development of the C_ToBI annotation conventions is based primarily on modern Hong Kong Cantonese, the proposal is for common levels of transcription for C_ToBI users at different sites for transcribing other varieties of Cantonese, such as Guangzhou (Canton) Cantonese, Zhongshan Cantonese, and so forth. The three properties that we outlined above as characteristic of Hong Kong Cantonese seem to hold in broad terms also for these other varieties. Variation that we have noticed in our own casual observations seems to involve only some paradigmatic differences in lexical tone and boundary tone inventory (and not in the dense syntagmatic specification of tone) as well as differing propensities for fused forms. (Fused forms seem to occur rather less frequently and only in more casual styles in less urban varieties.) Also, at this preliminary stage, development has been based primarily on constructed examples, although work is also in progress on an elicited spontaneous speech corpus. Again, however, it seems unlikely that further work with more different styles will change the broad outline of facts outlined above. In designing C_ToBI, therefore, we have tried to capture the generally shared structure, while leaving scope for change in detail as we compare transcriptions across different styles, and especially across different transcription sites. Cross-site comparison of transcriptions is desirable in order to arrive at an agreed-upon set of levels of transcription congenial to the community at large. Only in this way can we hope to be able to capture any sociolinguistically significant variation in prosody across the different varieties of
Cantonese and across the different styles that a native speaker of Cantonese controls in everyday life.

The challenge that Hong Kong Cantonese poses for prosodic annotation then comes from several facts that seem to be generally true of Cantonese in general. Moreover, these facts about Cantonese prosody make it different not just from languages such as American English and Tokyo Japanese, but also from other major varieties of Chinese, including Mandarin (as described in Peng et al., this volume). We will give an account of these facts in section 10.1. To capture these essential aspects of the language, we take advantage of the facilities that the ToBI framework has provided — namely, that it suggests only a basic differentiation between tonal specification and the prosodic groupings with which tones can be associated. Beyond this distinction between melody (Tones) and junctures (Break Indices), it allows for free proliferation of tiers and user-defined labels specific to the needs of the language in question and the interests of individual sites. To illustrate, we will lay out details of the C_ToBI annotation conventions, in section 10.2. The conventions were developed out of what we know already about the prosody of Cantonese and our initial hypotheses that require further testing. In section 10.3, we will close by listing some questions of interest that we plan to explore using C_ToBI.

10.1 FACTS ABOUT CANTONESE PROSODY

10.1.1 What do we mean by “Cantonese”? We use the term “Cantonese” in three senses. The first and oldest sense is in reference to the variety of Chinese spoken in and around the city of Canton (Guangzhou), the provincial capital of Guangdong Province in southern
China. Because many Canton City inhabitants have migrated to Hong Kong in the past 150 years or so (Hashimoto 1972:70), the variety of Cantonese spoken in Hong Kong is more similar to that of Canton than are varieties of the language spoken in towns closer to Canton City itself. Hence, socially prestigious “Standard Cantonese” can refer to the Cantonese spoken in Hong Kong as well as in Canton, even though over time a number of differences have emerged in the speech of these two localities (Bauer and Benedict 1997).

It is in this second usage of the term “Cantonese” that we focus here on modern Hong Kong Cantonese. Finally, the term “Cantonese” has further been extended to refer to the entire group of similar dialects that, together, form one of the major sub-classifications within Chinese, parallel to Mandarin, Wu, Min, and so forth. (One also encounters the term “Yue” for that usage — see, e.g., Hashimoto 1972; Yuan 1983; Grimes 1996.) Because of its close affiliation with the original regional standard of Canton City, the Standard Cantonese of Hong Kong can also serve as representative of that major sub-group within the Chinese language. However, it is important to recognize that many other varieties of Cantonese in this third sense are also found in Hong Kong and vicinity, and the present-day system cannot be understood without taking into account the rich sociolinguistics of contact among these varieties.

According to a 2001 report from the Hong Kong Census and Statistics Department, of the 6,708,389 inhabitants in Hong Kong, 89.2% speak Cantonese as their mother tongue, which is also the language of everyday use. Cantonese is also found in numerous overseas speech communities around the world, including Singapore, Malaysia, Vietnam, Indonesia, Thailand, Philippines, New Zealand, United Kingdom, United States, and Canada. In fact, the Summer Institute of Linguistics’ *Ethnologue: Languages of the World* (Grimes 1996)
gives an estimate of some 66 million speakers of Cantonese (“Yue”) in the world today, and ranks it as sixteenth in its list of top 100 languages by population. Ideally, we would like C_ToBI to be able to account for prosodic structure across this world-wide community of Cantonese speakers. At the least, it should describe the situation in Guangzhou, Hong Kong, and adjacent Cantonese-speaking areas.

10.1.2 The syllable. One important characteristic of Cantonese is the highly salient psychological reality of the syllable. The syllable is easy to define at the phonological level. The dense syntagmatic distribution of lexical tones marks the number of syllables in any stretch of speech more clearly than in any other variety of Chinese. Also, outside of fused forms, syllable boundaries are clearly identifiable from asymmetries in the distribution of onset and coda segments. That is, there is nothing like the large-scale ambiguities of segmentation within the stress feet of an English utterance. The morphology of the language is also highly conducive to thinking phonologically in terms of syllables, since more than any other major variety of Chinese, Cantonese is probably closest to the original Chinese morphological structure of a strong one-to-one correspondence between words and syllables.

The Cantonese syllable has an optional onset consonant and a rhyme, which (in addition to the obligatory tone specification) consists of either a simple vowel, a simple vowel followed by an optional coda consonant, a vowel-glide diphthong, or syllabic nasal. While there is a fairly rich inventory of consonants in onset position — including labial, dental, and velar nasals, and labial, dental, velar and labiovelar stops in both a voiceless aspirated and a voiceless unaspirated series — consonants in coda position are restricted to
the three nasals and labial, dental, and velar voiceless (unreleased) stops. The vowel inventory includes both front and back rounded vowels. The full inventory of consonants and vowels in Hong Kong Cantonese is given in Appendix I. At the level of the syllable, there are some co-occurrence restrictions. Most notably, labials and labiovelars almost never occur as onsets in syllables that have either labial codas or rounded vowels (Hashimoto 1972:110ff, 137ff; see also Newman 1987; Yip 1988; Bauer and Benedict 1997), with other places showing less robust Obligatory Contour Principle (OCP) effects.\(^1\) The rare exceptions to the labial OCP include onomatopoeia and loanwords.

