Bulletin of the *Transilvania* University of Braşov
Series IV: Philology and Cultural Studies • Vol. 13(62) Special Issue
https://doi.org/10.31926/but.pcs.2020.62.13.3.7

Reconstructing Chinese Pidgin English phonology on the basis of written sources

Ioana MIRON[1]

This paper is an attempt at outlining the phonology of Chinese Pidgin English, including its syllable structure, with an emphasis on the onset and the coda. Since Chinese Pidgin English is an extinct variety, the only available sources are written records such as magazine articles (e.g. in The Chinese Repository), literary works, travelogues, and letters. Reconstructing the phonology of Chinese Pidgin English on the basis of the orthography used in these sources raises the issues of the reliability of the sources and of the methodological implications. These are addressed in light of the caveats formulated by Mühlhäusler (1997), Baker and Winer (1999), Avram (2000), among others. In line with the principle of sociolinguistic accountability, all tokens in the samples of Chinese Pidgin English are included in the analysis as well as all the contexts where they might have appeared. In addition, a comparison will be made with other contemporary records of Chinese Pidgin English, with the phonology of Hong Kong English (Setter et al. 2010) and with the L2 phonology of Chinese learners of English. Such a comparison is certainly not the perfect equivalent of Rickford's (1986) "feedback from current usage", given that Chinese Pidgin English is no longer spoken. However, this approach is warranted by the so-called "uniformitarian principle" (Labov 1972), which posits that current patterns are similar to those that operated in the past.

Key-words: *Chinese Pidgin English, records, phonology*

1. Introduction

Chinese Pidgin English (henceforth CPE) no longer has a community of speakers. As such, reconstructing the phonology of this variety is possible only on the basis of textual attestations. This type of pursuit is not singularly applicable to CPE, but to other varieties as well, either for their historical reconstruction (e.g. Crowley 1993, Rickford 1995) or because they are extinct. Consequently, a sizeable body of

[1] University of Bucharest, ioanam1595@live.com

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

literature is available, outlining the advantages, caveats, and possible methodologies of such a linguistic analysis.

Before analyzing the data, it is first necessary to establish the degree of authenticity and reliability of such sources. These qualities are equally important, as it is possible for an authentic text not to be reliable, as is the case with many available sources of CPE or, more rarely, for a reliable document not to be fully authentic. Rickford's (1986) summarizes four criteria to have in mind: the author, the document, sociolinguistic accountability, and feedback from current usage. Other linguists have shed new light on these, but also added new criteria (Baker and Winner (1999), Crowley (1993), Lalla and D'Costa (1990), Mühlhäuser (1997), Avram (2000)).

As far as the author is concerned, we must primarily take into consideration their L1 and time of residence in the territory where the pidgin or creole is spoken. Baker and Winer (1999) distinguish between internal and external authors, and consider length of residence a crucial factor in their assessment. "Insiders" or "outsiders" with a minimum of five years' residence (Baker and Winer 1999, 105) are considered the most reliable authors. Other factors include level of education, profession and even, to a lower degree, their attitude (neutral, prejudicial or preferential) towards the speakers or the pidgin varieties. However, linguists should not be themselves prejudiced into favouring data provided by impartial authors in the detriment of those with negative attitudes towards the pidgin/creole speakers or languages.

A short history of the document (author, editors, number of editions, intended use and targeted audience) can also provide evidence in favor or against authenticity and reliability.[2] More important remains the content of the document and its assessment revolves around consistency – in spelling, syntax or meaning – at the level of a single document or when comparing with contemporary sources. This approach is widely supported, being discussed in Hancock (1977), Rickford (1991), Mühlhäuser (1997), and Avram (2000) among others. In our selection of primary sources, we also heavily relied on this, with the mention that especially in the case of pidgins, inconsistencies may indicate a particular stage in the development of the variety and not necessarily unreliability.

Sociolinguistic accountability (Labov 1972), posits that linguistic features be analyzed in all the contexts where they appear in a given source, and all other possible contexts of occurrence. A qualitative analysis of the data yields a more

[2] For example, the introduction of Hunt (1931, 5) states as reason for publication "the wholesome fun that is to be obtained from 'Babu' letters, […] this collection preserves a few typical specimens that more may share their humour". The editor's approach implies a filtering process for the materials, whereby only those deemed most humorous or deviant from Standard English were included in the collection.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

comprehensive and scientifically accurate linguistic image when corroborated by a quantitative analysis.

Feedback from current usage translates into a comparison with the modern-day version of the respective variety, helping with the interpretation of the data from the textual attestations. With CPE an extinct variety today, we can argue for a soft variant of feedback from current usage by comparing CPE with Hong Kong English or the English variety spoken today by L1 Chinese natives. Such an approach is warranted by Labov's uniformitarian principle, according to which current patterns are similar to those that operated in the past.

This paper is structured as follows. Section 2 argues for the selection of the texts listed under Primary Sources. Section 3 is concerned with the phonological reconstruction of CPE, with an emphasis on the consonantal system, and referencing the phonology of Hong Kong English and L1 Chinese – L2 English speakers. The syllable structure of CPE is discussed in Section 4. Section 5 presents the conclusions.

2. Selection of primary sources

The primary sources on which this discussion is based were selected for the following reasons: reliability, richness of data and the period when the data was collected. They are consistent from a document-internal, but also document-external point of view – a large proportion of lexical items (and syntactic structures) being corroborated by other contemporary sources and not only, of the same type or different type of texts.

Leland's glossary (1876) is validated by contemporary sources such as Hunter (1911): *belongey / belong*, *chop* (in *first chop*), *flin / flen* 'friend', *foolo* 'fool', *loose / loosum* 'lose', *numpa /no one* 'number one' (with the meaning of 'the best'), *olo* 'old', *one piece* (used as a classifier or indefinite article), etc. The majority of lexical items, with their multitude of meanings and idiosyncrasies of usage, are attested in travelogues across a century back. For example, *alla / allo* 'all', *hab got*, *catchee*, in Anderson (1795); *chin-chin* 'worship, adore, reverence, wish', *handsome, Joss* in Noble (1762); *have/hab* 'is', *got* 'have' (Keate 1788); *tinkee* 'think', *come, pigeon/pidgin* 'business', *Mandarin* (in Leland also in the variant *Mandalin*) in Hall (1826), etc. Airey's glossary (1906) is also consistent with Leland (1876).[3]

[3] Both Leland (1876) and Airey (1906) are in the category of "literary" sources of CPE – they are supposed to represent songs, poems and ballads written in this pidgin variety. While their glossaries are authentic and reliable, corroborated by other sources, the actual songs and ballads can be argued, on sociolinguistic grounds, to be the product of European authors. Arguments

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Denby (1911), a collection of letters and sketches, is also corroborated by contemporary sources, such as Hunt (1931): *can, must, chop, fault, make, man, me* 'I', etc. We find 19% of the vocabulary in older texts of this type, i.e. Anon. (1843), which is a good percentage, considering that the latter is poorer in CPE examples, and 35% lexical items in unrelated sources such as Leland (1876).

The majority of CPE sources come in the form of travelogues – diaries or memoires of British sailors, soldiers or diplomats travelling to the East. While this type of texts clearly surpasses in number literary pieces, letters or editorials, they are generally poor sources of CPE, with an average of 43 lexical items (not distinct) per thousands of pages. Moreover, sometimes they are unreliable, inconsistent in spelling (which poses difficulties for a phonological analysis), undated and possibly misleading in quotations.[4] Exception to this, from the point of view of richness of data, are Downing (1838) and Hunt (1931) from the travelogues analyzed so far, to which I will refer tangentially in the discussion below. The selected sources, although scarce in numbers, are in fact very rich in data and present the advantage that even inconsistencies, where sporadically present, are systematic across the same document.

Lastly, all five sources date starting with the latter half of the 19th century, when the CPE variety had arguably stabilized. Proof of this comes from the fact that even the European authors cease to refer to CPE as a barbaric tongue or an offense to the standard variety, but acknowledge its status: Anon. (1843, 80) calls it "Anglo-Chinese", Denby (1911, 376) quotes: "[…] using Pidgin English for lack of a better exchange". We can find another argument in the fact that Hunt (1931) corroborates 56% of the lexical items found in Anon. (1843). Such a high percentage of lexical items shared by sources a century apart confirms the fact that with the latter half of the 19th century CPE is in its most stable form. Lastly, after the second Opium War, with commercial trade being no longer confined to Canton, CPE travels upwards the eastern Chinese coast (reaching even Japan - Denby 1911, 376), serving as lingua franca between foreigners and Chinese, but also reportedly between Chinese speakers of mutually unintelligible dialects.

include the fact that the authors do not indicate how they collected the literary works, the fact that a pidgin is an unlikely choice to express poetic inclination, being nobody's mother tongue and restricted primarily to commercial interactions, and the fact that the Chinese viewed the British as red-haired barbarians with an inferior culture. For a larger discussion, see Miron (in press).

[4] For example, Hall (1826, 343), who attributes lexically and syntactically complex English productions to a Canton governor, in quotation marks, whereas upon closer analysis, it can be inferred that it is actually the work of an interpreter present in the room or simply the author's paraphrase.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

3. The phonology of Chinese Pidgin English

3.1. Consonants

Apart from the general considerations to bear in mind when assessing the authenticity and reliability of a text, each area of linguistic analysis (phonology, morpho-syntax, vocabulary, semantics, etc.) poses its unique challenges. One of the most common and, at the same time, most frequently discussed in the literature, as far as phonology is concerned, is anglicization. This refers to the tendency of European authors to employ English spelling rules when documenting samples of pidgin or creole varieties and, more importantly, to neutralize particular phonological features based on their L1. Among the root causes for this phenomenon, we can list "cultural blinkers" (Bickerton 2008, 213-214). It represents an inclination on the authors' part to record or even only perceive those features that are already familiar, and to discard features that are strange or unintelligible to them. Baker and Winner (apud. Avram 2015, 22) discuss the reverse situation, namely the possibility of what is most different or "bizarre" to be all the more salient and predominantly documented in early pidgin and creole records, resulting in a distorted image of the variety. The authors of the five primary sources selected are either certainly or arguably L1 English speakers[5]. Consequently, one or more of the risks discussed above may partially apply to their samples of CPE and must be taken into consideration for a phonological analysis.

