

On loanword truncation in Cantonese

Kang-kwong Luke · Chaak-ming Lau

Received: 30 April 2007 / Accepted: 6 July 2008 / Published online: 9 September 2008
© Springer Science+Business Media B.V. 2008

Abstract In spite of a powerful preference for bisyllabicity identified in previous research on loanword truncation in Cantonese, more new forms are increasingly found which have been truncated down to a monosyllable. An examination of a 1,400-word corpus of Cantonese loanwords collected over a span of 50 years reveals a significant increase in the number of loan verbs and adjectives in more recent times, as opposed to the almost exclusive adoption of nouns previously. Verbs, as opposed to nouns, are found to be much more prone to undergoing “monosyllabic truncation.” This is found to stem from an asymmetry between nouns and verbs in the native language. A preference for monosyllabicity, particularly in the case of verbs and adjectives, is confirmed via a study of a Cantonese translation of the Swadesh word list. A further investigation of a corpus of everyday conversations uncovers lexical statistics that may have been mirrored in the truncation process. Finally, the greater readiness for the importation of verbs in more recent times is explained in terms of Haugen’s “stages of bilingualism.”

Keywords Loanword phonology · Truncation · Cantonese · Word length · Noun–verb asymmetry

1 Introduction

In Hong Kong, in addition to the widespread use of older loanwords such as $tʃyː^{55}kuː^{55}lik^5$ (< ‘chocolate’), $siː^{22}tɔː^{55}pɛː^{55}lej^{25}$ (< ‘strawberry’), and $wɛn^{22}nej^{55}laː^{25}$ (< ‘vanilla’) in everyday conversation, it is becoming increasingly common to hear new loanwords like $jiː^{55}$ (< ‘email’), $jiːn^{55}$ (< ‘interview’), and $p^hɔw^{22}$ (< ‘professional’). These latter words have been so drastically transformed from their

K.-k. Luke (✉) · C.-m. Lau
Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong
e-mail: kkluke@hkusua.hku.hk

originals that it is virtually impossible to recognize their sources. This radical process, whereby loanwords are heavily reduced in the course of being adapted into a recipient language, has been referred to in the literature as loanword truncation (Ito 1990; Ito and Mester 1993; Silverman 1992).

The phenomenon of loanword truncation is not new in Cantonese. Examples like sa:⁵⁵si:²⁵ (<sarsaparilla) and k^wət⁵ (<quarter) were reported as early as the 1970s and 1980s (Bauer 1985; H.S. Cheung 1972; Chan and Kwok 1982; Y.S. Cheung 1986). However, there are indications that truncation is happening much more frequently in contemporary Hong Kong than before, and as we will show in this paper, the previously sporadic appearance of truncation may well be on its way to becoming a more productive process.

Investigations of truncation forms in Cantonese—and loanword adaptation more generally—from the point of view of loanword phonology have uncovered an important principle, namely, that of bisyllabic minimality (Kiu 1977; Silverman 1992; Yip 1993, 2002). To illustrate this principle, consider a few examples from Cheung (1986) and Silverman (1992):

- (1) Epenthesis: fluke > fu:²¹lɔk⁵ cream > kei:²²li:^msa:⁵⁵
 (2) Deletion: broker > pɔk⁵k'a²⁵ freezer > fi:⁵⁵sa:²⁵

The stop/fricative-liquid clusters in word-initial position in (1) and (2) have undergone different processes of adaptation. In (1), epenthesis applies, resulting in the addition of a syllable; in (2), the liquid is deleted. This differential treatment of similar consonant clusters is due, as Silverman (1992) and Yip (1993) rightly point out, to a requirement which specifies the minimal phonological unit to be bisyllabic. In accordance with this requirement, epenthesis is applied to (1) to turn the monosyllables into bisyllabic forms. In (2), deletion is preferred over epenthesis, because its application ensures that the original words' bisyllabicity is kept intact.

The minimality requirement is at once insightful and powerful and does hold true in the great majority of cases. Nevertheless, in the course of collecting more loanword data in contemporary Hong Kong, we were struck by an apparent trend for some loanwords to be truncated down to a single syllable, e.g., words like fɛ:k⁵⁵ (< 'fax'), k^hɛp⁵⁵ (< 'copy'), and fej²¹ (< 'fail'). According to the minimality requirement, such forms should not exist.¹ The fact that they do calls for an explanation.

In seeking a principled motivation for these monosyllabic forms, we noticed that many of them are verbs and adjectives. Might their apparent "immunity" from bisyllabic minimality have something to do with their word class status? Is the truncation process sensitive to word classes (e.g., does it apply more frequently to verbs than to nouns)? The fact that these questions may be worth asking can be illustrated with an initial example. The word 'beer' was first adapted from English

¹ It must be said that minimality is no by means the only criterion in Silverman or Yip's work. Well-formedness and other constraints must also be taken into account. However, the kind of monosyllabic truncated forms that we present in this paper, e.g., in (<interview), e (<email), etc., do not seem accountable in terms of syllable structure constraints.

into Cantonese as $pe:^{55}tsəw^{25}$, the second morpheme being a generic radical meaning ‘alcoholic drink’. The internal structure of $pe:^{55}tsəw^{25}$ suggests that the source item ‘beer’ was first turned into a single syllable ‘ $pe:^{55}$ ’, to which the extra syllable ‘ $tsəw^{25}$ ’ is then added. This extra morpheme is however semantically redundant, since beer is obviously a kind of alcoholic drink. The rationale for this operation can be captured very nicely by the principle of bisyllabic minimality. However, in recent years, a new usage has emerged whereby ‘beer’ is turned jocularly into a verb, meaning ‘drink beer’, in such expressions as ‘ $pe:^{55} jət^5 pe:^{55} k^həj^{25}$ ’ (beer one beer 3.SG, i.e., ‘Let’s go and have a drink’). In this sense of $pe:^{55}$ (as an action verb), it is possible to use the morpheme independently in its monosyllabic form. Indeed, when used as a verb, only the monosyllabic form is grammatical: it is not possible to say * $pe:^{55}tsəw^{25} pe:^{55}tsəw^{25}$ (verb reduplication) or * $pe:^{55}tsəw^{25} jət^5 pe:^{55}tsəw^{25} k^həj^{25}$. If it turns out that verbs and nouns do generally behave differently in this regard, what would a plausible explanation be for this? Would the difference have its source in the first language? These questions have led us to formulate two hypotheses:

- (1) Loan verbs (including adjectives),² as opposed to nouns, are more prone to truncation to a monosyllable.
- (2) This noun–verb asymmetry has its origins in the native language, i.e., verbs as opposed to nouns have a stronger preference for monosyllabicity in the L1.