10.1.3 Tone specification. As noted in section 10.0 above, different varieties of Cantonese differ in their lexical tone inventories, and “tone change” is one of the more salient variables in the sociolinguistic mix (see section 10.1.5). Hong Kong Cantonese today has the six phonemic tones illustrated in Figure 10.1 on the syllable /wai/. These tones are traditionally described using Yuan-ren Chao’s system of tone numbers, with ‘1’ for lowest pitch and ‘5’ for the highest pitch in the local pitch range. There are three level tones (/55/, /33/, /22/), two rising tones (/35/, /23/), and one falling tone (/21/).\(^2\) Some Cantonese speakers also have a high falling tone (/53/), which for most Hong Kong speakers has merged with the high level tone. In checked syllables (i.e. syllables closed by a stop) the three level tones are traditionally transcribed with single digits (i.e. /5/, /3/, /2/), iconically reflecting the shorter duration of the vowel. The C_ToBI labelling conventions adopt this traditional transcription of lexical tone, except that the iconic encoding of the durational difference between checked and non-checked syllables is made uniform by the addition of an “extra” digit to the rising and falling tones in syllables with all-sonorant
rhymes, as shown in Figure 10.1. (More detail on the labelling conventions for lexical tone is given in section 10.2.1(iii).)

Cantonese also has boundary tones. These are “extra” tones that can be added after the final lexical tone of an intonational phrase to produce various pragmatic effects. The C_ToBI conventions distinguish boundary tones from the lexical tones by transcribing them with labels that (following Pierrehumbert 1980) have become standard in autosegmental-metrical descriptions of other languages. Figures 10.2 through 10.4 illustrate three of the boundary tones of Hong Kong Cantonese, all added at the end of an utterance after a /33/ lexical tone on the utterance-final syllable. Observe the substantial lengthening of the final word, particularly for the complex boundary tone in Figure 10.4, where one could almost say that an extra lengthening is necessary to make room for these two “extra” tones at the phrase boundary. Figure 10.3 also illustrates an utterance-medial phrase boundary where there is substantial final lengthening, but no “extra” tone after the mid-level lexical tone on the particle /aa33/.

[FIGURE 10.2 HERE]

[FIGURE 10.3 HERE]

[FIGURE 10.4 HERE]
The boundary tones that we have observed so far in our recordings of Hong Kong Cantonese are inventoried further in section 10.2.1(iii). As with the lexical tones, we expect to discover variation and change in the inventory of boundary tones across varieties of Cantonese spoken in different regions and by different social groups. However, we do not expect varieties to differ from Hong Kong Cantonese in not having boundary tones. These tones and the other effects with which they are associated (such as the “phrase-final lengthening” just described) thus define a level of prosodic grouping above the word. This level is aptly termed the “intonational phrase” in keeping with the comparable level of prosodic hierarchy in German, Japanese, Greek, etc. (see chapters by Grice and Benzmueller, Venditti, and Arvaniti and Balthazani).

As Figures 10.2 through 10.4 well illustrate, Hong Kong Cantonese has a very dense syntagmatic specification of tone, with each syllable in a well-formed phrase bearing a lexical tone and the phrase-final syllable also often bearing a boundary tone that is a separate pragmatic morpheme specified for the phrase as a whole. Thus, each syllable is specified for one of the six tones in the language, and pragmatic particles are no exception. That is, by contrast to Taiwanese, Mandarin, and probably most other varieties of Chinese, Cantonese particles bear lexical tone. There are no unstressed (neutral tone) syllables, not even for these “little” function words. Cantonese has a particularly rich inventory of final particles. Depending on the treatment of closely related variant forms and of more or less conventionalised particle sequences, the count can be anywhere from about 30 (Kwok 1984) to as many as 206 (Yau 1980). This is in marked contrast to the seven particles commonly listed for Standard Mandarin (e.g. Matthews and Yip 1984:238). Since these final pragmatic particles are by definition final, they interact with tonal pragmatic
morphemes ("boundary tones") to yield very complex pragmatic effects (cf. Kwok 1984; Chan, Wong and Beckman 1998; Fung 2000). Precisely how these final pragmatic particles and tonal pragmatic morphemes interact leaves ample room for further research. The development of the C_ToBI annotation conventions should contribute a tool to help us understand the precise nature of the interaction.

10.1.4 Fusion forms. There are segmental effects in Hong Kong Cantonese that fuse two syllables together into a polysyllabic "word" in fast speech (Li 1986; Wong 1996) (see example (1b)). Thus, the facts about syllable structure outlined above in section 10.1.2 must be qualified to some extent, at least for the Hong Kong variety of Cantonese. This phenomenon of "fusion" has not been systematically studied. Indeed, there is little prior work on any aspect of Cantonese connected speech. Several examples of fused forms are listed in (1). (See also examples in Figures 10.4 through 10.7 in section 10.2.1(v)).

We define syllable fusion as follows: in a sequence of two syllables, there is substantial weakening or effective deletion of the oral gesture(s) of the segment(s) contiguous to the syllable boundary. More extreme fusion can simplify contour tones and "merge" the qualities of vowels that would be separated by an onset or coda consonant at more "normal" degrees of disjuncture between words. However, even at these more extreme junctures, fusion does not usually override the lexical tones of the syllables (Hashimoto 1972; Cheung 1986; Li 1986; Wong 1996), and it is only in the most extreme cases (e.g. mat5 je23 ‘what’ → me55 ‘what’) where we see true "tone loss". The fusion process, therefore, seems to be a gradient effect. Although it is possible that the most extreme cases of fusion may make the syllable count less determinate, fusion does not
seem to be a categorical change from two syllables to one in all cases, contra our own earlier description in Wong (1996).

(1) Examples of syllable fusion forms.
   a. know NEG
      sik5.m21 → si25.m21 → si5.m21
   b. in fact
      kei21.sat2 → ke21.at2 → ket21-2
   c. know
      zi55.dou33 → zi55.ou33 → ziu55-33

Li (1986) describes fusion as the postlexical counterpart of historical contractions resulting in “changed tones” (see next section). A complete account of fusion probably will need to place it in the context of the sociolinguistics of contact among these dialects. Our current preliminary account is in terms of the foot (cf. also Yip 1980; Duanmu 1990). That is, we ascribe the “monosyllabic” flavour of Cantonese to the preponderance of monosyllabic feet. Fused forms then are exceptions to the regular one-to-one correspondence between foot and syllable, and may presage a change toward a system with a more well-defined intermediate level of prosodic grouping between the syllable and intonational phrase.