Given the greater impact of articulators, consonantal sound qualities are more easily perceptible than vocalic qualities, even to the untrained or linguistically-naïve ear. Table 1 lists the consonants of three varieties that I will use as reference. The first is Cantonese (CAN), the native language of most speakers of CPE. While not solely responsible for the phonological system of pidgins and creoles, substrate influence has received due credit in the literature so far (Sebba, 1997). The second variety, Hong Kong English (HKE), is a soft variant of feedback from current usage, together with Chinese English (CE) as spoken by Mandarin natives.[6]

Table 1 lists the consonantal system of these varieties, according to Chan and Li (2000), Setter et al. (2010), and Mangalea (2019) for CAN, HKE and CE, respectively. This overview is used primarily for predictions that may be validated or invalidated by the data collected from the selected primary sources.

[5] Anon. (1836) was published in the Chinese Repository, which had only American editors and all contributors (anonymous as per the practice at the time) are understood to be native speakers of English. Anon. (1843) and Leland (1876) are British, whereas Denby (1911) is American, leaving Airey (1906) unverified, but presumed also to be an English native, judging by the introduction.

[6] In light of the fact that after 1850, CPE was no longer confined to Canton and the speakers might not all be native speakers of Cantonese.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 1. The consonantal systems of Cantonese, Hong Kong English and Chinese English

		Plosive	Nasal	Fricative	Affricate	Approximant
Bilabial	**CAN**	/p/, /pʰ/	/m/			
	HKE	/p/, /b/*	/m/			
	CE	/p/, /b/*	/m/			
Labiodental	**CAN**			/f/		
	HKE			/f/, (/v/)		
	CE			/f/, ?		
Dental	**CAN**					
	HKE			/θ/, (/ð/)		
	CE			(/θ/)		
Alveolar	**CAN**	/t/, /tʰ/	/n/	/s/	/ts/, /tsʰ/	/l/
	HKE	/t/, /d/*	/n/	/s/, /z/*	/tʃ/, /dʒ/*	/l/, /ɹ/*
	CE	/t/, /d/*		/s/	/ts/, ?	/l/, /ɹ/*
Postalveolar	**CAN**					
	HKE			/ʃ/, /ʒ/*		
	CE			/ʃ/, ?		
Palatal	**CAN**					/j/
	HKE					/j/
	CE					/j/
Velar	**CAN**	/k/, /kʰ/	/ŋ/			
	HKE	/k/, /kʰ/	/ŋ/			
	CE	/k/, /kʰ/	/ŋ/	/x/		
Labiovelar	**CAN**	/kw/, /kʰw/				/w/
	HKE					/w/
	CE					/w/
Laryngeal	**CAN**			/h/		
	HKE			/h/		
	CE			/h/		

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

3.1.1. Plosives

Regarding plosives, one prediction borne out by the data is the presence of voiceless stops shared with RP[7], i.e. /p, t, k/, especially in word-initial position, as can be seen the examples under (1), (2) and (3). This accords with HKE and CE as well, stemming from the frequency and distribution of these sounds in CAN.

(1) a. ***p****idgeon* 'business', ***p****iece*, ***p****lease*, ***p****utee* 'put'
Anon. (1836)

b. ***p****iece*, ***p****idgeon* 'business'
Anon. (1843)

c. ***p****apaman* 'barber', ***p****iecee* 'piece', ***p****lopa* 'proper'
Leland (1876)

d. ***p****idgin* 'business', ***p****lizzon* 'prison', ***p****ylong* 'thief'
Airey (1906)

e. ***p****ay,* ***p****lopper* 'proper', ***p****utee* 'put'
Denby (1911)

(2) a. ***t****ele* 'three', ***t****ime*, *to-day* 'today'
Anon. (1836)

b. ***t****alkee* 'talk', ***t****o*
Anon. (1843)

c. ***t****lee* 'three', ***t****im* 'time', ***t****op-side* 'up, upside, above', ***t****inkee* 'think'
Leland (1876)

d. ***t****opside* 'up, upside, above'
Airey (1906)

e. ***t****alkee* 'talk'
Denby (1911)

[7] RP is taken as the primary term of comparison because historically, the British were the second major influence on China (commercially and linguistically) after the Portuguese and on a par with the Dutch. American English may also have had an impact, although minor, seeing as they arrived later (Denby (1911) is the most recent selected source and has an American author) and most importantly, at a stage when CPE will have stabilized. On Hong Kong English, however, American influence is "on the increase" (Setter et al. 2010, 28).

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

(3) a. ***c**all-um* ‘call’, ***c**asion* ‘occasion’, ***c**ome*, *coolie* ‘servant, laborer’, ***c**umshaw* ‘gift, bribe’

Anon. (1836)

b. ***c**an*, ***c**ome*, ***c**umshaw* ‘gift, bribe’

Anon. (1843)

c. ***c**allee* ‘curry’, ***c**asion* ‘occasion’, ***c**atchee* ‘catch’, ***k**leen* ‘green’

Leland (1876)

d. ***c**ash* ‘copper currency’, ***c**ompladore* ‘compradore’, ***k**imono*

Airey (1906)

e. ***c**an*, ***c**atchee* ‘catch’, ***k**ick*

Denby (1911)

In Cantonese, /p, t, k/ appear not only in word-initial or word-medial (onset) position, but also in word-final position (unlike Modern Chinese[8], which only allows /n, ŋ/ in coda position). While the data corroborate syllable-initial distribution (see (4)), plosives in the coda are not as consistent as can be seen in (5):

(4) a. *pro**p**er, be**tt**er, Can**t**on, mas**k**ee* ‘nevermind’, *ta**k**ee* ‘take’

Anon. (1836)

b. *em**p**eror, slee**p**y be**tt**er, fo**k**y* ‘friend’, *tal**k**ee* ‘talk

Anon. (1843)

c. *plo**p**a* ‘proper’, *tal**k**ee* ‘talk’, *co**c**o* (unit of measure for weight), *dlin**k**ee* ‘drink’, *fai**t**ee* ‘quickly’, *fo-**t**in* ‘fourteen’

Leland (1876)

d. *com**p**ladore* ‘compradore’, *mas**k**ee* ‘nevermind’, *spe**c**ulum* ‘spectacles’, *fan**t**an* (gambling game), *li**tt**y* ‘little’

Airey (1906)

e. *plo**pp**er* ‘proper’, ma**k**ee ‘make’, mas**k**ee ‘nevermind’, mas**t**er, wa**t**er

Denby (1911)

(5) a. *cho**p*** ‘document, receipt’, *Eulo**p*** ‘Europe’, *te-lu**p*** ‘twelve’, *thin**k***, *si**k**-she* ‘six’*, go**t**, las**t**, jus**t***

Anon. (1836)

b. *figh**t**, ge**t**, ou**t**, ma**k**e*

Anon. (1843)

c. *cho**p*** ‘document, receipt’, *fa**k**e, ku**k**-man* ‘cook’, *co**t**-houso* ‘courthouse’, *hear**t***

Leland (1876)

[8] By Modern Chinese we understand 普通话 *Putonghua*, the standardized variety in mainland China, based in proportion of 80% on the northern, Mandarin dialect.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

d. *si**p**, to**p**side* 'up, upside, above', *chi**t*** (a form of I.O.U.), *li**ck**sha* 'rickshaw', *ya**k*** ('one' in Cantonese)

Airey (1906)

e. *cho**p*** 'document, receipt', *captain*[9], *faul**t**, go**t**, ou**t**, kic**k**, mista**k**e, sic**k***

Denby (1911)

One important note here is that while /p, t, k/ are to be expected in onset position, their distribution as such in the examples above reveals a common strategy in CPE of changing the English coda ending in /p, t, k/ into onsets by adding a paragogic vowel (e.g. *tal**k*** → *tal-**k**ee* most probably pronounced [tɔkɪ]). This happens even for single-consonant situations, not only in the case of clusters, which indicates a more restricted coda in CPE than CAN or even HKE. When they do appear in the coda, it can be observed that it is often in words of Cantonese origin (*chop*, *yak*), as repair strategies (*sik-she* instead of [sɪks]) or in arguable anglicized and unreliable forms (e.g. *heart* or even *out*, which inconsistently appears in Leland (1876) as *au-(sai)* and as *out(side)*).

The voiced plosives /b, d, g/ open up a larger discussion regarding voicing contrasts, relevant to all Chinese-substrate varieties with an English lexifier. This is because in Mandarin, Putonghua or Cantonese, voicing is not a phonologically distinctive feature, but aspiration is. In English, aspiration renders only allophones, not minimal pairs. Setter et al. (2010, 13) mention that:

> Hong Kong English clearly has a voicing contrast. This contrast is not unlike that of other varieties such as RP and General American, in that the chief difference between voiceless and voiced phonemes is the aspiration of voiceless phonemes.

Based on voice onset time (VOT), aspiration is empirically correlated to voice as follows: voiceless (aspirated) phonemes have greater values for VOT, whereas voiced (unaspirated) phonemes have much lower, possibly negative VOT values. This distinction can be clearly captured on a spectrograph, as opposed to voicing, which is more difficult to measure. This is why voiced plosives in HKE are recorded in Table 1 with an asterisk signaling this interpretation of the voiced sounds. At a first glance, in CE, /b, d, g/ are attested in syllable-initial distribution, and /b, d/ also in syllable-final position, as can be seen in (6) and (7), respectively:

[9] This word also occurs in Leland (1876) as *cabtun* seemingly replacing the voiceless bilabial stop with its voiced counterpart. This can be argued to be a mistake on the part of the author, and for the plosive to be in fact voiceless in this phonological context. Voice distinctions in all varieties of English with Chinese substrate are a debated topic (applicable to both HKE and CE).

