Origins of a Preposition: Chinese Pidgin English long and its Implications for Pidgin Grammar

Michelle Li
The University of Hong Kong
li.michellekl@gmail.com

Abstract
This paper examines the origins and grammatical properties of a preposition in Chinese Pidgin English – long – which has not received much discussion. The significance of long is that it is highly multifunctional and semantically versatile. Long is used to indicate a range of semantic roles: comitative, benefactive, malefactive and source. A second function of long is to mark coordination. It will be shown that a substantial part of the syntax and semantics of long can be attributed to substrate transfer of a corresponding Cantonese morpheme tung4 ‘with’. The creation of long does not conform to the traditional thesis of simply taking the phonetic form from the lexifier language and deriving the grammar from the substrate language. It will be argued that the emergence of long is a case of multiple etymologies which involves the recombination of phonological, syntactic and semantic features from both English and Cantonese. Findings from new CPE sources also suggest a need for re-examining the historical connections between CPE and other Pidgin English varieties of the Pacific region.

Keywords
Chinese Pidgin English; preposition; substrate transfer; multiple etymologies; conflaction

1. Introduction

Chinese Pidgin English (CPE) was a trade pidgin created around the mid-eighteenth century (Tryon, Mühlhäuser & Baker, 1996; Van Dyke, 2005: 81). It served as a lingua franca for interethnic communication, especially among Chinese and Europeans. CPE originated in Canton which was the only port opened for foreign trade in the eighteenth century. As a result of the Opium Wars, China was forced to open four more treaty ports along the China Coast. Foreign trade began to spread to cities such as Amoy (Xiamen), Foochow (Fuzhou), Shanghai and Ningpo (Ningbo), and so did CPE. Throughout its ca. 250 years of use, CPE remained a functionally restricted pidgin until its extinction around the 1960s.
The lexicon of CPE mainly consists of English vocabulary. There are, however, a number of vocabulary items originating from other languages, for example *savvy* ‘know’ (< Portuguese), units of measurement such as *catty, candareen* (< Malay), *chop* ‘seal, trade mark’ (< Hindi), etc. CPE lacks inflectional morphology and is basically a SVO language. CPE shows considerable influence from substrate (Cantonese) grammar, which is evident in the use of the classifier *piece*, serial verb constructions, topicalization, wh-in-situ questions, etc. (Hall, 1944; Baker, 1987; Baker & Mühlhäusler, 1990; Shi, 1991; Bolton, 2003; Ansaldo, Matthews & Smith, 2010; Matthews & Li, 2012). The preposition *long* discussed in this paper also demonstrates features that result from substrate transfer.

2. CPE Sources

CPE data for this study is drawn from two types of sources: English sources and Chinese sources, which are categorized according to the orthographies used to represent CPE, i.e. CPE expressions in English sources are represented by using English orthography, while CPE texts in Chinese sources are in Chinese script. These two types of sources also show a difference in genre. Chinese sources mainly consist of glossaries and teaching materials, whereas English sources include a variety of genres, for example glossaries, memoirs, travelogues, letters, periodicals, etc.

English sources comprise around 15,000 words (G. Smith, 2008: 68), which are divided into two sub-corpora according to dates of attestation: CPE1 (1721-1842) contains texts from early China-West contact to the close of the First Opium War (Baker, 2003a); CPE2 (1843-1990) starts with the opening of four more treaty ports, in addition to Canton, to the last decade of the twentieth century (Baker, 2003b). These sources span almost three centuries and are an invaluable database for analyzing the development and changes in CPE.

The major Chinese source used in this paper is *The Chinese and English Instructor* (hereafter the *Instructor*) (Tong, 1862), published in Canton. Tong Ting-kü 唐廷樞 (1832-1892) (also known as Tong King-sing 唐景星), author

---

1 These CPE corpora are not published, but have generally been made available to pidgin and creole scholars in electronic form by Philip Baker. In addition to the CPE texts, the corpora also contain background information such as dates and places of attestations, ethnic background of the speakers and page references in the sources.