10.1.5 Segmental and tonal alternations. In addition to the gradient variation that results from fusion, there are several categorical segmental and tonal alternations in the
Hong Kong variety of Cantonese that are of interest in the development of C_ToBI, for two reasons. First, some of these alternations must have arisen historically through the lexicalisation of connected speech processes and discourse-level phenomena described in the sections above, and thus give a hint of the kinds of prosodic change in progress to look for in modern Cantonese. Second, many of these alternations are socially meaningful consequences of the extreme linguistic diversity that has characterised Hong Kong as long as it has existed. (The territory was ceded to England and opened as a port city in 1842.)

Hong Kong’s history of rural-urban contact, with an accelerated pace of urban immigration in the last half century, has made it a mixing pot for Cantonese dialects, much like Shanghai in relationship to the Wu varieties much earlier. For example, in addition to Hong Kong Cantonese, other varieties encountered in Hong Kong include a number of dialects (e.g., Dongguan, Panyu, Shunde, Taishan, Kaiping) in which velars are merging with dentals in coda position (cf. Zhan and Cheung 1987; Zee 1999b). The resulting alternation between /t, n/ and /k, ng/ obscures further the less robust OCP effects with lingual consonants. The role of English as an administrative language further increased the linguistic diversity (see, e.g., the stratum of loanwords from English as exceptions to the labial OCP constraint). Nor is English the only non-Cantonese source of potential influence on Hong Kong Cantonese. There have always been speakers of other varieties of Chinese, such as Chiuchao (Chaozhou), Fukien (Fujian), Hakka (Kejia), Shanghainese, and Putonghua Mandarin (Hong Kong Census and Statistics Department 1996:22). With the change of government in 1997, there has been a further influx of such non-Cantonese Chinese speakers. And the trend is likely to increase with the implementation of an
immigration regulation in 2000 that would permit mainland Chinese to obtain approval of residency in Hong Kong at the rate of around 156 heads per day.

Examples of some of the segmental alternations found in Hong Kong Cantonese are shown in (2). (See Hashimoto 1972; Bauer 1979; Yeung 1980; Rao, Ouyang and Zhou 1981; Matthews and Yip 1994; Bourgerie 1990 for more examples.) These alternations occur on consonants in both onset and coda positions, as well as on the nuclear vowel. They may be conditioned lexically or occur across the board. These alternations are often said to be related to more casual versus more formal styles of speech. However, the relationship is not clear-cut. Much work needs to be done before we can adequately characterise the interaction between style and other factors such as the speaker’s dialect background and language attitudes in the currently very fluid sociological relationship between Hong Kong Cantonese, the mainland standard of Canton City, Putonghua, etc. (To aid researchers working on this characterisation, we recommend that sites that choose to include a phones tier in their C_ToBI transcriptions label the observed segmental alternate on the phones tier rather than regularising words to the “underlying” or “standard” form.)
(2) Segmental alternations in the Cantonese syllables (tones omitted). The symbol ‘∅’ represents zero onset.

a. Not lexically conditioned

Onset: 
n ~ l
lei ~ nei you

ng ~ ∅
ngo ~ o I

gw/kw ~ g/k gwong ~ gong bright (before the single vowel /o/)

Coda: 
n/t ~ ng/k tek ~ tet to kick

hon ~ hong sweat
b. Lexically conditioned

Onset:  
- k ~ h  \textit{k}eo\textit{i} ~ \textit{heoi}  s/he  
- k ~ g  \textit{kau} ~ \textit{gau}  to buy  
- c ~ j  \textit{caai} ~ \textit{jaii}  to step on  
- t ~ d  \textit{tit} (daa) ~ traditional Chinese  
  \textit{dit} (daa)  physiotherapy  

Coda:  
- ng ~ k  \textit{kwong} ~ \textit{kwok}  to widen  

Vowel:  
- i ~ e  \textit{ling} ~ \textit{leng}  to receive  
- a ~ aa  \textit{caak} ~ \textit{caan}  to test  
- e ~ u  \textit{mei} (lik) ~ charm  
  \textit{mui} (lik)  

The examples in (3)-(6) illustrate the tonal alternations, with the examples in (3) showing cases which had previously been analysed in terms of derivational processes resulting in \textit{bianyin} ‘changed tone’ (Chao 1947) or ‘modified tones’ (Whitaker 1955-56). That is, these are forms where speakers of some dialects (and classically educated Hong Kong speakers who command literary Cantonese) can still see a more or less transparent relationship between a verb and a derived noun, as in (3a), or between a semantically neutral form and a derived form carrying some extra connotation such as familiarity, diminution, derogation, etc. The derived form in each case has a high rising (/35/) tone or a high level (/55/) tone. Much has been written on changed tones since Chao’s (1947) \textit{Cantonese Primer} and Whitaker’s (1955-56) classic article (e.g. Hashimoto 1972; Wong 1941; Wong 1982; Matthews and Yip 1994; Bauer and Benedict 1997), but an adequate
description of these alternations today must take into account the changed status of literary Cantonese. That is, many younger Hong Kong Cantonese speakers may not identify the lexical stratum of literary readings for characters which gives classically educated speakers access to the historically “underlying” form in the alternations in (3).

The loss of a special status for literary Cantonese also affects the analysis of the examples in (4)-(6). Some of the alternations here arise from dialect borrowing of the literary Cantonese form into the spoken language — hence the relationship to casual versus formal style. Educated speakers who command literary Chinese have a clear sense of the different lexical strata, and think of the literary form as having the “underlying” or “true” lexical tone and the alternate form as having a “changed tone” here as well as in (3). Younger speakers, on the other hand, may not distinguish this type of dialect mixture from any other source of alternate forms. In addition to dialectal contact, sources may include collapsing of stylistic differences, tonal dissimilation, and reinterpretation of “nonce” tones that might have arisen from connected speech effects such as tonal coarticulation. (In keeping with the above recommendation for a phones tier, the C_ToBI convention is to label the observed tonal alternate on the tones tier rather than regularising to the “underlying” form — see section 10.2.1(iii).)
(3) Examples of tone change in Cantonese.
   
   a. paak3 → 35 to slap → racquet (nominalisation)
   
   b. daai22 → 55 big → small (diminution)
   
   c. wong21 → 35 yellow/surname → egg yolk

(4) Examples of tonal alternations in monosyllabic lexical items.

   a. guk2 ~ 35 bureau
   
   b. faan23 ~ 22 to overflood
   
   c. wui23 ~ 33 will (aux. verb)
   
   d. go35 ~ 55 (go33) that (one)

(5) Examples of tone alternation in only one of two related forms.

   a. jyu35 / jyu21 pin35 fish / fish fillet
   
   b. juk35 / juk2 aak35 jade / jade bracelet

(6) Examples of tonal alternations in reduplicative forms.

   a. zim22 zim22 ~ 35 gradually
   
   b. kam21 kam21 ~ 35 ceng55 hastily
However, whatever was their origin, none of these tonal alternations in modern Hong Kong Cantonese marks a level of prosodic grouping. That is, they are quite unlike the tonal alternations that define the sandhi group in Xiamen/Taiwanese (Chen 1987; Peng 1997), Fuzhou (Chan 1985), or Shanghainese (Jin 1985; Selkirk and Shen 1990). Thus, there is no obvious basis for prosodic grouping of Cantonese syllables other than the phenomenon of fusion that we have analysed in terms of the foot, and the boundary tones and other phrase final effects that mark the edges of intonation phrases.