The aims of this paper then are to explore how ‘monosyllabic truncation’ works in Cantonese and to attempt an explanation for it. We believe such an account, if correct, can supplement bisyllabic minimality and help build up a fuller picture of loanword adaptation in Cantonese.

The rest of the paper will be organized as follows. In Section 2, we report on two studies that were carried out to test the effect of word class on word length. In Section 3, an explanation is offered for the word class effect in terms of a noun–verb asymmetry in the L1. Section 4 concludes the paper with a summary of the main findings.

2 The effect of word class on word length

We first tackled Hypothesis 1 by carrying out two studies: an investigation of a corpus of loanwords, followed by an experiment in which subjects were tested on their word length preferences when presented with sentences in which the same truncation form is used sometimes as a verb and sometimes as a noun.

2.1 Investigation of a loanword corpus

In the first study, a corpus of loanwords was constructed and then examined with a view to obtaining statistics relevant to the questions of monosyllabicity versus bisyllabicity on the one hand, and verbs versus nouns on the other.

² It is commonplace among Chinese grammarians to treat adjectives as a sub-class of verbs due to the many similarities that verbs have in common with adjectives in the language. See, for example, Chao (1968), Li and Thompson (1981) and Norman (1988).

2.1.1 *Constructing the corpus*

Previous studies of loanwords in Cantonese were based largely on data collected before the 1980s. It so happens that the 1980s were a turning point in Hong Kong's recent history, particularly from the point of view of linguistic and social changes. This was a decade when Hong Kong moved, within a relatively short span of time, from a largely monolingual society into a bilingual one, where Cantonese and English are widely used in many domains of everyday life.³ As the level of bilingualism continued to rise, more and more loanwords were taken into the language, making significant augmentations to the old inventory. Against this backdrop, we set out to make a collection of loanwords which would contain not only pre-1980 forms but also post-1980 ones. Our overall aim was to include as many loanwords as possible. At the same time, we tried to ensure that information associated with each item was not only accurate but also rich enough to support a range of analyses.⁴ We have included all instances of loanwords, whether they are 'established loans' or 'nonce forms' (Poplack and Sankoff 1984; Poplack et al. 1988). We believe the distinctions between these categories are hard to maintain in practice. We therefore follow Myers-Scotton (1992) in assuming that both 'Embedded Language forms' ('established forms') and 'Code Switching forms' ('nonce forms') arise from the same processes.

2.1.2 *Data sources*

Our data were collected from a variety of sources, the main ones being H.S. Cheung (1972), Chan and Kwok (1982), and Bauer and Benedict (1997). Chan and Kwok (1982) conducted the first full-length study dedicated to Hong Kong loanwords. Included in the monograph was a list of loanwords from the period from early Hong Kong (i.e., mid-nineteenth century) up until the 1970s. Cheung's collection was from about the same period and, as a result, there was a fair amount of overlap between these two sources, but there were also items found only in one but not the other source. These two sources therefore complement each other very well. With the two word lists we have a collection that is representative of an early period. Bauer and Benedict's collection was made a little later and contains items beyond those found in the two earlier works. By crossing out items which were already found in the two earlier lists, we obtained 554 items representing a more recent period from the 1970s to about 1990. In the past few years, the authors have made a further collection of more recent loanwords from newspapers, magazines and university students' speech and writing from about 1990 to the present day. Putting all the data sets together gave us a collection of 1,447 loanwords, which form the basis of the present study.

³ For more details about these social and linguistic changes, see Luke and Richards (1982), Li (1999), and Luke (2005).

⁴ Information associated with each loanword in the database include (as far as possible): (1) the original word in the source language, (2) the Cantonese adaptation (spoken form), (3) written form (if any), (4) alternative forms (if any), (5) truncated forms (if any), (6) the source of the data and (7) the time period during which the word is likely to have entered Cantonese.

2.1.3 Findings

The corpus thus constructed contains 1,447 items, with 1833 variant forms. For reasons mentioned above, the collected items were divided into old loans (which appeared before 1990) and new loans (which appeared after 1990). The distribution in terms of old and new is: 554 old loans (38.3%, with 660 variants) and 893 new loans (61.7%, with 1,173 variants). As can be seen in Table 1 below, the most striking difference between the two time periods lies in the proportion of verbs and adjectives in the corpus. In the old period, most of the loanwords are nouns, with just a small handful of adjectives and verbs (588 noun variants, as opposed to 15 adjective variants and 20 verb variants); in the new period, loanwords fall into all three major word classes: 710 noun variants, 108 adjective variants and 305 verb variants. No doubt nouns still outnumber all other word classes, but the proportion of verbs and adjectives represents a massive increase from 5% in the old period to 35% in the new period.

As far as monosyllabic truncation is concerned, are there more monosyllabic verbs than nouns in the corpus, and is word class related to word length? Table 2 shows the distribution of *monosyllabic* truncated forms in terms of word classes.