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

(6) a. ***b**ack, **b**efore, **b**oy, **b**indee* 'bind', ***d**o, **d**ie, sen**d**ee* 'send', ***g**o, **g**ood, **g**ot*
Anon. (1836)

b. ***b**etter, **d**rive, man**d**arin, En**g**lish, **g**et, **g**ood, **g**ot*
Anon. (1843)

c. ***b**anjee* 'band' (of music), ***b**elongey* 'belong', ***d**o, can**d**areen* 'one penny', ***d**linkee* 'drink', ***g**lound* 'ground', ***g**ot, **g**ood*
Leland (1876)

d. ***b**o**b'b**ly* 'bobbery', ***b**oy, **d**evil, compla**d**ore* 'compradore', ***g**lyn**d**e* 'grind'
Airey (1906)

e. ***b**elong, **b**etter, **b**lains* 'brains', ***d**ie, **d**o, **d**ollar, **g**ive, **g**o, **g**ot*
Denby (1911)

(7) a. *ha**b**, trou**b*** 'trouble', *goo**d***
Anon. (1836)

b. *an**d**, goo**d**, han**d**some*
Anon. (1843)

c. *ca**b**-tun* 'captain', *si**d**e, bun**d*** 'quay', *gloun**d*** 'ground'
Leland (1876)

d. *topsi**d**e* 'up, above'
Airey (1906)

e. *be**d**room, insi**d**e, secon**d***
Denby (1911)

Upon closer analysis, however, the situation is not quite as straightforward, especially as far as the coda is concerned[10]. Firstly, many of these forms are examples of anglicization. For instance, 'slit-eye**d**' and 'soft-nose**d**', featuring /d/ in word-final position as result of a morphological process are the production of a native of the lexifier, who we cannot presume to be subject to Cantonese substrate influence in terms of voicing and aspiration. Additional argument for anglicized forms can be found in examples containing /ɹ/, especially in a cluster, as /ɹ/ rarely complies with the articulatory habits of any Chinese native.[11] Thirdly, inconsistent records give us one variant with voiced phonemes, one without them, and yet

[10] Setter et al. (2010 15) mention that in Cantonese codas, plosives are produced with "strong glottal reinforcement and not released", and that this pattern was transferred to HKE, resulting in a "perceived lack of contrast" (in voice / aspiration). Setter et al. (2010, 15) indicate that Cantonese natives "cannot actualize the systematic contrast between voiced and voiceless final plosives".

[11] Authors such as Leland (1875, 5) note "the almost insuperable tendency to pronounce *r* as *l*". Airey (1906, 9) mentions "[…] the Celestial finds it difficult to pronounce some consonants, and so he either substitutes another or omits it altogether. For instance, in the case of the 'r' he substitutes an 'l' […]. Examples from our sources supporting this include 'blains', 'dlinkee', 'glound', 'compladore' etc.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

another one with their voiceless counterparts, e.g. Leland (8176) *au-sai* for 'outside' and a plethora of forms with *side* (*bottom-side, come-this-side, top-side, inside, inside he heart, inside he mouth, longside, outside, roominside*, etc.). Lastly, by looking at those contexts where voiced plosives are distributed in RP, we discover a considerable number of examples where CPE replaces or deletes these phonemes (see (8)):

(8) a. *insi* 'inside', *joss* 'God'[12]

Anon. (1836)

b. *foky* 'friend'

Anon. (1843)

c. *banjee* 'band', *chilo* 'child', *flin* 'friend', *galanti* 'grand', *han-tun* 'hundred', *hankerchoo* 'handkerchief', *insy* 'inside', *kleen* 'green', *ko-au-sei* 'go outside', *ko-hom* 'go home', *lan-tun* 'London', *lau-tai* 'ladder', *olo* 'old', *papaman* 'barber', *pidgin* 'business', *seweitun* 'Sweden', *takta* 'doctor', *tinmak* 'Denmark', *ut* 'good'

Leland (1876)

d. *chilo* 'child', *joss* 'God'

Airey (1906)

Incidentally, note again the tendency to move the plosive in syllable-initial position by paragoge (e.g. *bindee*, *sendee*, *banjee*, *cango, galanti* 'grand', etc.). Baker and Huber (2001) include paragogic vowels among their list of worldwide features of English-lexifier pidgins and creoles and Avram (2015, 28) mentions default [i] in the analysis of Jamaican Creole.[13]

In addition to the bilabial, alveolar and velar plosives discussed so far, Table 1 also lists the labiovelar /kw, k^hw/ as present in CAN and HKE. Given that the articulatory gestures for /k/ and /w/ are simultaneous i.e. the back of the tongue touches the velum while the lips are rounded (Setter et al. 2010), they represent one sound (coarticulated consonants) with aspiration playing a phonologically distinctive role. The labio-velar plosive is attested in CPE as well, although marginally, and predominantly in words of Cantonese origin or phonetic calques:

[12] In CPE, *joss* with the meaning of God, is an example of Portuguese influence (*deus*), so that the /d/ is replaced in the onset, not the coda.

[13] Although outside the scope of the present paper, we can note that CPE displays a wide array of diagnostic features for pidgins, as listed in Baker & Huber (2001). Examples include Pacific features (*chinchin*, *chowchow*, *maski* – frequently recorded in our sources as *maskee* – *belong* for the genitive) and world-wide (*all same* for 'alike', *before time* 'formerly', *fashion* 'manner', *number one* 'best, chief', *suppose* 'if', *too much* 'a lot' – usually spelled *too muchee* – *WH fashion* for 'how', 'why', ZERO equative and predicative copula etc.).

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

(9) a. *kweisi* ‘crazy’

Anon. (1836)

b. *ahkwantsae* ‘gentleman’, *etalee kwoh* ‘Italian consulate’, *fan kwei* ‘foreign devil’, *makwa* ‘riding coat’

Leland (1876)

3.1.2. Nasals

Nasal consonants /m, n, ŋ/ are not expected to pose any difficulties for CPE speakers, given that they are attested in CAN, HKE, and CE alike. This prediction is borne out by the data, in the onset for /m, n/, and the coda for all three consonants, with examples in (10) and (11), respectively:

(10) a. *a**n**y*, ***n**oise, **n**other* ‘another’, ***n**umber, **m**an, **m**oon, **m**uchee* ‘much’

Anon. (1836)

b. ***n**o, **m**an, **m**andarin*

Anon. (1843)

c. ***n**aifoo* ‘knife’, ***n**o, **n**other* ‘another’, ***n**aiti* ‘night’, ***m**assa* ‘master’, ***m**askee* ‘nevermind’, ***m**oon* ‘month’

Leland (1876)

d. *a**m**ah* ‘woman servant’, *ki**m**o**n**o, **n**umber one* ‘best’, ***m**an**m**an* ‘slowly’, ***m**oon* ‘month’

Airey (1906)

e. ***n**o, **n**othing, **n**umber one* ‘best’, ***m**afoo* ‘broom’, ***m**akee* ‘make’, ***m**askee* ‘nevermind

Denby (1911)

(11) a. *bi**n**dee* ‘bind’, *chi**n**chi**n*** ‘worship, greetings’, *call-u**m*** ‘call’, *cu**m**shaw* ‘gift’, *lo**ng**, fi-te**ng*** ‘fifteen

b. *ca**n**, pidgeo**n*** ‘business’, *co**m**e, cu**m**shaw* ‘gift’, *handso**m**e*

Anon. (1843)

c. *casio**n*** ‘occasion’, *wa**n**chee* ‘want’, *fala**n**sai* ‘French’, *ko-ho**m*** ‘go home’, *ti**m*** ‘time’, *spillu**m*** ‘spoiled’, *ti**ng*** ‘thing’

Leland (1876)

d. *ma**n**ma**n*** ‘slowly’, *moo**n*** ‘month’, *sa**m**pan* ‘boat’, *speculu**m*** ‘spectacles’, *pylo**ng*** ‘thief’

Airey (1906)

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

e. *fashio**n**, folli**n*** ‘foreign’, *wa**n**chee* ‘want’, *co**m**e, nu**m**ber, Sha**ng**hai, thi**ng***
Denby (1911)

We can briefly note here that /n/ frequently remains alone in the coda after paragoge (discussed in section 3.1.1). This phenomenon, apart from preserving syllable weight, is also employed to resolve consonantal clusters (e.g. wanchee → wan-chee instead of ‘want’, or sendee → sen-dee instead of ‘send’). This can be interpreted as a repair strategy imposed by syllable structure constraints transferred from Cantonese, which does not allow liquids in the coda. /ŋ/ appears in words of both Cantonese and English origin, and although Cantonese allows it in the onset as well (e.g. 我 /ŋɔ/), the selected sources seem to attest it only in the coda of CPE syllables.