2 Tong (1862) is the most comprehensive Chinese source with CPE renderings, but this doesn’t mean that attestations of *long* are absent in other Chinese sources. For example, the
of the *Instructor*, was a Chinese merchant and comprador. Born in Tong Ka Village (唐家村) of Xiang Shan (香山) in Canton, Tong received an Anglo-Chinese education at the Morrison Education Society School from the age of ten and later continued his study at the London Missionary Society school at Hong Kong (C. Smith, 2005). Though the 6-volume book was primarily prepared for studying Standard English, pidgin renderings of English sentences appeared in the margins and main text of volumes 4 and 6 of the book (Selby & Selby, 1995). Bolton (2003: 176) states that the pidgin texts in the *Instructor* “are probably the largest single source for Chinese Pidgin English available to any Pidgin and Creole scholar.” Entries for the Chinese and English texts in the book were organized neatly and the phonetic transcription of the languages (Cantonese, English and CPE) was coded systematically. The layout of each entry was as follows: nineteenth century Cantonese was given on the left hand side. To the right of the Cantonese sentences were their romanized forms in the convention of *Dr Williams’ Tonic Dictionary of the Canton Dialect* (1856). English texts were given under the romanized Cantonese. The learning of English is aided by a “rigorous and very effective” phonetic transcription into Chinese characters, presented right next to the English translation (Selby & Selby, 1995: 124). Pidgin texts were not as systematically arranged as the Chinese and English texts. While the layout of the main texts (Chinese and English) was standardized, it seems that pidgin texts were inserted wherever space was allowed. Therefore, pidgin expressions can be found adjacent to the main texts, sometimes with a line separating the Chinese and English texts from the pidgin texts; they can also be found in the margins, just above the main texts. The pidgin text used in this paper is based on the transcription of Tong (1862) in Li, Matthews and Smith (2005).

3. The Preposition *long*

3.1 *Long* in CPE

Although a number of works have been done on the grammar of CPE, few studies focus on the use of prepositions. One of the earliest attempts at

preposition *long* (translated as Chinese *gàng6* 共 ‘with’) appears in one of the *Hungmou* ‘Redhaired’ chapbooks sold in Canton (Bolton 2003). In Shanghai, pidgin texts were found in a phrasebook entitled *Yinghua Zhujie* (Feng et al. 1860) prepared by six Ningpo merchants, while Yang Xun (1873) published verses containing CPE expressions in the Shanghai newspaper *Shen Bao*. *Long* was attested in both these sources from Shanghai.

3 It is not clear whether the pidgin renderings were added by Tong himself or another author.
providing a more comprehensive grammatical description of CPE is Hall (1944), in which a short description on the prepositions in CPE is given:

*Prepositions* are used to introduce adjectival or adverbial phrases with nouns or verbs as their centers; among the prepositions are *bîlôŋ, b(s)lôŋ ‘of, for,’ fû(r) ‘for, màski ‘in spite of,’ õlsem ‘like, as.’ *(Hall, 1944: 98)*

Some examples are then given: *mœn bîlôŋ lûk-sî hûrs ‘man for looking after horses, groom’; fû jû ‘for you’; õlsem màj ‘like me’; fû kiçensajd ‘for in the kitchen’ *(Hall 1944: 99).* Unlike Hall (1944) whose data are twentieth century CPE produced by English-speaking informants, Shi’s (1991: 22) analysis of CPE grammar is based on a wider basis of historical sources. However, he finds that no prepositions were being recorded in his corpus. English prepositions were attested but they were used as adjectives, nouns, or adverbs. The use of *long* as a preposition is mentioned in Baker (1987: 181) who shows that *long ‘with’* is attested as a comitative marker as early as 1831, but he goes on to say that “later attestations in CPE are so few in number as to suggest that *long* never became fully integrated in CPE.” Though admitting the existence of *long* in CPE, the findings by these authors seem to suggest that either *long* is not a preposition or *long* has never developed into a genuine preposition in CPE. With the availability of a larger database, there is no doubt that *long* is a genuine preposition in CPE and that it does not merely exist as a comitative marker, but a semantically more versatile function word.

### 3.2 Long in Melanesian Pidgin

In the Pacific region, *long* is also a multi-functional preposition. Some comparative studies on pidgins and creoles have shown that *long* is a worldwide feature. One such study is Clark (1979) which examines the lexical and grammatical features and the historical connection between Melanesian Pidgin and China Coast Pidgin (i.e. CPE). Two prepositional uses of *along* are examined: *along*¹ which functions as a comitative marker and *along*² a general preposition referring to meanings such as English *at, on, to,* etc. A contrast between CPE and Melanesian Pidgin in that CPE only has *along*¹ whereas Melanesian Pidgin has *along*².