As can be seen from Table 2, the majority of monosyllabic truncated forms are indeed (as predicted by Hypothesis 1) verbs (and adjectives) rather than nouns: 44 (57.9%) versus 23 (30.3%), and this pattern is due mainly to the relatively large number of monosyllabic verb forms found in the new period. Thus, although in the corpus as a whole, there are, as one would expect, far more nouns than verbs, when it comes to monosyllabic truncation, verbs, particularly those found in the new period, figure much more prominently than nouns. A comparison between the two time periods in Table 2 shows that monosyllabic truncated forms are far more prevalent in the new period ($N = 57$) than in the old period ($N = 19$), with the difference being statistically significant (Fisher's Exact

Table 1 Distribution of loanwords by word classes in different time periods

	Noun	Verbs	Adjectives	Others	Total
Old	588 (89%)	20 (3%)	15 (2%)	37 (6%)	660
New	710 (60%)	305 (26%)	108 (9%)	50 (4%)	1,173
All times	1,298 (71%)	325 (18%)	123 (7%)	87 (5%)	1,833

Table 2 Monosyllabic truncation in different word classes

	Noun	Verbs	Adjectives	Others	Total
Old	13 (68.4%) ^a	3 (15.8%)	0 (0.0%)	3 (15.8%)	19
New	10 (17.5%)	33 (57.9%)	8 (14.0%)	6 (10.5%)	57
All times	23 (30.3%)	36 (47.4%)	8 (10.5%)	9 (11.8%)	76

^a Of the 13 "old nouns", six are names of chemical elements which are highly unlikely to be used in contexts other than a textbook or a technical report. Hence, strictly speaking, only seven monosyllabic nouns were found in the old period comparable in status to those of the new period. No chemical elements or technical terms were included in the "new" part of the corpus

test, $n = 1,833$, $df = 1$, $p = 0.0071$). In terms of word classes, all loanwords are equally prone to truncation (Pearson Chi-square test $n = 1,833$, $df = 3$, $\chi^2 = 4.604$, $p = 0.2032$), and yet the distribution of loanwords that have undergone monosyllabic truncation is distinctly skewed in favour of verbs and adjectives as opposed to nouns: 44 out of 448 (9.82%) verbs and adjectives are monosyllabic truncated forms, as opposed to 23 out of 1,298 (1.77%) nouns (Pearson Chi-square test, $n = 1,746$, $df = 1$, $\chi^2 = 58.8$, $p < 0.0001$). With only four exceptions, all truncated verbs and adjectives are monosyllabic while the majority of truncated nouns are bisyllabic. This marked difference between nouns and verbs suggests a strong relationship between word class status and word length.

To get a sense of how truncation works differently within different word classes, we present in Tables 3 and 4 below some typical examples of truncated forms. It can be seen that verbs and adjectives are mostly truncated to a monosyllable, whereas truncated nouns are mostly bisyllabic.

Interestingly, when a loanword can function either as a verb or as a noun, it is always the verb that is truncated to a monosyllable. The nouns usually take a variant form that is bisyllabic. For examples, see Table 5.

In sum, our investigation of the corpus has yielded considerable supporting evidence for our first hypothesis. Specifically, it has shown: (1) that far more verbs and adjectives have been imported from English into Cantonese in contemporary Hong Kong than previously; (2) that, if and when they undergo truncation, they are almost always reduced to a single syllable; and (3) that there is a strong correlation

Table 3 Typical verb and adjective truncations

Source word	Cantonese adaptation	Truncated form
register (v.)	>[we:k ⁵ tʃɛs ²¹ ta: ²¹]	[we:k ⁵]
professional (a.)	>[p ^h ow ²² fɛ:t ⁵ fɔn ²¹ nou ²¹]	[p ^h ow ²²]
duplicate (v.)	>[tu:p ⁵ p ^h ik ¹ kej ²¹], [tu:p ⁵ p ^h ik ¹ kej ²¹]	[tɛp ⁵]
demonstrate (v.)	>[tɛ:m ⁵⁵ mɔ:n ²¹ stf ^w h ^{ej} ²¹]	[tɛ:m ⁵⁵]
differentiate (v.)	>[ti: ²² fɛ:n ⁵⁵ ji: ²¹ ej ²¹]	[ti: ²²], [ti: ⁵⁵]
factorize (v.)	>[fɛ:k ⁵⁵ hɔ: ²¹ wa:js ²¹]	[fɛ:k ⁵⁵]
interview (v.)	>[ji:n ⁵⁵ t ^h a: ²¹ wi:w ²¹], [ji:n ⁵⁵ t ^h a: ²¹ fi:w ²¹]	[ji:n ⁵⁵]

Table 4 Typical noun truncations

Source word	Cantonese adaptation	Truncated form
introduction	>[ji:n ²² t ^h ow ²² tɛk ⁵⁵ sɔn ²¹]	[ji:n ²² t ^h ow ²²]
library	>[la:j ⁵⁵ pa: ²¹ wi: ²¹]	[la:j ⁵⁵ pa: ²⁵]
physics ^a	>[fi: ⁵⁵ sɪk ¹ si: ²¹]	[fi: ⁵⁵ sɪk ¹]
biology	>[pa:j ²² ɔ: ⁵⁵ low ²¹ tʃi: ²¹]	[pa:j ²² ɔ: ⁵⁵]
configuration	>[k ^h ɔ:n ²² fɪk ⁵ ka: ⁵⁵ we:j ⁵⁵ sɔn ²¹]	[k ^h ɔ:n ²² fɪk ⁵]
inauguration	>[ji:n ²² ɔ:k ⁵ ka: ⁵⁵ we:j ⁵⁵ sɔn ²¹]	[ji:n ²² ɔ:k ⁵]

^a ‘‘Physics’’ and ‘‘Biology’’ are from Bauer and Benedict (1997)

Table 5 Noun–verb asymmetry

Source word	Verb	Noun	Note (Verb usage/Noun usage)
reply	[wi: ²²]	[wi: ²² p ^h la: ⁵⁵]	to reply/a reply
major	[mej ⁵⁵]	[mej ⁵⁵ tfoe: ²¹]	to major in a subject/a major
minor	[ma:n ⁵⁵]	[ma:n ⁵⁵ na: ²¹]	to minor in a subject/a minor
copy	[k ^h ɛp ⁵⁵]	[k ^h ɔ:p ⁵⁵ p ^h i: ²¹]	to copy/a copy
report	[p ^h ɔ:t ⁵⁵]	[wi: ²² p ^h ɔ:t ⁵⁵]	(Chan and Kwok 1982) to report/a report
fail	[fej ²¹]	[fej ²¹ low ²⁵]	to fail (an exam)/a fail (Cheung 1972)
tips	[t ^h i:p ⁵]	[t ^h i:p ⁵ si: ²⁵]	(Cheung 1972) to give a tip/a tip (or a piece of advice)

between word class and word length, such that verbs have a stronger tendency than nouns towards monosyllabicity.