3.1.3. Fricatives

As a starting point, we can note that based on the overview offered by Table 1, English fricatives surpass in number those in CAN, and even those attested in HKE and CE. Moreover, except for the velar /h/, all other classes of fricatives (labio-dental, dental, alveolar, postalveolar) come in pairs of voiced and voiceless phonemes in RP. Our first prediction is that excluding interdental /θ, ð/ and postalveolar /ʃ, ʒ/ (which are not available in CAN), voiceless fricatives /f, s, h/ would not pose any difficulties in onset position (see (12)):

(12) a. *be**f**ore, **f**i* ‘five’, *f**f**i-teng* ‘fifteen’, *wi**f**o* ‘wife’, ***s**ame, **s**ileek* ‘silk’, ***s**pose* ‘suppose, if’, ***h**ab* ‘have’, ***h**e*, ***h**ow*, ***h**ouse*
Anon. (1836)

b. ***f**ace, **f**ight, **f**oky* ‘friend’, ***f**etchee* ‘fetch’, ***s**ay, **s**ell, **s**’pose* ‘suppose, if’, ***h**ave, **h**e, **h**ow*
Anon. (1843)

c. *be**f**ore-tim* ‘formerly’, ***f**ai* ‘five’, ***f**acee* ‘face’, *ma**f**oo* ‘groom’, *all fa**c**ee* ‘face’, *con**s**oo* ‘consul’, ***f**a**ss**on* ‘fashionable’, ***h**ab* ‘have’, ***h**ow, **h**ankerchoo* ‘handkerchief’
Leland (1876)

d. ***f**antan* ‘gambling game’, ***f**ollin* ‘foreign’, ***s**am* ‘three’, ***s**ee, **s**ilver shoe* ‘form of payment’
Airey (1906)

e. ***f**ashion* ‘way, manner’, ***f**ollin* ‘foreign’, ***f**ive, o**ff**ice, all **s**ame* ‘alike’, *in**s**ide, **s**ick, **s**o, **h**ave, **h**e, **h**ow*
Denby (1911)

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

However, Anon. (1836) and Leland (1876) offer evidence that /f/ is attested not only in expected contexts of occurrence, but also in the onset of syllables resulting after paragoge, e.g. *waifo* → wai-fo, *naifoo* → nai-foo. While fricatives are not attested in the coda of CAN syllables, Setter et al. (2010, 17) provide examples of this distribution in HKE, and so does Mangalea (2019, 35-43). Examples like *waifo* and *naifoo* can be explained by constraints transferred from the substrate regarding syllable weight. Namely, if the nucleus is represented by a diphthong or a long vowel, the syllable should remain open (Ø coda). This is further exemplified by *fai/fi* for 'five', *ni* for 'nine' and supported by the attested counterexample *if* (short vowel). We can also find problematic examples, such as *fi-teng* 'fifteen' and *hankerchoo* 'handkerchief', which did not preserve the fricative in the coda, although the nucleus is a short vowel. For *fi-teng* we can argue that it was directly derived from 'five' (ending with /v/, which CPE dropped, resulting in *fi*) by adding the numeral suffix *teng* for 'teen', bypassing the irregular English form 'fifteen'. Alternatively, they can both be explained by substrate influence banning fricatives from the coda, and the fact that this does not apply for heavier syllables can indicate simply that syllable weight constraints outrank phonotactic ones.

Regarding /h/, Table 1 shows us that in CE, the laryngeal fricative undergoes fronting and often surfaces as [x], the velar fricative. Written data from all the sources perused so far (including travelogues, magazine articles, literature) do not offer any means of validating or invalidating this phoneme in CPE.

The samples below show us that /v/ is also attested in syllable-initial (13) and syllable final (14) position, although a more detailed discussion is in order:

(13) a. *co**v**er, sa**v**y* 'understand'[14], ***v**elly* 'very'
Anon. (1836)

b. *de**v**ilo* 'devil', ***v**elly* 'very', *sa**v**an* 'servant
Leland (1876)

c. *sa**vv**y* 'know', *sil**v**er shoe* 'form of payment'
Airey (1906)

d. *sa**vv**y* 'know'
Denby (1911)

(14) a. *go**v**'nor* 'governor', *ha**v**e*
Anon. (1836)

b. *dri**v**e, gi**v**e, ha**v**e*
Anon. (1843)

[14] From Portuguese *saber*, via Macao Portuguese, like *maskee*

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

c. *ha**v**e, lo**v**e-pidgin* ‘love affair’

Leland (1876)

d. *fi**v**e, gi**v**e, ha**v**e*

Denby (1911)

The first comment here is that part of these forms are spelled inconsistently, sometimes in the same source (e.g. Leland (1876) *devilo* / *debilo*, Anon. (1836) *hab*[15] */ have*, *fai / fi / five* for ‘five’, etc.). This suggests that many attestations of /v/ in the coda may be the result of anglicization. Secondly, there is a tendency to substitute [b] for /v/, yet another of the diagnostic features of pidgins manifested in CPE, corroborated by other contemporary sources such as Anon. (1857). Lastly, there are rare instances when fricative devoicing occurs, and /v/ surfaces as [f], e.g. *cofuh* ‘cover’.

As far as /z/ is concerned, English orthography once again hinders the analysis, because in many of the examples in (15), we have no means of establishing whether “s” represents the voiced or the voiceless phoneme:

(15) a. *noi**s**e*, *plea**s**e*, *spo**s**e* ‘suppose’

Anon. (1836)

b. *Chine**s**e, spo**s**e / s’po**s**e* ‘suppose’

Anon. (1843)

c. *loo**s**ee* ‘lose’, *lea**s**on* ‘reason’, *squee**z**e* ‘fine’

Leland (1876)

d. *pli**zz**on* ‘prison’, *squee**z**e* ‘extract money’

Airey (1906)

e. *soft-no**s**ed, suppo**s**e, u**s**e*

Denby (1911)

Moreover, part of the examples above are the productions of English natives (e.g. *casion*, *noise*, *please*, *soft-nosed*), whom we do not expect to devoice any sounds. Further comments are required for those forms where we do not find /z/, for instance, the widely-attested *pidgin* for ‘business’, which renders /z/ as an affricate (possibly /dʒ/, but the Cantonese-specific /ts/ is also a candidate). The same can be argued about *plizzon* – the orthography (identical to the Italian orthography for /dz/) indicates an affricate (either /dz/ or, most probably, /ts/). Replacement of /z/ is also attested in Leland (1876), i.e. *taushan* for ‘thousand’ or *hanszeman* for ‘husband’. For the latter, we can argue that its orthography indicates, once more,

[15] Incidentally, Leland (1876) corroborates the *hab* spelling. In addition, the widely attested *savvy* is also recorded as *subee* in Anon. (1836) or *sabbe / sabbee* in Anon. (1857).

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

the voiceless affricate [ts] especially that it is used as a reflex of voiceless /s/ as well: *chesze* for 'chest' in Leland (1876), or *wiszewun* for 'west wind', in Anon. (1836). The data also exemplify /z/ devoicing in *kweisi* for 'crazy' (Anon. (1836)) or deletion like in *chinee* for 'Chinese' (Leland (1876)).

The remaining fricatives, interdental /θ, ð/ and postalveolar /ʃ, ʒ/ require special attention, as none are attested in Cantonese. Substitution of /θ/ with [t] and /ð/ with [d] is a diagnostic feature of pidgins and creoles, and partially attested in HKE ([d] as a reflex for /ð/, as per Setter et al. (2010, 18) and in CE ([d] as a reflex for /ð/, as per Mangalea (2019, 37). However, these are not the only strategies. In HKE, /θ/ is rendered as [t] after stopping, or as [f] after fronting (Setter et al. 2010, 18). In CE, /θ/ is most commonly replaced by [s][16] whereas [l] and [z] can also surface as reflexes of /ð/. As far as CPE is concerned, at a first glance, the data indicate that both interdentals are attested as per (16):

(16) a. *no**th**er* 'another', ***th**ing*, ***th**at*, ***th**is*

Anon. (1836)

b. ***th**at*, ***th**en*

Anon. (1843)

c. *come-**th**is-side* 'arrive', *inside he mou**th*** 'secretly', *no**th**er* 'another', ***th**at*, ***th**is*

Leland (1876)

d. *dam**th**ing* 'damn thing'*, evele**th**ing* 'everything'*, no**th**ing*, ***th**at*, ***th**en*, ***th**ing*, ***th**ink*

Denby (1911)

However, problems do arise. In Anon. (1836) we find an example of deletion, in *e-too* 'Earth', most probably for reasons of phonotactic constraints (fricatives not being allowed in the Cantonese coda), or syllable weight. Leland (1867) offers an instance of stopping and devoicing (/ð/ → [t] in *mata* 'mother') or simply stopping (/θ/ → [t] in *t'atti* 'thirty', *t'attin* 'thirteen', *tau-shan* 'thousand', *tlee* 'three', *ting* 'thing', *ting-ki* 'thank you*, tinkee* 'think'). Moreover, in its glossary, there is an entry for 'that' and one for *"T'hat, a peculiar and common pronunciation of that"* (Leland, 1876, 134). It could be argued that this spelling apparently indicates aspiration (hence devoicing) and that subsequent forms spelled as "that" are examples of anglicization.

[16] While the fricative /s/ is recognized as part of the phoneme inventory of Putonghua and while there is consensus in the literature regarding the active articulator, the same cannot be said about the passive articulator. This varies between alveolar regions and lower teeth, and Mangalea (2019), based on Duanmu (2007, 25), classifies the Chinese /s/ as a voiceless dental fricative.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Regarding the remaining fricatives /ʃ, ʒ/, as per (17) below, they are both attested. Although not available in the phonemic inventory of Cantonese[17], at least the voiceless postalveolar has a strong presence, and not only in words of English origin:

(17) a. *aw-kuh-**sh**e* 'August', *ca**s**ion* 'occasion', *chay-**sh**e-mun* 'chessman', *cum**sh**aw* 'gift', *fa**sh**ion*, *fini**sh***, *sik-**sh**e* 'six', *six-**sh**e-teng* 'sixteen
Anon. (1836)

b. *cum**sh**aw* 'gift*, English*, *wish*
Anon. (1843)

c. *ca**s**ion* 'occasion', *cum**sh**aw* 'gift', *fini**sh**ee* 'finish', *fi**sh**ee* 'fish', *how fa**sh**ion* 'in what way, how come', *jo**sh*** 'God', *kam-ma-**sh**e-yun* 'commercial', ***sh**leep* 'sleep', ***sh**peakee* 'speak'
Leland (1876)

d. *ca**sh*** 'copper currency', *lick**sh**a* 'rickshaw', *sam**sh**u* 'rice alcohol', *silver **sh**oe* 'form of payment'
Airey (1906)

e. *a**sh**ore*, *fa**sh**ion* 'way, manner', ***Sh**anghai*, ***sh**oreside* 'by the sea'
Denby (1911)