---

¹ Contrary to Hall (1944), Clark (1979) and Baker (1987: 183) argue that *belong* is never a preposition in CPE. I agree with Clark and Baker since in the majority of cases, *belong* is either a verb or a copula. There are no clear cases of *belong* functioning like a preposition in the corpora.
Baker (1987) not only studies the use of *long* in CPE but also proposes historical relationships between the varieties of Pidgin English in the Pacific. Since Baker’s findings suggest that comitative *long* is not a fully incorporated preposition in CPE and is attested later than Pacific pidgins, he argues that CPE is unlikely to be the predecessor of comitative *along* in Pacific Pidgin English. However, the present study on *long* in CPE reveals that Baker’s statements may need revision with respect to the following points: First, Baker gives the year 1831 as the earliest attestations of *long* in Tilden’s journal. In fact in a journal prepared by Tilden dated 1819, *long* was already used as a comitative preposition.\(^5\) This certainly predates the first attestation (1827) of *along* in New South Wales. Second, given that *long* is commonly used in both English and Chinese sources Baker’s rejection of *long* being a preposition in CPE seems unjustified. Third, regarding the historical connection between CPE and Pacific pidgins, though it is still too early to establish confirmed relationships between these varieties of Pidgin English, the new findings suggest a new possibility for re-examining the connection between CPE and Melanesian Pidgin.

4. Functions of *long* in CPE

In this section, a descriptive account of the major meanings of *long* in CPE is given. These meanings can be categorized into two grammatical functions: preposition which is presented in section 4.1 and conjunction in section 4.2.

4.1 *Long* as preposition

Of the few prepositions that CPE has, *long* is probably the most important. The significance of *long* is demonstrated in its high frequency of occurrence and its versatility in meaning. The prepositional meanings of *long* are presented below.

4.1.1 Comitative

In early CPE texts, the complex form *along with*, instead of *long*, is found. This is shown in (1) attested in the *Memoirs of William Hickey*.

1. Maske Maske you come along with me (1769; Spencer, 1925: I.227)\(^6\)
   
   ‘Never mind, never mind, come along with me’

\(^5\) In the 1819 journal, there are eight attestations of *long*.
\(^6\) The year before the reference refers to the approximate date of attestation.
The form *along* as in (2) can be occasionally found in later texts.

(2) belong have got one piecee flend, my go along he (-1860; Anon, 1860: 100)
   ‘I have a friend, I can go with him.’

In most nineteenth and twentieth century texts, comitative meaning is mainly indicated by *long* or marginally *with*. Since *with* is absent from Chinese sources and the pronunciation makes it difficult for Chinese to produce the word, *with* may not be a CPE feature, or it may have been used by English speakers only. *Long*, as a comitative preposition, first appears in a journal recording one of Tilden’s voyages to China in 1819.

(3) Misser Tilden, you must go long my for catche chow chow tiffin (1819; Jenkins, 1944: 13)
   ‘Mr Tilden, you must go with me to get something to eat.’

Comitative is the most frequently occurring meaning associated with *long* in CPE. Among all the attested uses of *long* in the corpora, 38 out of the 60 (i.e. 63%) attestations of *long* are comitative.

### 4.1.2 Benefactive and Malefactive

*Long* takes on two semantically opposite meanings: benefactive as in (4) and (5) and malefactive as in (6) and (7).

**Benefactive**

(4) Long me catchee ten piecee coolie (-1862; Tong, 1862: IV.49)
   ‘Get ten coolies for me’

(5) my long you makee alla proper (-1862; Tong, 1862: IV.33)
   ‘I will put it straight for you’

**Malefactive**

(6) my too much fear some war ship mans want for make bobbily long china mans.
   (-1835-; Tilden 1834-36: 968)
   ‘I fear very much that the sailors want to make troubles for the Chinese.’

(7) talk mi so fashion mi kick up bobbery along you. (-1876; Clough, 1876: 11)
   ‘If you talk to me in this manner, I will make trouble for you.’