10.2 THE CANTONESE TONES AND BREAK INDICES (C_ToBI) 

ANNOTATION CONVENTIONS

As in all other systems that use the ToBI framework, C_ToBI requires that the transcription include the following three components: (i) a recorded audio signal; (ii) a fundamental frequency trace for the utterance; and (iii) a set of symbolic labels time-linked to the audio and F0 signals. The labelling platform that was used to make the C_ToBI transcriptions in the illustrations in this paper is xwaves and its associated xlabel function. However, any similar labelling platform that can provide the three basic components of the standard ToBI framework transcription may also be used. In developing C_ToBI, we have been careful to define the levels of transcription and inventory of symbolic levels in ways that allow us also to tap the more richly hierarchical field structures of Emu, a speech database labelling platform developed in the Speech Hearing and Language Research Centre at Macquarie University (Cassidy and Harrington 2001).
levels of transcription and symbolic labels for C_ToBI. The C_ToBI system currently specifies the six levels of transcription listed to the right of the xlabel windows displayed in Figure 10.2. These are for tagging (1) tones, (2) break indices, (3) any polysyllabic foot, (4) syllables, (5) words, and (6) miscellaneous phenomena such as an interval of coughing or an abrupt cutoff of phonation prior to repair of a speech error. The tones tier, words tier, break indices tier, and miscellaneous tier correspond to the original ToBI tiers of the same name. In C_ToBI, however, the words tier is not obligatory (see section 10.2.1(i)). In the following subsections, the six tiers in C_ToBI are described in turn: words tier, syllables tier, tones tier, break indices tier, foot tier, and miscellaneous tier. This section concludes with suggestions for future tiers.

(i) The words tier.

In the original ToBI system, the words tier is a word-by-word orthographic transcription. In keeping with the monosyllabic flavour of Cantonese (see section 10.1.2), the words tier in C_ToBI provides a syllable-by-syllable transcript of the utterance in the native orthography of the language. That is, every syllable is labelled with its corresponding Chinese character, except for syllables that have no native written form, which are instead transcribed in the roman alphabet (see section 10.2.1(ii)). This tier is not obligatory, because it is not functional on labelling platforms that do not allow Chinese character input using Big5, GB, Unicode, or some other comparable standard encoding system. Chinese character input is currently possible using the Emu labeller in Chinese Windows. In an Emu-based transcription, elements on this level are defined as “segments” (i.e. tags for
discrete intervals with both a beginning and end) rather than as “events” (tags which are associated with only one time stamp).

(ii) The syllables tier.
This tier provides an alphabetic transliteration for every element in the words tier. In addition, intervals of silence are flagged with the label <SIL>. The alphabetic transliteration scheme we currently use is the Jyutping Romanization Scheme (1993) developed by the Linguistic Society of Hong Kong (see Appendix I). Unlike the words tier, the syllables tier is obligatory. It serves the function of the words tier for sites that do not have a way to input and/or read Chinese characters. A primary motivation for having a words tier is to allow users of a database to efficiently search for, say, all instances of a given lexical item. Since the syllables tier stands in for the words tier at sites without Chinese character input, we recommend that labels on this tier be made uniform across a speech database. One goal for the C_ToBI labelling community, therefore, should be to develop a freely available communal online character dictionary that regularises the segmental and tonal alternations described in section 10.1.5 to a single “dictionary” form for that morpheme.

(iii) The tones tier.
The tones tier tags the lexical tones and the boundary tones of an utterance (see section 10.1.3). It transcribes the lexical tones syllable by syllable using the traditional Chao-number descriptions with some modification. Specifically, dynamic tones on syllables with all-sonorant rhymes are distinguished from dynamic tones on checked
syllables by doubling the first number in the transcription of the non-checked syllables. Thus, /35/ is transcribed as /335/, and /21/ as /221/ in non-checked syllables, to distinguish them from /35/ and /21/ on checked syllables. (Although /23/ and /53/ do not occur on checked syllables in Hong Kong Cantonese, we double the first digit here as well, to make the transcription of non-checked syllables uniform.) The inventory of lexical tones in Hong Kong Cantonese is given in Table 10.1. This list includes two contour tones on checked syllables that are not usually considered part of the inventory of lexical tones in the older linguistics literature — namely, /35/, which is traditionally analysed as a “changed tone” as in example (3a) above, and /21/, which surfaces occasionally on onomatopoeic words. Here, we use “lexical tones” as a cover term for the full set of tones that contrast monosyllabic morphemes in Hong Kong Cantonese, ignoring intuitions of classically educated speakers who command literary Cantonese.

[TABLE 10.1 HERE]

The tags for boundary tones differ from those for lexical tones. We use “H” and “L” and the “%” diacritic from Pierrehumbert (1980), as in the ToBI framework conventions for other languages. In our preliminary analysis, Hong Kong Cantonese has an inventory of six different phrase-final boundary types. Table 10.2 lists the labels for these types and identifies a figure illustrating it. Note that some of these types are distinguished by accompanying durational and/or voice quality effects — for example, “H%” versus “H:%” in Figures 10.2 versus 10.7, or the “-%” that marks truncation at phrase-end. These labels reflect our preliminary understanding of the “surface” features that contrast the boundary
types; they are not meant as a definitive phonological analysis. The phrase-initial boundary “%fi” differs in function from the other boundary tags. It identifies the initial syllables of “pragmatic idioms” that involve the choice of both a phrase-final boundary effect and a pair of particles that “frame” the construction. Identifying the beginning of the idiom as well as the end will allow us to investigate other potential prosodic markers for these idioms. For example, it is possible that there is a separate pitch range specified for the idiom as a whole.