We can see that /ʃ/ is distributed both in onset and in coda position, in words of Cantonese origin, such as *samshu*, Indian origin, such as *cumshaw*, and English origin, such as *cash, finishee* etc. There are limited contexts, such as *shleep* and *shpeakee*, where the voiceless fricative appears unexpectedly in a complex onset, represented in English by biconsonantal, s-initial clusters. However, while clusters are illicit in the substrate, [ʃ] as a reflex of /s/ does not serve as a reduction strategy, nor does it repair the sonority scale generalization violation (which doesn't even apply to /s/ followed by [+sonorants], like the liquid /l/). Other sources, such as Hall (1826), list *speak* in its original English spelling, as does Anon. (1843) with *sleep*. Like other fricatives, /ʃ/ also appears in syllable-initial position after paragoge, most probably because of phonotactic constraints transferred from Cantonese, e.g. *fi-shee* 'fish' or *sik-she* 'six'. This unclear distribution, together with occasional inconsistent spelling and a tendency to raise /ʃ/ to the alveolar [s] (e.g. *fasson* 'fashion'*, russki* 'Russian' in Leland (1876) or *finis* 'finish' in Anon. (1857), *sip* 'ship' in Airey (1906), etc.) suggest linguistic variation. The voiced postalveolar /ʒ/ is available in only one context of appearance ('occasion')[18], and based solely on this, it is attested. However, a more thorough discussion is not possible in the absence of more data.

[17] Standard Chinese does have the retroflex /ʂ/, with the point of articulation just before the palate, and the tip of the tongue bent backwards. This explains why /ʃ/ is easily articulated in CE.

[18] Mangalea (2019) had no context available for /ʒ/ in the selected corpus (Speech Accent Archive), so this is a topic for further research.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

3.1.4. Affricates

Predictions are difficult to formulate regarding the affricates /tʃ/, /dʒ/ for the following reasons. Firstly, these phonemes are not attested in Cantonese, as can be seen from Table 1. However, we can still rely on the substrate to understand possible transfers or replacement strategies, and we would expect much more articulatory ease when it comes to the voiceless affricate, giving that both /t/ and /ʃ/ seem to be attested in CE. Another possible prediction is raising the fricative to /ts/, which is part of the CAN phonemic inventory. Secondly, feedback from current usage (HKE and CE) does not seem provide a clearer image. Setter et al. (2010, 14) lists both affricates in HKE, but without a lengthier discussion, whereas Mangalea (2019, 38) concludes that /tʃ/ is attested in CE, but with no data available to analyze /dʒ/.

The samples in (18), exemplifying the voiceless affricate, clearly outnumber those in (19), focusing on the voiced counterpart. However, even regarding /tʃ/ we can make a number of comments. It is attested in words of both Cantonese origin / transliterations (e.g. *chay-she-mun* 'chessman', the renowned *chinchin* 'greetings, to worship', or *chow-chow* 'eat, food') and in English words such as *China*, *catchee* 'catch' or *hankerchoo* 'handkerchief'. In the case of the former, it can be argued that /ts/ is a better candidate than /tʃ/ because of Cantonese influence. As far as the latter is concerned, no problems seem to arise with /tʃ/ in onset position. In coda position, paragoge occurs, shifting /tʃ/ or /ts/ from the coda to the onset, as is the case with many sounds analyzed so far and subject to phonotactic constraints (*catchee, muchee, wantchee,* etc.).

(18) a. *catchee* 'catch', *chay-she-mun* 'chessman', *che-na-wi-le* 'January', *China, chin-chin* 'greetings, worship'*, chop-chop* 'quickly', *chow-chow* 'eat'
Anon. (1836)

b. *China*, *chinchin* 'goodbye', *fetchee* 'fetch', *such*
Anon. (1843)

c. *catchee* 'catch', *char* 'chair', *chee* 'long', *che-sze* 'chest, box', *chilo* 'child', *chin-chin* 'worship', *chinee* 'Chinese', *chit* 'letter', *chop* 'stamp, seal', *chop-chop* 'fast', *chow-chow* 'food, to eat', *hankerchoo* 'handkerchief', *ma-chin* 'merchant', *midchiman* 'midshipman', *muchee* 'much', *wantchee* 'want'
Leland (1876)

d. *chilo* 'child', *chin-chin* 'greetings', *chit* 'a form of IOU', *chopchop* 'quickly', *chow* 'food'
Airey (1906)

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

e. *catchee* ‘catch’, *China, chop-chop* ‘quickly’, *chow-chow* ‘to eat’, *wanchee* ‘want’

Denby (1911)

The voiced affricate /dʒ/ is much scarcer, on the one hand because of a limited number of contexts of occurrence, and on the other hand, arguably due to articulatory difficulties and the lack of voice contrasts in Cantonese. Part of the examples below (e.g. *ginger*) are the production of native speakers of English. Moreover, in certain contexts where we would expect the voiced affricate, it appears devoiced, as in *che-na-wi-le* for ‘January’.

(19) a. ***g**in**g**er*

Anon. (1836)

b. *pi**g**eon* ‘business’, *sol**d**ier*

Anon. (1843)

c. *ban**j**ee* ‘band’, *pi**dg**in* ‘business’

Leland (1876)

d. *pi**dg**in* ‘business’

Airey (1906)

e. *stlan**g**e* ‘strange’

Denby (1911)

3.1.5. Liquids

While the occurrence of the lateral /l/ is undisputed in CPE, both in onset and coda position, the rhotic poses considerable difficulties. The tendency to replace it with the lateral is reported in four out of the five selected sources, as can be seen in the examples in (20):

(20) a. *Eulop* ‘Europe’, *mally* ‘marry’, *velly* ‘very’

Anon. (1836)

b. *flin* ‘friend*, galanti* ‘grand’, *glound* ‘ground’, *velly* ‘very’, *kali* ‘care’, *kleen* ‘green’, *mandalin* ‘mandarin’, *melican* ‘American’, *leason* ‘reason*, plopa* ‘proper’, *tlee* ‘three’, *ulup* ‘Europe

Leland (1876)

c. *compladore* ‘compradore’, *follin* ‘foreign*’, glynde* ‘grind’, *licksha* ‘rickshaw’, *plizzon* ‘prison’

Airey (1906)

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

d. *blains* ‘brains’*, evelething* ‘everything’, *follin* ‘foreign’, *plopper* ‘proper’, *stlange* ‘strange’

Denby (1911)

3.1.6. Glides

The two RP glides, /j/ and /w/, are attested in all varieties from Table 1, i.e. Cantonese, Hong Kong English and Chinese English (as spoken today by L1 Chinese natives)[19]. Both sounds are attested in CPE as well, as can be seen from (21) and (22):

(21) a. *Eulop* ‘Europe’, *few*, *secure*, *you*

Anon. (1836)

b. *you*

Anon. (1843)

c. *fan-yun* ‘foreigner’, *kam-ma-she-yun* ‘commercial’

Leland (1876)

d. *yak* ‘one’, *yee* ‘two’

Airey (1906)

e. *use*, *yes*, *you*

Denby (1911)

(22) a. *che-na-wi-le* ‘January’, *lum-wun* ‘eleven’[20], *one*, *wanchee* ‘want’, *way* ‘away’, *what*, *which*, *wi-fo* ‘wife’, *wi-sze-wun* ‘West wind’

Anon. (1836)

b. *why*, *wish*

Anon. (1843)

c. *what*, *wata* ‘water’, *seweitun* ‘Sweden’, *waifo* ‘wife’*, walkee* ‘walk’, *wantchee* ‘want’*, wailo* ‘go away’

Leland (1876)

d. *number one* ‘the best’, *whilo* ‘go’

Airey (1906)

e. *wanchee* ‘want’*, water, what*

Denby (1911)

[19] Regarding the approximant /w/, Setter et al. (2010, 13), list it both as bilabial and as velar, considering it a labio-velar in HKE.

[20] *-wun* corresponds to the *-en* spelling, as with other compounds such as *sum-wun* for ‘seven’, and similarly to how *-eng* corresponds to *-een* from *sum-wun-teng* ‘seventeen’ or *ta-teng* for ‘thirteen’.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

The glides are attested only in onset position, as in RP[21]. For examples such as *aw-kuh-she* 'August', *chow-chow* 'eat, food', *cow*[22] or *faw / fowlo* for 'fowl', the nucleus is a diphthong. Moreover, the *fowlo* variant strengthens our assumption regarding syllable weight constraints in CPE, namely, if the nucleus is a long vowel or a diphthong, the coda must be Ø, so that [l] moves to the onset of the new syllable created via paragoge.

As opposed to RP, which allows consonant-glide combinations in the onset, CPE seems to consider these illegitimate. Examples include *sutemeet* 'sweetmeat', *seweitun* 'Sweden', *tunti* 'twenty', *tuilip* 'twelve', where we notice vowel epenthesis or /w/-vocalization. The same is applicable to /j/ preceded by a consonant, in situations like *culio* / *quiri* for 'curious'. The former exemplifies deletion, whereas the latter, replacement, in a transcription which we assume to be /kwɪlɪ/. *Quiri* 'curious' and *kweisi* 'crazy' could indicate that one CG combination is allowed (i.e. /kw/), however, based on substrate influence, we can argue that this is actually the coarticulated consonant. In Cantonese, the articulatory gestures for /k/ (plosive release of the airflow) and /w/ (rounding of the lips) are simultaneous, so that /kw/ is actually one sound, and not a CG combination[23].

4. Syllable structure

Preliminary observations regarding the syllable structure of CPE must make reference to the syllable structure of the substrate, i.e. Cantonese. In the eventuality of any samples recorded in the selected primary sources coming from a speaker with L1 Mandarin, we can also take into consideration the structure of *Putonghua* syllables.