A point to note is that these two meanings of *long* are found either in the Chinese or English sources. Benefactive *long* is attested in the Chinese sources only whereas malefactive *long* is found in English sources only (see Table 1). Another difference between benefactive *long* and malefactive *long* is shown in

---

7 An anonymous reviewer points out that while European languages often employ identical markers for benefactive and malefactive roles, Asian languages generally use different markers or constructions for these functions. In a cross-linguistic study of benefactive and malefactive
their syntax. As seen in (4-7) benefactive *long* is attested exclusively in preverbal position while malefactive *long* is found in postverbal position. This syntactic discrepancy will be discussed in section 7 where an examination of the contributions of lexifier and substrate languages will be presented.

4.1.3 Source

This function is equivalent to English *from*. Like benefactive *long*, it is only found in the Chinese source.

(8) My long you takee alla (-1862; Tong, 1862: VI.8)  
   ‘I will buy the whole from you.’
(9) my long he borrow (-1862; Tong, 1862: IV.57)  
   ‘I borrowed from him’
(10) I wantchee long you buy something (-1862; Tong, 1862: VI.1)  
   ‘I want to buy something from you’

All occurrences of source *long* are attested in preverbal position as in (8-10), which matches the word order of source PP in Cantonese grammar (Matthews & Yip, 2011).

<table>
<thead>
<tr>
<th></th>
<th>Preverbal</th>
<th>Postverbal</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comitative</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese sources</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>English sources</td>
<td>2</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td><strong>Benefactive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese sources</td>
<td>10</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>English sources</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Malefactive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese sources</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>English sources</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese sources</td>
<td>7</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>English sources</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>29</td>
<td>31</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 1 Distribution of the prepositional functions of *long* in CPE
A summary of the prepositional meanings of *long* is given in Table 1 which shows the number of attestation for each meaning and their distribution in Chinese and English sources.

4.2 *Long* as Coordinating Conjunction

A second function of *long* is coordination. Before looking at how *long* behaves as a conjunction in CPE, it should be mentioned that three forms of coordination are found in the corpora: a) juxtaposition; b) English conjunction *and*; and c) *long*. Juxtaposition is by far the most prevalent means of coordination in both English and Chinese sources. English *and* is a minor strategy which is overwhelmingly found in English sources.\(^8\)

Conjunction *long* is first attested in Tilden’s journal dated 1819.

(11) Willum! hab you plenty see my wifes long she my daughters?  \(^9\) (1819; Jenkins, 1944: 19)

‘William, have you seen my wives and my daughters?’

Other examples such as (12) from Hunter (1855) also point to the coordinating function of *long*.

(12) My long you No. 1 good flen (Hunter, 1855: 193)

‘You and I are very good friends.’

Conjunction *long* is primarily found in English sources. Coordination in the *Instructor* is predominately indicated by juxtaposition; however, there is one possible case of conjunction *long* as seen in (13a).

(13a) my long you go see he (-1862; Tong, 1862: VI.79)

‘We will go together to have an interview.’

---

\(^8\) There is only one instance of *and* in *The Chinese and English Instructor;*

My piecee goodsee number one good and strong (-1862, Tong, 1862: VI.5)

‘My piece goods are good and strong.’

\(^9\) I agreed with an anonymous reviewer that (11) could have a comitative interpretation with the translation ‘William, have you seen my wives together with them my daughters?’ I think the meanings of (11) is related to the interpretation of the pronoun *she* that follows *long*. In the comitative reading, *she* would be referring to the wives. However, in the same text, there are other instances of [pro NP] structure whose pronoun is co-referential with the following NP, as in (a):

(a) Willum! hab she my five piece daughters too much handsome? (1819, Jenkins, 1944: 19)

‘William, are my five daughters beautiful?’

So depending on the referent of the pronoun, both interpretations are plausible for (11). As for (12), the context suggests that a coordination reading is more likely.
Arguably a comitative interpretation of in (13a) is also possible; however, note that the phrase *my long you* are both translated as *we* in English (13a) and *daai6gaa1 ‘we’* in Cantonese (13b). This suggests that a conjunction reading of *long* may be more appropriate.

A closer look at the constituent type of the coordinants shows that *long* is used to link noun phrases only. This is evident from (11-13) and (14-15) below.

(14) Dat paper hab all same same as money: for make contente quie, and for talke talke to he for no to make burn city - houses - and spoil jonck ship in liber long boat outside in sea. (1833-; Tilden, 1833-34: 833)  
‘These papers were the same as money. They were used to make the ghost happy and to ask him not to burn the city and houses, and not to destroy the junks in the river and the boats in the sea.’