2.2 An experiment

Following the corpus study, an experiment was conducted in order to find out whether native speakers would show any word length preferences with regard to loanwords from different word classes. We believed it would be useful to obtain this kind of behavioral data to complement the corpus data.

2.2.1 Subjects

Twenty undergraduates and postgraduates from the University of Hong Kong were recruited. All of them were highly bilingual in Cantonese and English, having studied English for at least 15 years.

2.2.2 Stimuli

Thirty-six sentences were generated by using nine monosyllabic loanwords selected from the database for their ability to function both as nouns and as verbs (e.g., *pass*, *list*, and *pause*). All the nine words end with a fricative or consonant cluster which is disallowed by Cantonese phonotactics. Each loanword was used in two sentences, once as a verb and once as a noun. The pronunciation of each loanword was manipulated so that it was either epenthesized (by i: after /s/ and u: after /f/), resulting in a bisyllabic form, or not epenthesized, resulting in a monosyllabic form. Thus, a total of four sentences (two word classes × two forms) were generated from each loanword. For example, the four sentences generated for the word *pass* were given in Table 6.

2.2.3 Procedure

Each trial consisted of two sentences, which contained the same loanword in the same phonological form but used as different word-classes (such as sentences (a)

Table 6 Sample test sentences

(a) 1syll-V	tsɔŋ ⁵⁵ jy: ⁵⁵	<u>p^ha:s⁵tsɔ:²⁵</u>	kɔ: ³³	ha:w ²⁵ si: ²³	
(b) 2syll-V	tsɔŋ ⁵⁵ jy: ⁵⁵	<u>p^ha:³⁵si:²¹tsɔ:²⁵</u>	kɔ: ³³	ha:w ²⁵ si: ²³	
	finally	pass-Perf	Cl	exam	
	'(I) finally passed the exam.'				
(c) 1syll-N	kəm ⁵⁵ ts ^h i: ³³	wuj ²² ha:w ²⁵	jew ²³	lɔk ² fɔ: ⁵⁵	<u>p^ha:s⁵</u>
(d) 2syll-N	kəm ⁵⁵ ts ^h i: ³³	wuj ²² ha:w ²⁵	jew ²³	lɔk ² fɔ: ⁵⁵	<u>p^ha:³⁵si:²¹</u>
	this-time	cert-exam	have	six	pass
	'I got six passes in this certificate exam.'				

and (c) above). The stimuli were presented auditorily. Participants were asked to judge which sentence was more acceptable. The forced-choice paradigm was used so that participants had to make a choice between the two sentences. No time pressure was applied, although participants were encouraged to respond within 20 s.

2.2.4 Result

The result of the experiment shows that when the test words are used as verbs, the monosyllabic forms are significantly more acceptable (102 out of 168 accepted) than when the same words are used as nouns (78 out of 192) (Pearson Chi-square test, $\chi^2 = 14.464$, $p < 0.05$). By manipulating two variables—noun versus verb and monosyllabic form versus bisyllabic form—we were able to test native speakers' reaction to monosyllabic forms in different syntactic positions. The result of the experiment has given further support to our first hypothesis, i.e., that word class status has an effect on word length.

3 Noun–verb asymmetry in Cantonese

3.1 Noun–verb asymmetry in the L1

Thus verbs, more so than nouns, appear to have a stronger tendency to allow truncation beyond the bisyllabic template. But why should this be so? Might it have some basis in the L1? There are indications that a noun–verb asymmetry may well exist in the native language. Lü (1963) first pointed out the existence of an asymmetry in Mandarin Chinese Verb–Object combinations whereby it is possible for the verb to be monosyllabic and the noun to be bisyllabic but not the other way round, i.e., when the verb is bisyllabic and the noun monosyllabic, the result is usually unacceptable. Thus, while combinations like *mai fangzi* 'buy house' and *zuo lunchuan* 'take boat' are fine, phrases like **goumai fang* and **chengzuo chuan* are not acceptable. The same pattern holds true for Cantonese, as can be seen in the following examples, suggesting that verbs and nouns may have different preferences for word length (Table 7).

Another clue suggesting a possible preponderance for Cantonese verbs to be monosyllabic comes from an experiment carried out on native speaker preferences in the face of a choice between two synonymous verbs: one monosyllabic, the other

Table 7 Asymmetry in the internal structure of VO compounds

Mono-V + bi-N (1+2)	*Bi-V + mono-N (2+1)
kɔ:t ³³ wɔ: ²¹ ts ^h ɛ:ŋ ⁵⁵ ‘reap young-crop’ (reap the crop while it’s young: an idiom used by gamblers to mean take one’s gain and leave)	*sɛw ⁵⁵ kɔ:t ³³ wɔ: ²¹ ‘reap crop’
ti:w ³³ lɛj ²¹ ma:ŋ ⁵⁵ ‘fish small-fish’ (pick up small fish one by one: an idiom used by taxi drivers to mean pick up passengers)	*pɔw ²² tʃɔk ⁵⁵ jy: ²⁵ ‘catch fish’
fa:t ³ fa: ⁵⁵ ti:n ⁵⁵ ‘come-out madness’ (becoming mad, usually said of women who are crazy about men)	*pa:w ³³ fa:t ³ ti:n ⁵⁵ ‘explode madness’
tɛj ³³ ŋa:n ²³ mej ²¹ ‘shave eye-brow’ (shave eyebrow,: an idiom meaning to humiliate someone)	*ts ^h a:n ²⁵ ts ^h ɔj ²¹ mej ²¹ ‘remove eye-brow’

bisyllabic (Lam 2006). In this experiment, 56 subjects were asked to indicate their preference, if any, between two versions of a sentence: one containing a monosyllabic verb and the other a largely synonymous bisyllabic verb, e.g., ŋa:w³³ versus tsɛŋ⁵⁵lɛn²² ‘argue’, jɔw⁵⁵ versus ta:m⁵⁵sɛm⁵⁵ ‘worry’. The result was a clear preference for monosyllabicity in six out of the ten verbs tested. For three of the four remaining verbs, the subjects were also in favour of the monosyllabic verb, even though the difference in score between the monosyllabic and the bisyllabic versions was less marked.