The maximal syllable structure in Cantonese is (C)V(C), according to Chan and Li (2000, 75) and Setter et al. (2010, 25), which entails the following possible configurations: V, CV, VC, and CVC. For *Putonghua*, Duanmu (2007, 71) discusses CGVX, where X stands for a vowel (Ø-coda, resulting in a long vowel or a diphthong in the nucleus) or a consonant. Moreover, in his analysis, consonant-glide (CG) combinations in Chinese emerge as a single sound in the surface representation. The immediate conclusion is that the substrate, be it Cantonese or Mandarin, does

[21] For Cantonese, Setter et al. (2010, 25) note that "diphthongs are treated as vowel + coda, where the coda is a glide". In this interpretation, *faw* 'fowl' is a CVG example.

[22] Indicating female genre, such as *cow-chilo* 'little girl', or the animal, such as *cow-oil* for 'butter'.

[23] This also happens in *Putonghua.* Duanmu (2007) argues that /ɕ, tɕ, tɕʰ/ (transliterated in pinyin as 'x, j, q') are coarticulated CG combinations, namely that the articulatory gestures for /s, ts, tsʰ/ are simultaneous with /j/. This analysis tɕ, tɕʰ/ are the palatalized allophones of the dentals /s, ts, tsʰ/, is supported by the fact that these sounds are in complementary distribution.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

not allow clusters, either in the onset, or in the coda. This is in stark contrast to the lexifier, which has the following maximal syllable size: (C)(C)(C)V(C)(C)(C)(C) (e.g. *strengths* /strɛŋkθs/[24]).

Consequently, one prediction that has already been anticipated in the analysis of the consonantal system is that clusters will tend to be reduced in CPE. In what follows, we will analyze repair strategies and the contexts where they apply, after an overview of single onsets and codas.

4.1. Simple onsets

Both Cantonese and Mandarin are very permissive from the point of view of the onset. Cantonese allows the complete consonantal inventory (/p, pʰ, t, tʰ, k, kʰ, kw, kʰw, f, s, h, ts, tsʰ, m, n, ŋ, l, j, w/), whereas Mandarin, all consonants, with the exception of /ŋ/ (/p, pʰ, t, tʰ, k, kʰ, f, s, ʂ, ʐ, ts, tsʰ, tʂ, tʂʰ, ɕ, tɕ, tɕʰ, m, n, l, x/). Moreover, the onset is optional in both languages[25].

As far as CPE is concerned, the following consonants are attested in simple onsets: /p, b, t, d, k, g, kw[26], m, n, f, v*, s, z*, θ*, ʃ, tʃ, l, j, w/. Based on the data analyzed, as far as /v, z, θ/ are concerned, there is considerable variation and, in many contexts, they were deleted or replaced: *devilo* / *debilo* 'devil', *hab / have* for 'have', *fai / fi / five* for 'five', *cofuh* 'cover'. Moreover, /ð, ʒ, dʒ/ were not taken into consideration because of a limited number of contexts of occurrence, insufficient for a compelling argument, and even at this level, there is variation.

4.2. Simple codas

As opposed to onsets, substrate codas are very restricted. Cantonese allows only /p, t, k, m, n, ŋ/[27] (Chan and Li (2000), Setter et al. (2010)) and *Putonghua* only /n, ŋ/ (Duanmu (2007)).

[24] There is inter-speaker variation with respect to the occurrence of [k].

[25] Although as far as Mandarin is concerned, Kratochvil (1968, 32) argues against this, noting that when the onset is not realized by a consonant, zero onset (breathed transition from silence to sound) or glottal stop (glottis closure released plosively) are the candidates. Duanmu (2007, 73) argues that these are "unintended gestures", needed by the vocal tract to produce the vowel in the absence of a consonant/glide.

[26] The phonological distinction operated by aspiration in the case of /kw/ and / kʰw/ is difficult to assess based solely on written sources.

[27] The discussion of voice vs aspiration, or voice interconnected with aspiration based on VOT, no longer applies in this context. In Cantonese codas, the three plosives are produced with glottal reinforcement and are not released.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

At a first glance, throughout the five selected sources, the following consonants are attested in CPE simple codas: /m, n, ŋ, t, d, p, s, k, l, ʃ, b, v, f, z, g, θ/, which greatly exceeds any predictions that may be formulated based on a restrictive substrate. The aforementioned list comprises not only nasals or voiceless plosives, but also voiced plosives, voiced and voiceless fricatives and the liquid /l/. However, a closer quantitative analysis reveals the patterns visible in Tables 2-6 (dedicated to each source). The *Expected – NO* columns count the number of phonological contexts where in RP we would expect a certain sound to surface, and in CPE, it does not, because of phenomena which will be discussed henceforth.

Table 2. Codas in Anon. (1836)

Coda	Expected		Occurrences	% Unexpected
	NO	YES		
⌀ coda	**13**	**60**	**73**	**18%**
-n	6	20	26	23%
-m	2	8	10	20%
-ŋ	3	7	10	30%
-s		9	9	0%
-t	3	6	9	33%
-k	2	6	8	25%
-p	2	5	7	29%
-l		4	4	0%
-tʃ		3	3	0%
-z		3	3	0%
-b	2		2	100%
-v		2	2	0%
-d		1	1	0%
-f		1	1	0%
-ʃ		1	1	0%
Grand Total	**33**	**141**	**174**	**19%**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 3. Codas in Anon. (1843)

Coda	Expected		Occurrences	% Unexpected
	NO	YES		
⌀ coda	**5**	**20**	**25**	**20%**
-n		8	8	0%
-l		7	7	0%
-t		6	6	0%
-m		5	5	0%
-v		3	3	0%
-z		3	3	0%
-s		2	2	0%
-ʃ		2	2	0%
-d		1	1	0%
-f		1	1	0%
-k		1	1	0%
-tʃ		1	1	0%
Grand Total	**5**	**63**	**68**	**7%**

Table 4. Codas in Leland (1876)

Codas	Expected		Occurrences	% Unexpected
	NO	YES		
⌀ coda	**36**	**49**	**85**	**42%**
-n	18	61	79	23%
-m	11	15	26	42%
-t	1	20	21	5%
-d	1	17	18	6%
-s	1	12	13	8%
-p	1	8	9	11%
-k		8	8	0%
-ŋ		7	7	0%
-b	2	2	4	50%
-l	1	1	2	50%
-v		2	2	0%
-z		2	2	0%
-g		1	1	0%
-ʃ		1	1	0%
-θ		1	1	0%
Grand Total	**73**	**215**	**288**	**25%**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 5. Codas in Airey (1906)

Coda	Expected		Occurrences	% Unexpected
	NO	YES		
⌀ coda	**4**	**12**	**16**	**25%**
-n	1	9	10	10%
-m		6	6	0%
-p		5	5	0%
-s		4	4	0%
-l	1	1	2	50%
-d		1	1	0%
-k		1	1	0%
-ŋ		1	1	0%
-ʃ		1	1	0%
-t		1	1	0%
-z		1	1	0%
Grand Total	**6**	**47**	**53**	**11%**

Table 6. Codas in Denby (1911)

Coda	Expected		Occurrences	% Unexpected
	NO	YES		
⌀ coda	**7**	**35**	**42**	**17%**
-n	1	14	15	7%
-ŋ		9	9	0%
-m		8	8	0%
-s		7	7	0%
-t		7	7	0%
-d		5	5	0%
-k		5	5	0%
-p		4	4	0%
-l		3	3	0%
-v		3	3	0%
-z		2	2	0%
Grand Total	**8**	**111**	**119**	**7%**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Firstly, we can notice that in all five sources, open syllables prevail. Secondly, the nasals / m, n, ŋ/ are a constant in all five sources, and at the top of the list in the order of frequency, followed by (predominantly voiceless) plosives and fricatives, with the voiced counterpart attested much more infrequently. In accordance with sociolinguistic accountability, the *Expected* column isolates those contexts where a certain phoneme is not attested in CPE although attested in the lexifier, and vice-versa (a phoneme is attested in CPE although it is not in RP). Yet another common denominator in the data is that unexpected occurrences are linked, in a large proportion (from 17% to 42%), to open syllables, which suggests stronger phonotactic constraints on the coda rather than the onset.

For example, /tʃ/, /l/ can appear in single codas in the lexifier, but in CPE they move to the onset of a new syllable created by paragoge, e.g. (23) and (24), respectively. Given that vowel epenthesis is much more often employed for the reduction of illicit clusters, we will discuss the quality of the vowel in subsequent sections.

(23) a. *muchee* 'much'
Anon. (1836)
b. *fetchee* 'fetch'
Anon. (1843)
c. *too-muchee* 'too much'
Leland (1876)
d. *catchee* 'catch'
Denby (1911)

(24) a. *alla* 'all', *call-um* 'call
Anon. (1836)
b. *smellum-wata* 'cologne', *spillum* 'spill'
Leland (1876)

The larger proportion of the changes affecting syllable structure has to do with consonantal clusters, which will be reviewed in subsequent sections. There are, nonetheless, many instances when vowel epenthesis and, more rarely, consonant deletion operate for reasons of syllable weight. Substrate syllables are predominantly light or heavy at most (never super heavy), not allowing consonants to follow diphthongs. Further proof that this phonotactic constraint transferred to CPE is the fact that although Cantonese does not phonemically differentiate

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

between long and short vowels, with the exception of /a/ and /a:/[28], long vowels, too, are part of open syllables in CPE. The samples in (25) respectively exemplify strategies of preserving the syllable weight in case of a diphthong and in case of a long vowel in the nucleus:

(25) a. *fi* 'five', *insi* 'inside', *ni* 'nine', *sze-kay-le-sze* 'scale', *sze-taw* 'stove', *takee* 'take', *wifo* 'wife'

Anon. (1836)

b. *au-sai* 'outside', *tim*[29] 'time', *boilum* 'boil', *facee* 'face', *faitee* 'faith', *houso* 'house', *go-hom* 'go home', *makee* 'make', *naifoo* 'knife', *naiti* 'night', *piecee* 'piece, CL', *waifo* 'wife'

Leland (1876)

c. *litey* 'write'

Airey (1906)

d. *all li/ly* 'all right', *makee* 'make'

Denby (1911)

(26) a. *talkee* 'talk'

Anon. (1843)

b. *cot-houso* 'court-house', *talkee*[30] 'talk', *largee /largey* 'large', *loosee* 'lose', *one piecee* (classifier), *walkee* 'walk'

Leland (1876)

c. *talkee* 'talk'

Denby (1911)

4.3. Clusters

The discussion regarding clusters is mandatory not only in light of the attention that this subject has received in the literature on pidgins and creoles so far (see e.g. Avram 2005), but also with reference to Hong Kong English as a soft variant of feedback from current usage. Cantonese and *Putonghua* do not allow any clusters at all, neither in the onset, nor in the coda, while in Hong Kong English, both syllable-initial and syllable-final clusters undergo simplification, the latter more so than the former. A natural prediction, under these circumstances, is that clusters may be illicit in CPE as well.