The phrase-final intonational patterns described in Table 10.2 can occur on a sentence that does not have a final particle, and stand alone to indicate the pragmatic relationship between the sentence and its discourse context. They can also occur with (at least some of) the final particles to produce more complex pragmatic effects, combining the meaning of the intonational pattern with that of the particle, or particle sequence. As noted above, these effects include both tonal phenomena (an “extra” boundary tone can be added after the lexical tone of the last syllable), and rhythmic/voice-quality phenomena (elongation versus truncation of the last syllable, associated at least sometimes with a contrast between breathy versus creaky or checked voice source). Figures 10.8a and 10.8b
give a minimal pair of examples ending with the final particle ge335. The speaker in (10.8a) asks a question with the question particle ge335. Notice that no extra tone is added at the end of the particle, so that it is the particle alone that causes it to be interpreted as a question. By adding the L% at the end of the particle in (10.8b), the speaker implies further that the answer of the question is already known, as in an English tag question with a falling H* L-L% tune.

When labelling using the xwaves/xlabel platform, a lexical tone label is aligned with the associated words and syllables tier labels, except when the last lexical tone in a phrase needs to precede a boundary. In that case, sites can choose to specify that the lexical tone label be placed just before the boundary tone, or place it earlier — e.g., at an appropriate inflection point in the F0 contour. Translating this practice into the Emu labelling system, the tones tier could be treated as an independent stream of “segments” — in which case the last lexical tone and the boundary tone will arbitrarily divide the last syllable into two intervals. Another possible conversion schema for going between the two labelling platforms treats the tones tier of the xwaves/xlabel platform as a flattened projection of tone labels for two different levels of the prosodic hierarchy. That is, in an Emu labelling template the lexical tones could be specified as one of several independent label fields for syllables, and boundary tones, similarly, would be a label field for intonational phrases. This takes advantage of the more richly hierarchical database
structures available with Emu, to encode the different tonal domains directly into the representation of the prosodic hierarchy.

(iv) The break indices tier.
The set of break indices in the C_ToBI system encodes the levels of grouping identified in sections 10.1.2 through 10.1.4 — namely the syllable, the foot, and the intonational phrase — and the conventions for annotating break indices were developed in association with the foot tier, the syllables tier and the tones tier. That is, because of the relationship between “words” and syllables in the metrical structure of Cantonese (see sections 10.1.2 and 10.2.2-10.2.3), we can think of the break indices in terms of inter-syllabic juncture, and a break index must be aligned at the end of every label on the syllables tier. Table 10.3 describes the inventory of break labels and diacritics.

[TABLE 10.3 HERE]

Break index 0 marks the weakest disjuncture, for syllable boundaries that are foot internal — i.e. the juncture between syllables in a fusion form, which we now analyse as foot boundary erasure (cf. sections 10.1.4 and 10.2.1(v)). Identification of the syllable juncture in a fusion form may not be easy. (Of course, this is a problem for the words and syllables tier labels as well as for the break indices tier.) The fact that syllable fusion in general does not override the lexical tone helps segmentation in cases such as jai5-22 (fusion form of jat5 hai22 ‘or’). Here, the labels can be placed at the F0 inflection point. However, in cases where the sequence of lexical tones for the fused syllables does not provide an obvious inflection (e.g., in a succession of two identical level tones) and the
spectrographic display does not give much clue either (e.g. \textit{zai22-22} as the fusion form of \textit{zau22 hai22} ‘is’, transcribed on the foot tier as \textit{t+sai22-22}) the ‘0’ can only be placed arbitrarily around the midpoint of the vowel. Break index 1 then marks “ordinary” syllable juncture, which is also the end of a foot. Break index 2, marking intonation phrase break, is obligatory when a boundary tone is placed on the tones tier.

As in other ToBI framework conventions, these break index numbers can be modified by the ‘-’ uncertainty diacritic. Thus, ‘1-’ marks uncertainty about whether sufficient segmental weakening has occurred to ‘erase’ the foot boundary, and ‘2-’ marks uncertainty about whether there is an intonational phrase boundary.

The last two tags are modelled after the AmEng\_ToBI break-index diacritics for disfluent cutoff or prolongation. The label ‘c’ marks abrupt cutoff of the syllable, which may or may not be accompanied by a glottal stop, before the sentence continues. The lexical tone may or may not be truncated. The label ‘p’ marks prolongation of syllables internal to the intonational phrase arising from lexical search and other kinds of hesitation, ground-holding, etc. The diacritic ‘p’ is not to be used when perceived emphasis is marked for certain syllable(s) on the foot tier. In that case, syllable elongation is a means for focal emphasis, and is marked by ‘*’ on the foot tier (see next section).

In the current C\_ToBI annotation conventions, we have defined these as stand-alone labels rather than as diacritics for the break index numbers. Given the dense syntagmatic distribution of tones in Cantonese and our still very incomplete understanding of syllable fusion, it is impossible to make a principled distinction between ‘1c’ and ‘2c’ or between ‘1c’ and ‘0c’ at this time. Also, while there is nothing in theory to preclude a
distinction between ‘1p’ and ‘0p’, we have not seen any examples to date of prolongation
of the segmental material for the non-final syllable within a fused form.

In translating from xlabel tiers into an Emu labelling system, break indices could be
treated as an independent stream of events. Alternatively, the break indices tier could be
replaced by an explicit hierarchy of syllables nested within feet, which are in turn nested
within intonational phrases (see section 10.2.1(iii)). The latter move would emphasise the
hierarchical relationship among the different metrical units that the different break index
values mark, and would allow the user to take better advantage of Emu’s hierarchical
querying language.

(v) The foot tier.
The foot tier is used primarily to tag fusion forms, which we analyse as erasure of the foot
boundary. We transcribe these feet with a phonetic transcription of the resulting fused form.
(See Appendix I for our set of phonetic symbols, i.e., “phones” symbols5.) The dash
notation in the transcription of the form $la_{221}-22$ in Figure 10.4 indicates the concatenation
of lexical tones in a fusion form. We do not transcribe duration of lexical tones in a fusion
form, since this can be measured in the acoustic signal; but truncated rise, as well as falling
tone undershoot will be captured. Hence, truncated /23/ and undershoot of /21/ (or 223 and
221 respectively in the C_ToBI transcription of lexical tones; see Table 10.1) will both be
transcribed as 22 in a fusion form. (See Figures 10.5, 10.7 and 10.8a for examples of
truncated rise and Figure 10.8b for examples of falling tone undershoot.)

We also use the foot tier to tag emphasised syllables and phrases — i.e.
monosyllabic “words” or longer sequences that are set off by an expanded pitch range,
longer duration, clearer articulation, and other hallmarks of focal prominence (or phrase-level “stress”). Understanding the relationship between the domain of these effects and the prosodic hierarchy of syllable, foot, and intonation phrase is an important issue both for the ToBI framework and for phonological theory in general. However, there has been no systematic research on this question for any variety of Cantonese. Thus, our (preliminary) decision to tag these effects on the foot tier is quite arbitrary, and should not be interpreted as a claim that the foot of Cantonese is to be identified with the foot in “stress” languages such as English and Mandarin. The mark for emphasis is a ‘*’ and the grammar is similar to that on the miscellaneous tier (see section 10.2.1(vi)). That is, there are paired labels ‘<*’ and ‘>*’ for the beginning and end of an emphasised stretch that is longer than a syllable, and the unitary label ‘<*>’ for a single emphasised syllable. When the first syllable of a fused form is emphasised, we mark it with ‘<*>’ in front of the phonetic transcription. An example is given in Figure 10.5.