(15) Ayah! hab too much wind! long lain. Quie hab make ship jump downside in water. ten piece large china jonck ship - long some piece eulope ship. (-1835-; Tilden, 1834-36: 945) 
‘Alas! The wind was too strong. And there was prolonged raining. The ghost had sunk the ships: ten large China junks and some European ships.’

Unlike English *and* which is used to link all types of phrasal and clausal constituents, this property of *long* reflects the ways coordination is marked in Cantonese. This parallelism between CPE and Cantonese with respect to coordination will be discussed in detail in section 6.2.

5. The Lexical Origins of *long*

This section analyzes the processes that lead to the creation of *long* in CPE. It will be shown that *long* is not simply a form borrowed from the orthographically similar English *along*, but results from the recombination of features from the contributing languages. The closest candidate that matches CPE *long* seems to be English *along* (adverb, preposition). This is supported by CPE corpora where besides the form *long*, variants such as *along* and *‘long* are found. In view of the fact that English *along* has quite different meanings from CPE *long*, the question one needs to address is: how does CPE *long* derive from the English adverb/preposition *along*? In the following, I propose that the predecessor of CPE *long* may have been derived from the complex preposition *along with* and/or the adverb *along* which provide the semantic basis for the creation of *long* in CPE.
The prepositional meanings of English *along* are shown in (16-17); however, the meanings differ from the ones used in CPE.

(16) If you walk along this street, you’ll find the cinema.
(17) We want to plant roses along this path.

It could be hypothesized that comitative *long* is a metaphorical extension of the above meanings of *along*; however, given that English uses the complex form *along with* to denote comitative, this combination seems to provide a better match for comitative *long* in CPE. This hypothesis is supported by two facts: first, semantically *along with* has an intrinsic comitative meaning which is compatible with the comitative meaning of CPE *long*; second, historically, as has been shown in section 4.1.1, comitative is initially marked by *along with*, then the form *(a)long* emerges as the comitative marker in nineteenth century texts. Let’s compare the eighteenth century text in (1) repeated here as (18) with a nineteenth century text (19):

(18) Maske Maskee you come along with me (1769; Spenser, 1925: I.227)
‘Never mind, never mind, come along with me’

The complex form *along with* as a comitative marker is attested among the earliest CPE texts. The form comitative *long* does not appear until early nineteenth century as shown in (19).

(19) Misser Tillen, you must go long my for catche chow chow tiffi   n (1819; Jenkins, 1944 : 13)
‘Mr Tilden, you must go with me to get something to eat’

From a historical point of view, the development from the complex *along with* to a more simple form *(a)long* is a plausible explanation for the creation of *long* in CPE.

Apart from the complex form *along with*, the adverb *along* as used in (20) might be another factor that contributes to the emergence of *long*.

(20) Come along!

In (20), though no comitative NP is overtly expressed the meaning of accompaniment is implicit. Therefore, (20) could be analyzed as having a structure similar to *along with* NP, which can be represented as (21a) or (21b):

(21a) \[ \text{VP come \{PP along with NP\}} \]
(21b) \[ \text{VP come along \{PP with NP\}} \]

The structure in (21a) shows that *along* and *with* form a complex preposition, heading the PP which in turn forms a VP with *come*. In (21b) the PP is headed by the preposition *with* and *along* is a particle attached to the verb.
come. While (18) can be interpreted either way, (21b) is clearly the analysis for (20). So the hypothesis here is that both the explicitly expressed comitative NP in along with and the implicit accompaniment associated with the adverb along contributed to the selection of along as the comitative marker and eventually led to the creation of a new form long in CPE.

The use of along alone to refer to comitative also has a source in older form of English. According to the Oxford English Dictionary, along may be used for the sense of “in company, as a companion, with (someone)” as in the example in (22).

(22) 1882 HOWELLS Lady of Aroostook 137 ‘Our Captain’s wife…was not along,’ said Lydia. ‘Not along?’ repeated Mrs. Erwin. ‘Who were the other passengers?’ (Oxford English Dictionary Online)

This usage of along is dated, but the fact that it can be used in such context in the nineteenth century suggests that it could also be a model for CPE long.