[28] *Putonghua* does not differentiate at all.

[29] And in many compounds with 'time', e.g. *before-tim* 'previously', *one-tim* 'at one time, once', *nother-tim* 'another time', *that-tim* 'then' etc.

[30] *good-talkee / talkee leason* 'reasonable talk', *handsome-talkee* 'eloquence'

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

4.3.4. Onset clusters

In the selected sources, onsets are available in a reduced number of contexts, as compared to codas. They are, nonetheless, attested, as can be seen from (27):

(27) a. *please*, *proper*, *small*, *stay*, *steal*, *troub*
Anon. (1836)

b. *drive, great*, *sleepy*, *small*
Anon. (1843)

c. *plopa* 'proper', *dlinkee* 'drink', *glound* 'ground', *smellum* 'smell', *spilum* 'spill'
Leland (1876)

d. *chopsticks*, *close*, *glynde* 'grind', *plizzon* 'prison'
Airey (1906)

e. *blains* 'brains', *break*, *from*, *plopper* 'proper'
Denby (1911)

The majority are two-consonant clusters; however, three-consonant clusters are also marginally attested: *stlange* 'strange' (Denby 1911), *squeeze* (Airey 1906), *squeeze* (Leland, 1876).

As in the case of HKE, CPE appears to also reduce onset clusters more rarely than coda clusters. Only between 10-20% of the unexpected occurrences of CPE onsets differ from RP because of reduction strategies, as can be seen from tables 7–11.

Table 7. Onset Analysis of Anon. (1843)

Onset Clusters	Occurrences	Expected	
		NO	YES
-st	4		4
N/A	2	2	
-kw	1	1	
-pl	1		1
-pr	1		1
-sk	1		1
-sm	1		1
-sp	1		1
-tr	1		1
Grand Total	**14**	**3**	**11**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 8. Onset Analysis of Anon. (1843)

Onset Clusters	Occurrences	Expected
-sm	2	2
-dr	1	1
-fr	1	1
-gl	1	1
-gr	1	1
-sl	1	1
-sp	1	1
Grand Total	**8**	**8**

Table 9. Onset Analysis of Leland (1876)

Onset Clusters	Occurrences	Expected	
		NO	YES
N/A	7	7	
-gr	5	3	2
-bl	3	1	2
-pl	3	2	1
-fl	2		2
-kw	2	1	1
-sp	2		2
-br	1		1
-dl	1	1	
-fr	1		1
-kj	1		1
-pr	1		1
-skw	1		1
-ʃl	1		1
-sm	1		1
-tl	1	1	
Grand Total	**33**	**16**	**17**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 10. Onset Analysis of Airey (1906)

Onset Clusters	Occurrences	Expected	
		NO	YES
-pl	2	1	1
-gl	1	1	
-kl	1		1
N/A	1	1	
-sp	1		1
-st	1		1
-swk	1		1
Grand Total	**8**	**3**	**5**

Table 11. Onset Analysis of Denby (1911)

Onset Clusters	Occurrences	Expected	
		NO	YES
-bl	1	1	
-br	1		1
-fr	1		1
N/A	1	1	
-pl	1	1	
-skr	1		1
-sl	1		1
-stl	1	1	
-vr	1		1
Grand Total	**9**	**4**	**5**

N/A entries mark cluster reduction, for which the most common strategy is vowel epenthesis, e.g. *sitop* 'stop' (Anon. 1836), *kolock* 'clock', *patili* 'priest'[31] (Leland 1876), *evelething*[32] 'everything' (Denby 1911). There are also examples of /w/-vocalization, for example *sutemeet* 'sweetmeat' (Anon. 1836), *tui-lip* 'twelve' and *tun-ti* 'twenty' (Leland 1876). Apart from this, the most common phenomenon, corroborated by multiple sources, is the [l] reflex for /ɹ/, as can be seen in (28):

[31] From Portuguese *padre*.
[32] Which we argue to be an example of cluster reduction by vowel-copying ([e]) based on the spelling.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

(28) a. *all plopa* 'in good order', *dlinkee* 'drink', *glound* 'ground'
Leland (1876)

b. *glynde* 'grind', *plizzon* 'prison'
Airey (1906)

c. *blains* 'brains', *plopper* 'proper', *stlange* 'strange'
Denby (1911)

4.3.4. Coda Clusters

Syllable-final clusters are marginally attested, as can be seen from the examples below:

(29) a. *box*[33], *just*, *last*, *must*, *think*
Anon. (1836)

b. *and*, *fools*, *handsome*
Anon. (1843)

c. *bund*, *first-chop* 'best quality', *glound* 'ground', *handsome talkee* 'eloquence', *willow-waist*
Leland (1876)

d. *chopsticks*, *follin devils*, *joss sticks*
Airey (1906)

e. *blains*, *fault*, *must*, *second*, *soft-nosed*, *stlange*, *think*
Denby (1911)

Even a three-consonant cluster is attested in Airey (1906), i.e. *cents.* As can be seen, some of these clusters are the result of morphological processes (plural and the participial *-ed* suffixes, the latter being the production of a native speaker of English). Tables 12-16 quantify the frequency of repair strategies for coda clusters, which are, in a large proportion, illicit:

[33] Nonetheless, counterexamples are also available, such as *fixee* 'fix' (Anon. 1836).

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 12. Cluster Reduction Frequency in Anon. (1836)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
⌀ coda	**73**	**13**	**8**	**62%&**
-n	26	6	6	100%
-k	8	2	2	100%
-p	7	2	2	100%
-b	2	2	1	50%
-m	10	2	1	50%
-ŋ	10	3	1	33%
-t	9	3	1	33%
-d	1			
-f	1			
-ks	1			
-l	4			
-ŋk	1			
-s	9			
-dʒ	1			
-st	3			
-tʃ	3			
-v	2			
-z	3			
Grand Total	**174**	**33**	**22**	**67%**

Table 13. Cluster Reduction Frequency in Anon. (1843)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
⌀ coda	**25**	**5**	**2**	**40%**
-d	1			
-f	1			
-k	1			
-l	7			
-ls	1			
-m	5			
-n	8			
-nd	2			

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
⌀ coda	**25**	**5**	**2**	**40%**
-s	2			
-ʃ	2			
-t	6			
-tʃ	1			
-v	3			
-z	3			
Grand Total	**68**	**5**	**2**	**40%**

Table 14. Cluster Reduction Frequency in Leland (1876)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
-n	79	18	11	61%
⌀ coda	85	36	7	19%
-l	2	1	1	100%
-s	13	1	1	100%
-b	4	2		0%
-d	18	1		0%
-g	1			
-k	8			
-m	26	11		0%
-nd	5	1		0%
-nt	1			
-ŋ	7			
-p	9	1		0%
-ʃ	1			
-st	2			
-t	21	1		0%
-v	2			
-z	2			
-θ	1			
Grand Total	**288**	**73**	**20**	**27%**

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Table 15. Cluster Reduction Frequency in Airey (1906)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
∅ coda	**16**	**4**	**2**	**5**
-l	2	1	1	100%
-n	10	1	1	100%
-d	1			
-k	2			
-ks	2			
-ls	1			
-m	6			
-ŋ	1			
-p	5			
-s	4			
-ʃ	1			
-t	1			
-z	1			
Grand Total	**53**	**6**	**4**	**67%**

Table 16. Cluster Reduction Frequency in Denby (1911)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
-n	15	1	1	100%
∅ coda	42	7	1	14%
-d	5			
-ft	1			
-k	5			
-l	3			
-lt	1			
-m	8			
-nd	1			
-ndʒ	1			
-ns	1			
-nts	1			

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Coda	Occurrences of Coda	Unexpected Occurrences	Occurrences of Cluster Reduction	% Cluster Reduction / Unexpected Occurrences
-ŋ	9			
-ŋk	1			
-p	4			
-s	7			
-st	1			
-t	7			
-v	3			
-z	2			
-zd	1			
Grand Total	**119**	**8**	**2**	**25%**

ø coda in the tables above indicates an open syllable. We can notice that when cluster reduction occurs, most often it results in open syllables[34], or /n/-single codas, and to a lower degree, in obstruent / liquids-single codas. Moreover, a large proportion of unexpected occurrences (50-100 %) are connected to repair strategies for illicit clusters. The preferred strategy seems to be vowel epenthesis, as can be seen in (30). Marginal strategies comprise /w/-vocalization (e.g. *sutemeet* for "sweetmeat") or deletion of word-final /t/ and /d/, very often operating simultaneously with vowel epenthesis:

(30) a. *bindee* 'bind', *sendee* 'send', *sileek* 'silk', *wanchee* 'want'

Anon. (1836)

b. *olo* 'old', *patili* 'priest', *savan* 'servant', *tinkee* 'think', *tau-shan* 'thousand'

Leland (1876)

c. *chilo* 'child', *glynde* 'ground'

Airey (1906)

d. *wanchee* 'want'

Denby (1911)

[34] Based on this and the numerous instances of syllable structure phenomena that have to do with syllable weight, we can hypothesize that syllable-weight constraints generally outrank phonotactic ones.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Vowel epenthesis is consequently a very common phenomenon, surfacing as a result of both syllable-weight and phonotactic constraints. Regarding the quality of the vowel, the default intrusive vowel is /i/, which is documented as a common strategy in many English-lexifier pidgins and creoles, however not the only one. We also have marginal instances of vowel copying (e.g. *alla* for 'all', *evelething* 'everything').