The existence of nearly synonymous couplets, such as the verbs used in Lam’s experiment, is due to the structure of the Cantonese lexicon. Like most other languages, Cantonese has a lexicon that is far from uniform or monolithic in terms of the sources of the lexical items and the extent to which they are affected by different phonological constraints. A distinction, at least in principle, between the core of a lexicon and its periphery is now generally accepted.⁵ Ito and Mester (1993), for example, recognize four lexical strata in Japanese: Yamato (native), Sino-Japanese, mimetic and foreign. We believe that the Cantonese lexicon is similarly organized into four strata: native Cantonese words, Mandarin words, mimetic words (including onomatopoeia and baby talk) and loanwords. For most of the couplets that have been examined, the monosyllabic version tends to belong to the native stratum while the bisyllabic version belongs to the Mandarin stratum.⁶

We suspect that the preference for bisyllabicity as captured in the minimality requirement may not apply with equal force to verbs and nouns. The propensity for bisyllabicity identified in previous studies, based largely on data from the old period, may turn out to be less than a general tendency, in that its domain of application may be confined to loan nouns. If it is true that verbs, unlike nouns, remain to some extent exempted from the minimality requirement, then the fact that

⁵ For an alternative view, see Rice (1997) and Inkelas et al. (1997).

⁶ Mandarin Chinese has also been reported to have “dual vocabulary” (Duanmu 2007, pp. 161–65). There is, however, a major difference between Mandarin and Cantonese in this respect. In Mandarin, both the monosyllabic and bisyllabic variants are native, except that one is derived from the other either via expansion (from a monosyllabic to a disyllabic word, as in *mei* > *meitan* ‘coal’) or via truncation (from a bisyllabic word to a monosyllabic “pseudo-word,” as in *dianshi* > *dian* ‘TV’). In Cantonese, however, the monosyllables tend to be native while the bisyllabic words tend to have been imported from Mandarin.

bisyllabicity was thought to be a general tendency may be the consequence of a simple fact, namely, that loanwords in the old period were almost exclusively nouns.

Interestingly, in spite of the general preference for bisyllabicity in Mandarin, Duanmu (2000, p. 149) has also noticed the existence of some monosyllabic forms. On the basis of a word list compiled by the National Language and Script Reform Committee (1959), he found verbs to have a stronger tendency than nouns towards monosyllabicity. This tendency becomes even more prominent if a further distinction is made between basic verbs ('denoting bodily and daily activities') and modern verbs (e.g., those denoting 'political and legislative activities'). On the Reform Committee's word list, 73% of the basic verbs, as opposed to 2% of the modern verbs, are monosyllabic. There is, however, a fundamental difference between Mandarin and Cantonese in this regard. While many monosyllabic verbs are still in use in modern Mandarin, they have all come from an earlier period. As far as one can see, the old monosyllabic tendency has all but died out. Duanmu (2000) reported in the same publication a 'total lack of monosyllabic words in the new vocabulary' (p. 150). In Cantonese, however, not only are many old basic verbs still in everyday use, the monosyllabic tendency also seems to be alive and kicking. For example, new words (e.g., slang words) are created from time to time, not a few of which come in the form of a single syllable, e.g., jy:²⁵ 'embarrassed', k^hɛw⁵⁵ 'to date the opposite sex', ts^hi:w²¹ 'trendy', sɔ:k³ 'gorgeous (girl)'.

A possible scenario emerging from these considerations goes roughly as follows: at the same time as a tendency towards bisyllabicity developed in Cantonese in modern times, affecting mostly nouns, an age-old preference for monosyllabicity in verbs and adjectives remained latent in the language. This is usually not a very prominent characteristic, but as more and more verbs are adopted from English (and go through a truncation process), it comes to the fore and becomes more noticeable.

3.2 Testing the second hypothesis

We turn now to the testing of our second hypothesis, namely, that verbs as opposed to nouns have a stronger preference for monosyllabicity in the native language. In order to zoom in on the most typical native words, we took Swadesh's (1955) list of 207 words and translated them all into Cantonese.⁷ The Swadesh list was chosen because it is widely used in linguistic fieldwork to test whether words in different languages are derived from the same historical sources. For this reason, the word list was constructed to represent a language's most basic vocabulary, i.e., words which are the least likely to have been borrowed from another language. An inspection of the word list in Cantonese translation revealed that the great majority of the words are monosyllabic, with the verbs being the most strongly inclined towards monosyllabicity: 90.6% of the verbs (as opposed to 72.1% of nouns) are monosyllabic,

⁷ We realize that some of Swadesh's items can be translated in more than one way. Whenever there are alternatives, we have selected the most colloquial one. To the extent that other alternatives are possible, our statistics may not be 100% accurate. The results can be refined by double-checking the translations with more native speakers.

Table 8 Word length of a Cantonese translation of the Swadesh word list

	1 syll.	%	2 syll.	%	Total	Mean length
Nouns	62	(72.1)	24	(27.9)	86	1.28
Adjectives	34	(97.1)	1	(2.9)	35	1.03
Verbs	58	(90.6)	6	(9.4)	64	1.09 ^a

and their average length is 1.09 syllables. We therefore conclude, on the basis of the Swadesh-list study, that words in the native stratum, particularly verbs, are characterized by monosyllabicity (Table 8).