In translating this tier into a hierarchical Emu labelling template (as suggested in earlier sections), the fusion form transcriptions could be treated as another field of tags for the foot tier segments. The emphasis tags, however, would need to be treated as a separate stream of segment labels, since the relationship between the domain of these effects and the prosodic hierarchy is not yet known. This highlights one of the principal advantages of the looser structure of the original ToBI framework. While the hierarchy of break indices is amenable to interpretation in terms of the prosodic hierarchies proposed in work such as Nespor and Vogel (1986), Selkirk and Shen (1990), or Keating et al. (in press), the interpretation is not enforced. That is, the architecture of independent parallel tiers does not
require that the domain of an effect that is tagged on another tier be identified with any particular break index level unless there is sufficient theory-external evidence for it.

(vi) *The miscellaneous tier.*

As in other ToBI systems, this tier is used to mark breaths, coughs, disfluencies, etc., using labels with ‘<’ and ‘>’ to mark roughly the beginning and the end points of the labelled effect — e.g., ‘laugh<’ and ‘laugh>’ for an interval where speech is interrupted by laughter. Finer specifications may be devised according to site-specific needs.

(vii) *Other tiers.*

We recommend that at sites where there are resources to do a phoneme-by-phoneme segmental labelling, there also be a phones tier. Appendix I lists our proposed inventory of phones symbols. (Both the IPA and Jyutping symbols are given.) We also recommend the eventual development of a sociolinguistic variables tier and a codes tier, given the facts in section 10.1.5. The detailed working out of these other tiers should be modelled on annotation conventions in other communities which have adopted the ToBI framework where development of such tiers is well underway.

### 10.3 DISCUSSION AND FUTURE WORK

The conventions for transcribing spoken Cantonese corpora developed thus far encode what we know already about Cantonese prosody, as well as our “best guess” hypotheses about phenomena that need to be further tested before the labelling conventions can be
“frozen” for large-scale database development. In this section, we summarise the hypotheses that seem most solidly grounded in current knowledge, and the facts on which they are based, before listing further questions for research.

We know, for instance, that boundary tones come at the edges of prosodic constituents larger than the word. In some speech styles, these edges can be marked by a very salient elongation of the final syllable’s rhyme, and these may involve utterances with or without final particles. Particles in sequence can easily exceed a second in duration (Chan 1998), a phenomenon that seems to be particularly salient in Cantonese. A deliberate truncation can also occur here, with a definite pragmatic effect that seems to interact with sentence particle meaning in the same way that boundary tones do. For example, the final particle zek5 in this variant — with checked syllable and shorter duration — is a stronger way to voice a complaint than the corresponding open syllable variant, ze55, which can be elongated for different pragmatic effects (Chan 1996 and sources cited therein; Fung 2000). Further work is needed to explore the precise nature of these interactions, and what stylistic and other constraints there are on both truncation and prolongation.

We also know that fused forms exist, and that they are more common at faster tempo (see Wong 1996). Li (1986) also notes style of speech, frequency of use, and segmental make-up as possible factors for contraction/fusion. There may also be speech register and regional dialectal influences. Further research is needed to understand different aspects of this phenomenon, including the extent to which it occurs across the different varieties of Cantonese. Besides Hong Kong Cantonese, fusion has also been reported for Toisan (Taishan) Cantonese (Kong 1984), for example. We currently analyse fusion as an erasure of the foot boundary, because the syllable edge becomes blurred
without a necessary reduction in syllable count in all cases (contra an earlier analysis in Wong 1996). Further corpus work combined with psycholinguistic experiments should be helpful in testing these two competing analyses.

Also, potential constraints from emphasis/stress on the potential for fusion forms need to be explored. Our native speaker intuition is that if just one of the syllables in a fused form is emphasised, it has to be the first. If this turns out to be true in the databases that we develop with this annotation tool, then we will have evidence of the possible beginnings of a more Mandarin-like stress system, albeit realised in segmental lenition without tone deletion.

Prosodically annotated speech corpora of Cantonese are essential for answering these and other questions. We hope that results of corpus studies will feed back into the development of the labelling system, appropriately tuning the current set of labels, and rendering C_ToBI an efficient tool for using annotated speech corpora for diverse research purposes and technological applications.

ACKNOWLEDGEMENTS

We thank Steve Cassidy and Jonathan Harrington of the Speech and Hearing Research Centre at Macquarie University for advice in the implementation of the EMU speech database labelling platform at our site. We also thank Sun-Ah Jun and an anonymous reviewer for providing us with thoughtful suggestions and comments.

REFERENCES

Arvaniti, A. and Balthazani, M. (this volume), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora'.

29


Accessible from the World Wide Web:

Beckman, M. E. and Jun, S.-A. (1996), ‘K-ToBI (Korean ToBI) Labeling Convention, version 2.1’, Department of Linguistics (Ohio State University) and Department of Linguistics (University of California, Los Angeles).

Accessible from the World Wide Web:


Bruce, G. (this volume). ‘Intonational Prominence in Varieties of Swedish Revisited’


Grice, M., Baumann, S., and Benzmüller, R. (this volume), ‘German Intonation within the Framework of Autosegmental-Metrical Phonology’.


Accessible from the World Wide Web:


Hong Kong Census and Statistics Department (2001), *Population Census* (Hong Kong Special Administrative Region).

Accessible from the World Wide Web:

Ethnicity and language use –

<http://www.info.gov.hk/censtatd/chinese/hkstat/fas/01c/01c_index.html>

[Accessed 2 January 2002].


Jun, S.-A. (this volume), ‘Korean Intonation and Prosodic Transcription’.


Kong, B. (1984), ‘Assimilation, contraction, and word fusion in Toisan,’ manuscript (University of British Columbia).


Leung, C.-S. (1992), ‘A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong,’ M. Phil. thesis (The Hong Kong Polytechnic University).


Wong, S.-L. (1941), A Chinese Syllabary Pronounced According to the Dialect of Canton (Hong Kong: Chung Hwa Book Co. Ltd.).