5. Conclusions

Based on the primary sources selected, the following conclusions can be drawn regarding the phonology of Chinese Pidgin English. As far as the consonantal system is concerned, /p, t, k, f, s, h, ʃ, tʃ, m, n, ŋ, l, j, w/ are generally attested, as well as the coarticulated consonants /kw, k^hw/. Although voiced phonemes, especially the plosives /b, d, g/ are attested, it is apparent that voicing contrasts are not easily realized by CPE speakers. Proof of this comes from anglicized examples, inconsistencies in the sources, which may indicate linguistic variability, and phenomena such as replacement (/v/ → [b], /θ, ð/ →[t]), deletion, devoicing (e.g. *cofuh* 'cover') fronting (*mata* 'mother'), etc. The interdentals and the voiced postalveolar fricative (/ð, θ, ʒ/) represent a topic for further research as the contexts of occurrence were either insufficient in number or inconsistent. The rhotic approximant is predominantly replaced by the lateral one.

Single-consonant onsets are permissive in CPE, allowing /p, b, t, d, k, g, kw, m, n, f, v*, s, z*, θ*, ʃ, tʃ, l, j, w/, with /v, z, θ/ marked with an asterisk because of phenomena such as replacement or deletion. Single-consonant codas, however, are much more restricted because of phonotactic constraints transferred from the substrate. Nasals and voiceless plosives are most frequently attested in single-consonant codas, followed by voiced and voiceless fricatives, and very rarely affricates and the liquid /l/. As the data has shown, between 18% and 42% of unexpected coda phenomena have resulted in open syllables, whereas the etyma have a consonant or a cluster.

While two-consonant and even three-consonant clusters are attested, repair strategies often operate on them. From among the repair strategies vowel-epenthesis is the most common, but /w/-vocalization and consonant deletion are also attested. Very importantly, vowel-epenthesis does not occur only as a repair strategy for illicit clusters or because of phonotactic constraints not allowing

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

certain phonemes in the coda, as in other Pacific English-lexifier pidgins and creoles (Avram 2005: 196-204), but also for reasons of syllable weight. In the substrate, if the nucleus is represented by a diphthong or a long vowel, the syllable should remain open (Ø coda), and this constraint seems to have transferred to CPE (*waifo* 'wife', *naifoo* 'knife').

Primary Sources

Anderson, Aeneas. 1795. *A Narrative of the British Embassy to China, in years 1792, 1793, and 1794*. London: J. Debrett.

Anon. a. 1836. "Jargon Spoken at Canton: How It Originated and Has Grown into Use; Mode in Which the Chinese Learn English; Examples of the Language in Commercial Use Between Foreigners and Chinese." In *Chinese Repository* IV: 428-435.

Anon. b. 1843. *The Last Year in China, to the Peace of Nanking: As Sketched in Letters to His Friends, by a Field Officer*. London: Longman, Brown, Green and Longmans.

Airey, Frederick W. I. 1906. *Pidgin Inglis Tails and Other Plaited and Drawn by F. W. I. Airey*. Shanghai: Kelly and Walsh.

Denby, Jay. 1911. *Letters from China and Some Eastern Sketches*. London: Murray and Evenden.

Downing, Charles T. 1838. *The Fan-Qui in China in 1836-7*. London: Henry Colburn.

Keate, George. 1788. *An Account of the Pelew Islands, Situated in the Western Part of the Pacific Ocean, Composed from the Journals and Communications of Captain Henry Wilson, and Some of His Officers, Who, in August 1783, Were There Shipwrecked*. London: G. Nicol.

Hall, Basil. 1826. *Voyage to Loo-Choo and Other Places in the Eastern Seas, in the Year 1816*. London: Hurst, Robinson.

Hunt, Cecil. 1931. *Honoured sir —from Babujee*. London: Philip Allan and Co.

Hunter, William C. 1911. *The 'Fan Kwae" at Canton before Treaty Days 1825–1844. By an Old Resident*. Shanghai: Kelly and Walsh.

Leland, Charles. G. 1876. *Pidgin-English Sing-Song or Songs and Stories in the China-English Dialect. With a Vocabulary*. London: Trübner and Co.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Noble, Charles F. 1762. *A Voyage to the East Indies in 1747 and 1747. Containing an Account of the Islands of St. Helena and Java. Of the City of Batavia. Of the Government and Political Conduct of the Dutch. Of the Empire of China, with a Particular Description of Canton; and of the Religious ceremonies, manners and Customs of the Inhabitants*. London: T. Becket and P. A. Dehondt.

References

Avram, Andrei A. 2000. "On the Phonological Interpretation of Early Written Records of English-Based Pidgins and Creoles." *Lancaster University Working Papers of the Centre for Language in Social Life* 117:1-15.

Avram, Andrei A. 2005. *On the Syllable Structure of English Pidgins and Creoles.* Bucharest: Editura Universităţii din Bucureşti.

Avram, Andrei A. 2015. "The Oldest Jamaican Creole Text: A Linguistic Analysis." In *Languages in Contact 2014*, ed. by Piotr P. Chruszczewski, Richard L. Lanigan, John R. Rickford, Katarzyna Buczek, Aleksandra R. Knapik and Jacek Mianowski, 21-46. Washington, D.C: Wydawnictwo Wyższej Szkoły we Wrocławiu, Wrocław.

Baker, Philip and Lise Winer. 1999. "Separating the Wheat from the Chaff. How Far Can We Rely on Old Pidgin and Creole Texts?" In *St Kitts and the Atlantic Creoles. The Texts of Samuel Augustus Mathews in Perspective*, ed. by Philip Baker and Adrienne Bruyn, 103-122. London: University of Westminster Press.

Baker, Philip and Magnus Huber. 2001. "Atlantic, Pacific, and World-wide Features in English-lexicon Contact Languages." *English World-Wide* 22: 157-208.

Bickerton, Derek. 2009. *Bastard Tongues. A Trailblazing Linguist Finds Clues to our Common Humanity in the World's Lowliest Languages*. New York: Hill and Wang, A division of Farrar, Straus and Giroux.

Chan, Alice Y. W. and David C.S. Li. 2000. "English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems." *Language, Culture and Curriculum* 13 (1): 67-85.

Crowley, Terry. 1993. "Père Pionnier and Late Nineteenth Century Bislama." *Journal of Pidgin and Creole Languages* 8 (2): 207-226.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Hancock, Ian F. 1977. “Recovering Pidgin Genesis: Approaches and Problems.” In *Pidgin and Creole Linguistics*, ed. by Albert Valdman, 277-294. Indiana: University Press, Bloomington.

Holm, John. 1994. *Pidgins and Creoles*, vol. II, *Survey*. Cambridge: Cambridge University Press.

Holm, John. 2000. *An Introduction to Pidgins and Creoles*. Cambridge: Cambridge University Press.

Kratochvil, Paul. 1968. *The Chinese Language Today, Features of an Emerging Standard*. London: Hutchinson University Library.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lalla, Barbara. 1986. “Tracing Elusive Phonological Features of Early Jamaican Creole.” In *Focus on the Caribbean*, ed. by Manfred Görlach and John A. Holm, 117-132. Philadelphia: John Benjamins.

Lalla, Barbara and Jean D’Costa. 1990. *Language in Exile. Three Hundred Years of Jamaican Creole*. Tuscaloosa: University of Alabama Press.

Lohrli, Anne. 1973. *Household Word. A Weekly Journal 1850-1859 Conducted by Charles Dickens*. Toronto and Buffalo: University of Toronto Press.

Malcolm, Elisabeth L. 1973. “The Chinese Repository and Western Literature on China 1800 to 1850.” *Modern Asian Studies* 7 (2): 165-178.

Mangalea, Ioana. 2019. *The Phonology of Chinese English. A Case Study of Mandarin Users*, B.A. thesis, University of Bucharest.

Miron, Ioana (in press). “An Overview of the Data Sources for Chinese Pidgin English”. *Analele Universității din București. Limbi si literaturi straine* LXVI (2).

Mühlhäuser, Peter. 1997. *Pidgin and Creole Linguistics, expanded and revised edition.* London: University of Westminster Press.

Rickford, John R. 1986. “Short Note.” *Journal of Pidgin and Creole Languages* 1 (1): 159-163.

Rickford, John R. 1991. “Contemporary Source Comparison as a Critical Window on the Afro-American Linguistic Past.” In *Verb Phrase in Black English and Creole*, ed. by William F. Edwards and Donald Winford, 302-322. Detroit: Wayne State University Press.

Rickford, John R. 1995. “Textual Evidence on the Nature of Early Barbadian Speech, 1676-1835.” *Journal of Pidgin and Creole Languages* 9 (2): 221-255.

Sebba, Mark. 1997. *Contact Languages. Pidgins and Creoles*. Basingstoke: Macmillan.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)

Setter, Jane, Cathy S.P. Wong, and Brian H.S. Chan. 2010. *Dialects of English. Hong Kong English*. Edinburgh: Edinburgh University Press.

Winer, Lise. 1984. "Early Trinidadian Creole: The *Spectator* texts." *English World-Wide* 5 (2): 181-210.

BDD-A32205 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-26 16:59:46 UTC)