3.3 Lexical frequencies and their possible effect on loanword truncation

A final study was carried out to find out about the lexical frequencies of monosyllabic versus bisyllabic verbs in connected speech as a step towards exploring possible effects that lexical statistics may have on loanword truncation. For this purpose a 190,000-word corpus, the *Hong Kong Cantonese Corpus (HKCanCor)*, was used (Luke 2007).⁸ This is a collection of Cantonese conversations recorded in Hong Kong in the late 1990s which has been fully transcribed and grammatically annotated. An investigation of the corpus revealed that the average word lengths are 1.384 for adjectives ($N = 7,338$), 1.771 for nouns ($N = 18,682$) and 1.273 for verbs ($N = 44,612$). Within the verbs, 26.2% are bisyllabic, and 73.2% are monosyllabic. In other words, as far as verbs are concerned, monosyllabic ones outnumber bisyllabic ones in everyday speech by a ratio of about three to one. Table 9 below summarizes the full set of results of the Hong Kong Cantonese Corpus study.

Recent work on stochastic phonological knowledge has shown interesting effects of lexical frequencies on speakers' handling of new words, nonce probes and loanwords (Albright 2002; Ernestus and Baayen 2003; Hayes and Londe 2006; Hsieh and Kenstowicz, this volume). For example, Hayes and Londe have shown how knowledge about phonological patterns' frequency of occurrence appears to be internalized by native speakers, who then make use of this knowledge in deciding what patterns to use when faced with a new task, e.g., adapting a loanword. Examining the Hong Kong Cantonese Corpus allowed us to see how the noun-verb asymmetry and the monosyllabicity preference for verbs mentioned in the previous sections actually play out in connected speech. The picture that emerges from Table 9 appears to be consistent with the predictions of the theory of stochastic phonological knowledge. Further research should be carried out to ascertain the extent to which the phonological patterns found in Cantonese loanword truncation mirror lexical statistics in the L1.

3.4 Historical considerations

Having found some supporting evidence for both our hypotheses, we are now ready to consider a possible explanation of loanword truncation in Cantonese. To do this

⁸ The actual figure is 187,395 words.

Table 9 Frequencies of major word classes in the Hong Kong Cantonese corpus

	1 syll.		2 syll.		3 syll.		4+ syll.		Total
N	6,317	33.9%	10,584	56.8%	1,464	7.9%	257	1.4%	18,622
A	4,618	62.9%	2,641	36.0%	59	0.8%	20	0.3%	7,338
V	32,693	73.3%	11,670	26.2%	222	0.5%	27	0.0%	44,612

satisfactorily, we will need to situate our linguistic observations within the context of changes in Hong Kong's level of bilingualism over time. In this connection, we find Haugen's idea of "stages of bilingualism" very useful. According to Haugen (1950), three stages can be distinguished in the development of bilingual communities for the purpose of studying lexical borrowing: (1) a pre-bilingual period, (2) a stage of adult bilingualism, and (3) a period of childhood bilingualism. (Haugen 1950, pp. 216–217) The "old" and "new" time periods used in our data analysis would seem to correspond to Haugen's earlier (stages 1 and 2) and later stages (stages 2 and 3) respectively. In Hong Kong, one significant change brought about by a higher level of bilingualism would seem to have been a more ready importation of verbs and adjectives from English, as opposed to the almost exclusive importation of nouns in earlier stages. Since, as we have shown in this section, commonly used verbs have a tendency to be monosyllabic in the native language, thus verbs imported from English in more recent times would have been modelled on the native verbs when put through the truncation process. We believe this is the main reason why, in spite of the powerful minimality requirement, loan verbs have been truncated beyond the bisyllabic template to a monosyllable.

This account can also explain changes in the handling of phonotactic discrepancies from the old period to the new period, e.g., stop/fricative-liquid clusters. Recall that in the old period, there was a preference for epenthesis when words containing these clusters were adapted into the language. In the new period, however, there is an increasingly strong trend towards deletion. Thus, words like 'print' and 'click' in the old period would almost certainly have been transformed via epenthesis. In the new period, however, the output is a monosyllabic word beginning with a single aspirated stop: p^hi:n⁵⁵, k^hik⁵. The reason for this should now be clear: as verbs, these words are exempted from the usual bisyllabic requirement and have a tendency to truncate to a monosyllable. Under these circumstances, deletion is favoured over epenthesis as it will produce the right results.⁹

3.5 Word length in Cantonese and Mandarin

Before drawing this paper to a close, we wish to comment briefly on the similarities and differences between Cantonese and Mandarin with regard to word length. Duanmu (2000, 2007) offers an extensive discussion on "the word length problem" in Mandarin, by which is meant the skewed patterns observed in the

⁹ Smith (2006) discusses several examples of re-borrowings in Japanese, where the same source word takes two or more forms with different meanings, e.g., strike > *sutoraiiki* (baseball term) and later, >*sutoraiiku* (refuse to work).

use of two variant forms of “the same word” (a monosyllabic form and a bisyllabic one), such that in some linguistic contexts, only one form but not the other is possible. For example, as we have seen above, in a VO combination, it is not possible to combine a disyllabic V with a monosyllabic O, unless the noun in question is a “fixed length word,” i.e., a monosyllabic word that has no disyllabic variant. On the other hand, in a N1N2 compound, it is possible for N1 to be disyllabic and N2 monosyllabic but not the other way round (e.g. *meitan dian* ‘coal shop’ is fine but **mei shangdian* is not). To explain these patterns, Duanmu (2007, p. 146) proposes a metrical account based on stress assignment that is sensitive to syntactic functions.¹⁰ According to his “non-head stress rule,” in a syntactic construction, stress falls on the non-head as opposed to the head. Thus, in a VO combination, where the verb is the head and the object the non-head, the former is more likely to be monosyllabic while the latter bisyllabic. In contrast to this, in a N1N2 compound, the modifier (N1) is more likely to be disyllabic while the head (N2) more likely to be monosyllabic. The same asymmetrical patterns regarding VOs and NNs are found also in Cantonese.