34


APPENDIX I

A correspondence table showing the Jyutping Romanization Scheme (1993) for Cantonese
<http://www.hku.hk/linguist/lshk/>, the phonetic symbols (the “phones” symbols) that we use in transcribing fusion forms on the foot tier (and for sites having a phones tier), and the corresponding IPA symbols (length distinction of nuclear vowels are indicated (cf. Lee 1999; Zee 1995, 1999a, 1999c)).

<table>
<thead>
<tr>
<th>Consonants</th>
<th>IPA</th>
<th>Phones</th>
<th>Jyutping</th>
<th>Vowels</th>
<th>IPA</th>
<th>Phones</th>
<th>Jyutping</th>
<th>Diphthongs</th>
<th>IPA</th>
<th>Phones</th>
<th>Jyutping</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p+H</td>
<td>p</td>
<td>i</td>
<td>i</td>
<td>u</td>
<td>i</td>
<td>ui</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>b</td>
<td>y</td>
<td>y</td>
<td>ei</td>
<td>y</td>
<td>ei</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>t+H</td>
<td>t</td>
<td>u</td>
<td>u</td>
<td>ei</td>
<td>i</td>
<td>oi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>d</td>
<td>i</td>
<td>i</td>
<td>ai</td>
<td>ai</td>
<td>ai</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>k+H</td>
<td>k</td>
<td>u</td>
<td>U</td>
<td>ai</td>
<td>Ai</td>
<td>aai</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>g</td>
<td>e</td>
<td>E</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
<td>æ</td>
<td>R</td>
<td>œ</td>
<td>œ</td>
<td>œ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>ø</td>
<td>&amp;</td>
<td>eo</td>
<td>eo</td>
<td>eo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ŋ</td>
<td>N</td>
<td>ng</td>
<td>ŋ</td>
<td>O</td>
<td>o</td>
<td>ou</td>
<td>ou</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>u</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>s</td>
<td>a</td>
<td>A</td>
<td>aa</td>
<td>øy</td>
<td>&amp;y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>l</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>@</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tsʰ</td>
<td>tsʰ</td>
<td>t+s+H</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>T+S</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>t+s</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>T+S</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kʰw</td>
<td>kw+H</td>
<td>kw</td>
<td>kw+H</td>
<td>kw</td>
<td>kw</td>
<td>kw</td>
<td>kw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kʰw</td>
<td>kw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllabic nasals</td>
<td>IPA</td>
<td>Phones</td>
<td>Jyutping</td>
<td>Others</td>
<td>IPA</td>
<td>Phones</td>
<td>Jyutping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>M</td>
<td>m</td>
<td>æ</td>
<td>X</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ŋ</td>
<td>ŋ</td>
<td>N</td>
<td>ng</td>
<td>ŋ</td>
<td>W</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tsʰ</td>
<td>tsʰ</td>
<td>t+s+H</td>
<td>c</td>
<td>tsʰ</td>
<td>T+S</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>t+s</td>
<td>z</td>
<td>tsʰ</td>
<td>T+S</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kʰw</td>
<td>kw+H</td>
<td>kw</td>
<td>kw+H</td>
<td>kw</td>
<td>kw</td>
<td>kw</td>
<td>kw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kʰw</td>
<td>kw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td>gw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>IPA</td>
<td>Phones</td>
<td>Jyutping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>@</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>æ</td>
<td>X</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ŋ</td>
<td>W</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tsʰ</td>
<td>T+S</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tsʰ</td>
<td>T+S</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>S</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>q</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~</td>
<td>~*</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
* The symbol “~” is to be added after the segment that is nasalised.
### Table 10.1  C_ToBI transcription of lexical tones.

<table>
<thead>
<tr>
<th>Tones</th>
<th>Non-checked syllable</th>
<th>Checked syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level tones:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High level</td>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>Mid level</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Low level</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td><strong>Rising tones:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High rising</td>
<td>335</td>
<td>35</td>
</tr>
<tr>
<td>Low rising</td>
<td>223</td>
<td>--</td>
</tr>
<tr>
<td><strong>Falling tones:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low falling</td>
<td>221</td>
<td>21</td>
</tr>
<tr>
<td>High falling*</td>
<td>553*</td>
<td>--</td>
</tr>
</tbody>
</table>

* For those speakers who make a distinction between /55/ and /53/.
Table 10.2 The inventory of boundary tones in C_ToBI.

<table>
<thead>
<tr>
<th>Tone types</th>
<th>Descriptions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>L%</td>
<td>fall from the final lexical tone</td>
<td>Figure 10.3</td>
</tr>
<tr>
<td>H%</td>
<td>rise from the final lexical tone</td>
<td>Figure 10.2</td>
</tr>
<tr>
<td>H:%</td>
<td>rise from the final lexical tone, with a short plateau at the very end of the rise; incredulity reading accompanied</td>
<td>Figure 10.7</td>
</tr>
<tr>
<td>HL%</td>
<td>final rise and then fall from the final lexical tone</td>
<td>Figure 10.4</td>
</tr>
<tr>
<td>%</td>
<td>phrase-end with no extra tone</td>
<td>Figure 10.5</td>
</tr>
<tr>
<td>-%</td>
<td>truncated rise of the final lexical tone</td>
<td>Figure 10.6</td>
</tr>
<tr>
<td>%fi</td>
<td>frame-initial boundary used to mark the initial particle in phrase-framing particle pairs such as “mat5 ......me55? (rhetorical question)”</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 10.3 Levels of disjuncture and other diacritics used in the C_ToBI system.

<table>
<thead>
<tr>
<th>Break index / diacritics</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>foot internal syllable boundary</td>
</tr>
<tr>
<td>1</td>
<td>end of a syllable that is also end of foot</td>
</tr>
<tr>
<td>2</td>
<td>intonation phrase end</td>
</tr>
<tr>
<td>1-</td>
<td>uncertainty between 0 and 1</td>
</tr>
<tr>
<td>2-</td>
<td>uncertainty between 1 and 2</td>
</tr>
<tr>
<td>c</td>
<td>an abrupt, disfluent cutoff of phonation</td>
</tr>
<tr>
<td>p</td>
<td>prolongation at a disfluency (&quot;hesitation pause&quot;)</td>
</tr>
</tbody>
</table>