As for the disappearance of with in the complex, the pronunciation of with poses some difficulties to Chinese speakers whose native languages lack dental fricatives. This could be one reason why with is only attested in English sources. Another factor is substrate reinforcement. Cantonese has a morpheme tung4 ‘with’ which is phonologically and semantically close to CPE long. Therefore, Cantonese tung4 may have acted as a catalyst for reinforcing the selection of (a)long. The association of English along and Cantonese tung4 can be explained in terms of the principle of perceptual salience which predicts that “a feature of the L1 is more likely to be transferred if there is a potentially stressed free morpheme in the L2 that can be used or reanalyzed according to the L1 pattern.” (Siegel, 1999: 24) Also, since long receives stress on the second syllable, it is perceptually more salient than with. Thus, having a morpheme along in the lexifier that refers to ‘accompaniment’ and a phonologically and semantically similar morpheme tung4 ‘with’ in the substrate maximizes the preference of long to with.

A conjunction-like property of along with in context such as (23) could also explain the coordinating function of long in CPE:

(23) Tom, along with Sally, will come to the ceremony.

This use of along with in (23) is functionally similar to the conjunction and in English, so that (23) can be rephrased as ‘Tom and Sally will come to the ceremony.’ A structurally similar example (24) is attested in the CPE corpora:

(24) Tom and Sally will come to the ceremony.
The near identical structure in (23) and (24) shows that the coordinating function, like the preposition *long*, could also have a source in the complex preposition *along with*. A reinforcing factor comes from the substrate – Cantonese which uses *tung4* ‘with’ not only as a preposition but also a conjunction as seen in (25):

> (25) ni1 di1 man4gin2 tung4 ni1 fung2 seon2 …
> 
> dem cl document and dem cl letter
> 
> ‘these documents and this letter …’

The analysis of the lexical origins of *long* shows that the processes involved in the creation a new lexical item in CPE involves more than simply borrowing a form from the lexifier and giving it the grammar of the substrate language as proposed in the relexification hypothesis (Lefebvre, 1998). In the case of CPE *long*, it clearly shows recombination of uses from both lexifier and substrate languages.

### 6. Substrate Transfer

In previous sections, it has been mentioned that *long* shows a number of features that are also present in the substrate language. This section focuses on three aspects of *long* in which substrate transfer is evident: (i) functions of *long*, (ii) comitative-conjunction syncretism, and (iii) word order of *long*-PPs. These aspects will be examined with respect to the grammar of CPE *long* and Cantonese *tung4*.

#### 6.1 Transfer of Functions from Cantonese

Let’s look at the meanings of Cantonese *tung4* in (26-29).

**Comitative**

> (26) ngo5 tung4 maa4maa4 heoi3 gung1jyun2
> 
> 1sg with mother go park
> 
> ‘I went to the park with my grandmother.’

**Benefactive**

> (27) ngo5 soeng2 tung4 go2 go3 neoi5jan2 zou6 faan1 di1 je5
> 
> 1sg want for dem cl woman do again cl thing
> 
> ‘I want to do something for that woman.’
Malefactive

(28) ?ngo5 tung4 keoi5 zai3zou6 zo2 hou2do1 maa4faan4
1sg for 3sg make ASP many trouble
‘I have created many problems for her.’

Source

(29) li1 tiu4 kwan4 ngo5 tung4 go2 go3 neoi5jan2 maa15 ge3
DEM CL dress 1sg from DEM CL woman buy SFP
‘I bought this dress from that woman.’

The semantic mapping between Cantonese tung4 and CPE long is clear when we compare the range of meanings that tung4 expressed from (26) to (29) with those expressed by long in (1-10) in section 3. Similar to long, Cantonese uses an identical morpheme to represent the above semantic roles, thus influence of Cantonese grammar on CPE long is apparent. English, on the contrary, uses different prepositions to mark the above semantic roles. Since comitative meaning of long seems to be derived from both English along (with) and Cantonese tung4, I will argue in section 7.3 that this is a case of semantic confl ation.

6.2 Comitative-Conjunction Syncretism

Further evidence in support of the substrate playing a significant role in the formation of CPE long is shown in the ability of this morpheme to serve both a comitative meaning as well as a conjunction function. In section 4.2 it has been mentioned that a second function of long is coordination. This feature is unlikely to be an influence from English which uses different markers for comitative (30) and coordinating conjunction (31).

(30) Patrick is having dinner with his brother.
(31) Patrick and his brother went out for dinner.