Duanmu (2007) goes one step further and proposes to use his metrical account to explain not only the VO and NN patterns, but also discrepancies between different word classes in terms of their preferred word length. Thus, with his non-head stress rule, he believes it is possible to explain why there is a significantly higher proportion of monosyllabic verbs in Mandarin than monosyllabic nouns. According to Duanmu (2007, p. 185), ‘nouns are mostly disyllabic because they mostly occur in non-head positions, and verbs are mostly monosyllabic because they mostly occur in head positions.’ Whether this account would work for the word class asymmetry in Cantonese is not clear. In order for a proper assessment of its applicability to be made, a number of questions must first be looked into—and much more closely than is possible in the space of this paper. In the interest of further research, we will simply mention what we believe to be the four most salient questions.

First, stress is, to say the least, not very well understood in Cantonese. Since relatively little work has been done on this topic, it is not at all clear whether Cantonese has stress, and if so, how it should be defined, what forms it can take, and how it can best be treated. Clearly, in the absence of a full account of stress in Cantonese, it is impossible to properly assess Duanmu’s theory which takes stress as the starting point.¹¹

¹⁰ Duanmu’s account is actually rather more complex than the summary given here, involving as it does not only the non-head stress rule but also foot structure, information structure and other mechanisms. For want of space and for the purposes of the present discussion, the simplified account given here should hopefully suffice.

¹¹ The first remarks regarding stress were made by Chao in his 1947 book *Cantonese Primer*. His view is that there is no stress in Cantonese. Cheung (1986) and Bauer and Benedict (1997) give overall accounts of Cantonese phonology and remark briefly on stress or metrical structures. Luke, Lee and others explore the complex interactions between stress on the one hand, and F0, duration and intensity on the other (Luke et al. 2001, 2002). Lai (2002, pp. 101–118) proposes to handle stress with an iambic foot structure. Bauer et al. (2004) report on a study of the acoustic properties of focus stress in the context of a sentence. However, in spite of these efforts, there is still no general agreement among phonologists as to whether there is stress in Cantonese and, if so, how it should be defined and treated.

Second, the nature of “dual vocabulary” appears to be fundamentally different in the two varieties. In Mandarin, whatever the historical relationship may be between a word’s two (or more) variant forms, one can assume that they are both native forms of the same language. In Cantonese, however, as has already been pointed out above, the monosyllabic forms tend to be native forms, but the bisyllabic ones have mostly been imported from Mandarin. As a result, the possibility that different phonological mechanisms may have been at work in producing the different variant forms cannot be ruled out. Indeed, whether the different forms should be regarded as “variant forms of the same word” becomes a tricky question.

Third, while the noun–verb asymmetry may look, on the face of it, somewhat similar in Cantonese and Mandarin, there are significant differences between them which should make one cautious about making generalizations across the two varieties. For example, according to Duanmu (2007, p. 185), “[while] in the new vocabulary [in Mandarin] both verbs and nouns are mostly disyllabic,...in the old vocabulary most verbs are monosyllabic while nouns are usually disyllabic.” This is clearly not the case in Cantonese, where one finds new verbs and adjectives that are monosyllabic, as well as old nouns that are monosyllabic (as in the Swadesh list above). Indeed, the percentage of monosyllabic nouns in the Swadesh list (72%) is incomparably higher than the percentage of monosyllabic nouns found on Duanmu’s list (16%) (2007, p. 182). Thus, the possibility cannot be ruled out that Cantonese may have a word length “design” that is fundamentally different from Mandarin’s.

Finally, it seems to us that Duanmu’s treatment of adjectives may need looking into more closely. In his 2007 book, adjectives (in Mandarin) are said to be mostly bi- and poly-syllabic (41% of them are found to be monosyllabic, so presumably 59% are bi- or poly-syllabic). This is considered unproblematic for Duanmu as adjectives are regarded as typically serving non-head functions (i.e., as modifiers of nouns). In our own study, however, adjectives were found to be much more similar to verbs than to nouns: they have at least as strong a tendency, if not stronger, than verbs towards monosyllabism. Since adjectives occur at least as often in a predicate position (head) as a modifier position (non-head), our finding would seem to be entirely in line with Duanmu’s theory, but it would go against Duanmu’s own analysis of Mandarin adjectives, which for him should be predominantly disyllabic. Whether there are fundamental differences between the phonology and syntax of adjectives in Cantonese and Mandarin is, in our view, yet another question awaiting further investigation.

Because of space limitations, it is not possible for us to go into a more in-depth discussion of any of the above questions, which must be left to further research.

4 Conclusion

This paper began with some observations about loanword truncation in Cantonese; in particular, the truncation of newer loanwords to a single syllable. These monosyllables represent something of an exception to the well-known minimal

word requirement reported in previous publications which specifies a preference for bisyllabicity. Through a corpus investigation and an experimental study, a word class effect on word length was identified, i.e., verbs as opposed to nouns were found to be more likely to undergo monosyllabic truncation. An account of this new phenomenon was then developed in terms of a noun–verb asymmetry which appears to be latent in the native language, in spite of a modern preference for bisyllabicity established under the influence of Mandarin. In order to test this hypothesis (of a traditional preponderance towards monosyllabicity), a study was carried out using Swadesh's list of 207 words, the results of which confirmed the hypothesis. A further investigation was carried out on the Hong Kong Cantonese Corpus which uncovered lexical statistics that might have been mirrored in the truncation process. Finally, a diachronic perspective was adopted to explain changes in loanword adaptation patterns in terms of Haugen's stages of bilingualism.

Acknowledgements We would like to thank the Hong Kong Research Grants Council (HKU 7271/03H) and the State Key Laboratory for Brain and Cognitive Sciences, University of Hong Kong for their support in carrying out this research. An earlier version of this paper was presented at the ICEAL conference held at the University of Toronto in November 2006. We would like to thank participants at that conference for their questions and comments. We would also like to express our appreciation to the editors of this special volume and an anonymous reviewer for making very perceptive and helpful comments on earlier drafts of the paper. In addition, the first author would like to thank Nanyang Technological University for providing excellent facilities, and for colleagues in the Linguistics programme for their hospitality and support during his stay in 2006–2007, when this paper was written.