Figure 10.1a  A display of three of the six phonemic tones in modern Hong Kong Cantonese /55, 335, 33/, as illustrated by the syllable /wai/ (in Jyutping romanization, see Appendix I for the scheme) with F0 traces. Annotated at the end of each tone are Chao-type tone numbers as modified in the C_ToBI labelling conventions, with ‘1’ to ‘5’ in ascending pitch height (e.g., /55/ is high-level; /221/ is low falling, etc.) (cf. Table 10.1).
Figure 10.1b A display of three of the six phonemic tones in modern Hong Kong Cantonese /221, 223, 22/, as illustrated by the syllable /wai/ with F0 traces. Annotated at the end of each tone are Chao-type tone numbers as modified in the C_ToBI labelling conventions, with ‘1’ to ‘5’ in ascending pitch height (e.g., /55/ is high-level; /221/ is low falling, etc.) (cf. Table 10.1).
Figure 10.2 F0 contour of the utterance “Keoi223 waa22 hai22 faai33 wai33? (S/he said it was (the word) ‘pleasant’?)”, with the pragmatic boundary tone H% attached to the final tone, indicating seeking confirmation, with probably a connotation of surprise (figure transcribed in C_ToBI; function of each tier of transcription in the C_ToBI system is noted on the right margin of the tiers).
Figure 10.3  F0 contour of two prosodic phrases “Bin55 go33 aa33[final particle]? Lei221 laa33[final particle]. (Who is it? Coming.)” marked by two phrase boundaries. Notice that the final tone of the second utterance is attached with the pragmatic boundary tone L%. Demonstrated here is a declarative use of L% (figure transcribed in C_ToBI).
Figure 10.4 F0 contour of the utterance “O223 jyun221 loi221 ha22 wai33. (Oh, I see, so it was (the word) ‘fear’.)”, with the pragmatic boundary tone HL% attached to the final tone. This use of HL% is a typical way to express connotation of “discovery” in Hong Kong Cantonese (figure transcribed in C_ToBI).
Figure 10.5 F0 contour of the prosodic phrase “Gam335 keoi223 go33 men335 (Then, his name)”, with no extra tone added after the final syllable men335. Notice that lengthening can also occur without adding any extra tones at the boundary (figure transcribed in C_ToBI).
Figure 10.6 F0 contour of the utterance "Zoi33 faam55 uk5 kei335 (And then go home) ", with truncated rise of the final syllable kei335 (figure transcribed in C_ToBI).
Figure 10.7 F0 contour of the utterance "Keoi22 jau22 waa22 hai22 faai33 wai33? (She then said it was (the word) ‘pleasant’?)", with the pragmatic boundary tone H:% attached to the final tone. The speaker asks with incredulity (figure transcribed in C_ToBI).
Figure 10.8a F0 contour of the utterance "Dim335 gaai335 lei223 sik5 keoi223 ge335[final particle]? (How come you know him/her?)". A neutral information-seeking question, where ge335 is a question particle (figure transcribed in C_ToBI).
Figure 10.8b  F0 contour of the utterance "Jyun221 loi221 lei223 sik5 keoi223 ge335[final particle] (So I see you know him/her)". With L% added after the final particle ge335, the speaker indicates that the answer of the question has been made known to her. The question particle is now said in combination with assertiveness expressed by the L tone pragmatic morpheme (figure transcribed in C_ToBI). Notice how creaky the voice quality of the L% can be (such creakiness may cause pitch tracking failure and result in pitch-doubling as can be seen at the phrase-end of this utterance).
1 See Wong (forthcoming), ‘A corpus study of the Cantonese lexicon’.

2 The choice of /35/ versus /23/ for the two rising tones follows Chao (1947). Note, however, that the literature on Cantonese is not unanimous in assigning these specific descriptors. Bauer and Benedict (1997:144), for example, argue for /25/ as the transcription of the mid-rising tone in modern Hong Kong Cantonese, on the grounds that the tone now starts at a mid-low level rather than at the middle of the speaker’s pitch range. Bauer (personal communication) suggests further that Chao’s transcription is of an older variety that distinguished an underlying /35/ from a morphologically conditioned “changed tone” (bianyin). That is, the lowered onset may be the residue of a neutralising sound change. Our own general observation is that speakers differ in their production of the two rising tones, especially in uttering the citation tones in sequence using the same syllable. For some speakers, the mid-rising tone is produced with overall higher pitch than the low-rising tone. For other speakers, however, it seems that the two rising tones have a similar starting point – mid-low – and where they differ is in slope and final tonal target; that is, the mid-rising tone rises much more sharply and ends much higher in pitch than the low-rising tone. The differences may be due in part to sociolinguistic factors, a possibility that requires further research. The C_ToBI system is intended to promote such research, by providing a means to tag phonological contrasts in spoken language corpora which are large and varied enough to provide good statistical control of sociolinguistic and discourse factors that can influence pitch range and target. That is, terms such as “/35/” should not be interpreted as phonetic transcriptions. They cannot substitute for the fundamental frequency trace or other comparably fine-grained phonetic representations. Rather /35/ versus /23/ should be understood merely as mnemonic tags into the actual speech data — a tool for keeping track of the phonological categories that are relevant when extracting finer-grained phonetic detail from quantitative models of all of the factors that affect fundamental frequency.

3 Note that in these cases, the fusion typically results in lexicalisation of the contracted form. That is, most native speakers probably would identify me55 ‘what’ as a separate lexical item.

4 Pragmatic functions of the phrase-final boundary types are suggested in the captions in the figures.

5 The “phones” symbols, being all ASCII symbols, were devised with a view to facilitating information exchange (across platforms, operating systems, etc.), database queries, and input efficiency.
ABSTRACT

This paper introduces the C_ToBI (Cantonese Tones and Break Indices) conventions for modern Cantonese. These conventions, developed within the Autosegmental-Metrical approach of the ToBI framework, are designed for use in annotating and exploring tone and juncture phenomena in spoken Cantonese corpora. Tone and juncture phenomena of especial interest for prosodic typology include: the rather strict monosyllabicity of Cantonese wordforms; the absence of contrast between “stressed” and reduced (“neutral tone”) syllables; and the extremely dense syntagmatic tonal specification, including non-segmental boundary tones. All three characteristics set Cantonese apart from Mandarin Chinese. Another phenomenon of interest is that, despite the existence of syllable fusion in Cantonese, there seems to be no reliable categorical markings of intermediate levels of prosodic grouping between the syllable and the intonational phrase, unlike in Greek, Korean, and the Wu varieties of Chinese. The C_ToBI conventions proposed here are intended to facilitate the development of the large prosodically-annotated speech corpora that are needed to address these issues and the many other outstanding questions concerning the prosodic structure of modern Cantonese.

(172 words)
INDEX

Cantonese
Chinese
Guangzhou
Zhongshan
Taishan
C_ToBI
syllable fusion
fused forms
lexical tones
boundary tones
tonal alternations
changed tones
foot tier
syllables tier
particles
final particles

KEY WORDS / PHRASES

autosegmental analysis
Cantonese
Chinese
Chinese dialects
intonation
metrical analysis
prosody
spontaneous speech annotation
syllable fusion
ToBI