References

- Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78(5): 684–709.
- Bauer, Robert S. 1985. The expanding syllabary of Hong Kong Cantonese. *Cahiers Langues de l'Asie Orientale* 14: 99–111.
- Bauer, Robert S., and Paul K. Benedict. 1997. *Modern Cantonese phonology*. Berlin, New York: Mouton de Gruyter.
- Bauer, Robert S., Kwan-hin Cheung, Pak-man Cheung, and Louise Ng. 2004. Acoustic correlates of focus-stress in Hong Kong Cantonese. In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001, Tempe*, 29–49. Arizona State University Program for Southeast Asian Studies Monograph Series Press.
- Chan, Mimi, and Helen Kwok. 1982. *A study of lexical borrowing from English in Hong Kong Chinese*. Hong Kong: Hong Kong University Press.
- Chao, Yuen Ren. 1947. *Cantonese Primer*. Cambridge, MA: Harvard University Press.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Cheung, Hung-nin Samuel. 1972. *Xianggang Yueyu Yufa De Yanjiu* [A study of the grammar of Hong Kong Cantonese]. Hong Kong: Chinese University Hong Kong.
- Cheung, Kwan-hin. 1986. *The phonology of present-day Cantonese*. Dissertation, University College London.
- Cheung, Yat Shing. 1986. Xianggang Guangzhouhua Yingyu Yinyi Jieci de Shengdiao Guize [On the tone system of loanwords from English in Hong Kong Cantonese]. *Zhongguo Yuwen* 1: 42–50.
- Duanmu, San. 2000. *The phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duanmu, San. 2007. *The phonology of Standard Chinese* (2nd edition). Oxford: Oxford University Press.
- Ernestus, Mirjam, and Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79: 5–38.
- Haugen, Einar. 1950. The analysis of linguistic borrowings. *Language* 26: 210–231.

- Hayes, Bruce, and Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23: 59–104.
- Hsieh, Feng-fan, and Michael Kenstowicz. 2008. Phonetic knowledge in tonal adaptation: Mandarin and English loanwords in Lhasa Tibetan. *Journal of East Asian Linguistics*. doi:10.1007/s10831-008-9027-7.
- Inkelas, Sharon, Orhan Orgun, and Cheryl Zoll. 1997. The implications of lexical exceptions for the nature of grammar. In *Derivations and constraints in phonology*, ed. Iggy Roca, 393–418. Oxford: Oxford University Press.
- Ito, Junko. 1990. Prosodic minimality in Japanese. CLS 26-II: *Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239.
- Ito, Junko and Armin Mester. 1993. Japanese phonology: Constraint domains and structure preservation. In *The handbook of phonological theory*, ed. John Goldsmith, 817–838. Cambridge, MA: Blackwell.
- Kiu, K. L. 1977. Tonal rules for English loan words in Cantonese. *Journal of the IPA* 7(1): 17–22.
- Lai, Esther Yuk Wah. 2002. *Prosody and prosodic transfer in foreign language acquisition: Cantonese and Japanese*. Muenchen: Lincom Europa.
- Lam, Chu Shan Erica. 2006. Monosyllabism in Cantonese: A study of Cantonese verbs. Ms., Final year essay, Department of Linguistics, University of Hong Kong.
- Lee, Wai-sum, Fangxin Chen, K.K. Luke, and Liqin Shen. 2002. A phonetic study of the prosodic properties of bisyllabic, trisyllabic, and polysyllabic compounds in Hong Kong Cantonese. In *Proceedings of the 1st International Conference on Speech Prosody*, April 2002 in Aix en Provence, France.
- Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, David. C. S. 1999. The functions and status of English in Hong Kong: A post-1997 update. *English World-Wide* 20(1): 67–110.
- Lü, Shuxiang. 1963. Xiandai Hanyu and shuang yinjie wenti chu tan [A preliminary study of the problem of mono- and disyllabic expressions in modern Chinese]. *Zhongguo Yuwen*, 1963(1): 11–23.
- Luke, Kang Kwong. 2005. Cong shuangyu he shuangyan kan Xianggang shehui yuyan bianqian [Sociolinguistic Changes in Hong Kong: Bilingualism and Diglossia]. *Journal of Chinese Socio-linguistics* 1: 82–89.
- Luke, Kang Kwong. 2007. *Hong Kong Cantonese Corpus*. <http://www.hku.hk/hkcancor/>. Accessed 15th April 2007.
- Luke, Kang Kwong and J. C. Richards. 1982. English in Hong Kong: Functions and status. *English World-Wide* 3(1): 47–64.
- Luke, Kang Kwong, Fangxin Chen, Wai-sum Lee, and Liqin Shen. 2001. A phonetic study of the prosodic properties of bisyllabic compounds in Hong Kong Cantonese. In *Proceedings of the Conference on Phonetic Sciences*, October 2001, Beijing.
- Myers-Scotton, Carol. 1992. Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development* 13: 19–39.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Poplack, Shana, and David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22: 99–135.
- Poplack, Shana, David Sankoff, and Christopher Miller. 1988. The social and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104.
- Rice, Karen. 1997. Japanese NC clusters and the redundancy of postnasal voicing. *Linguistic Inquiry* 28: 541–551.
- Silverman, Daniel. 1992. Multiple scansion in loanword phonology: Evidence from Cantonese. *Phonology* 9: 289–328.
- Smith, Jennifer. 2006. Loan Phonology is not all perception: Evidence from Japanese loan doublets. *Japanese/Korean Linguistics* 14: 63–74.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21: 121–137.
- Yip, Moira. 1993. Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics* 2: 261–291.
- Yip, Moira. 2002. Perceptual influences in Cantonese loanword phonology. *Journal of the Phonetic Society of Japan* 6: 4–21.