Cantonese, on the other hand, uses tung4 for these two functions.

(32) ngo5 maa4maa1 soeng2 tung4 nei5 sik6 caan1 faan6.
1sg mother want with 2sg eat CL meal
‘My mother wants to have dinner with you.’
(33) ngo5dei6 gam1maan5 sik6 gai1 tung4 haa1.
1pl tonight eat chicken and shrimp
‘We have chicken and shrimps for dinner.’

Grammaticalization of coordination markers from comitative markers is cross-linguistically very common, especially in African languages and Asian languages like Cantonese (Mithun, 1988). Examining the ways languages
marks noun phrase coordination, Stassen (2000) classifies the world’s languages into two main types: \textit{and}-languages and \textit{with}-languages. Since English formally differentiates ‘with’ and ‘and’, it is an \textit{and}-language. Cantonese is a \textit{with}-language because it possesses a coordinating conjunction which is identical in form with a word meaning ‘with’. Given that CPE demonstrates a typological feature that is only present in Cantonese but not in English, this is a strong indication that this syncretism pattern is transferred from the substrate.

Substrate transfer is also evident in the ways different constituent types are coordinated in English and Cantonese. Cantonese has a number of conjunctions and selection of the appropriate conjunction is determined by the types of conjoint constituents. \textit{Tung4} ‘with’ is used primarily for noun phrase coordination as in (34); \textit{jau6} ‘also’ is used for coordinating verb phrases as in (35) and adjectives as in (36); \textit{tung4maai4} ‘with’ is often employed for clausal coordination as in (37):

(34) ping4gwo2 tung4 hoeng1ziu1 ngo5 dou1 zung1ji3 sik6 apple and banana 1sg also like eat

‘I like to eat apples and bananas.’

(35) keoi5 jau6 coeng3go1 jau6 tiu3mou5 3sg also sing also dance

‘She sang and danced.’

(36) li1 tiu4 kwan4 jau6 gwai3 jau6 lou5tou2 dem cl dress also expensive also old-fashioned

‘This dress was expensive and old-fashioned.’

(37) ngo5 heoi3 zo2 syu1guk2 tung4maai4 tai2 zo2 tou3 hei3 1sg go asp bookshop and watch asp cl film

‘I went to a bookshop and watched a film.’

The translations of the above Cantonese sentences show that English uses \textit{and} for all types of coordination. In all instances of \textit{long} functioning as a coordinating conjunction, it occurs exclusively with noun phrases. Based on typological similarity and the selection of conjoint constituents, CPE \textit{long} demonstrates a strong parallelism with Cantonese.

6.3 Word Order of long-PPs

In Table 1 we saw that \textit{long} has variant word order, i.e. it can precede or follow the verb. This variation could be due to retention of the lexifier order and/or substrate influence. Here, I focus on how the syntax of PP in Cantonese is reflected in \textit{long}. In English, the canonical position of PPs is postverbal; while in Cantonese, PPs typically precede the verb. Given that both preverbal and
postverbal long-PPs are attested in CPE, as seen in the contrast between (38) and (39), the preverbal option must have come from Cantonese syntax because this option is not allowed for in English.

Preverbal long-PP

(38) he long one gentleman talkee (-1862; Tong, 1862: VI.39)
   ‘He is talking with a gentleman’

Postverbal long-PP

(39) he more better takee two piecee coolie long he (-1862; Tong, 1862: VI.55)
   ‘He had better take 2 coolies with him’

Also, Cantonese possesses a word order combination: [PP V] and [VO] which is cross-linguistically rare (Dryer, 2003). The preverbal PP order as shown in (38) is clearly not commonly found in most of the world’s languages; hence an apparent answer to the preverbal order of PP is substrate transfer. Transfer of word order is even more evident in cases where Cantonese is the only model for long. Comitative long can be attributed to both Cantonese tung⁴ and English along (with), so that what we find in the corpora is that long is predominantly postverbal in English sources and either preverbal or postverbal in Chinese sources. On the contrary, benefactive and source long are not based on English use of along (with), so the only possible source of grammar is Cantonese tung⁴ which only occurs in a position preceding the verb.

7. Multiple Etymologies of CPE long

The preceding section has shown that Cantonese has a significant influence on the grammar of long, it seems that the conventional thesis that the lexifier provides the form of the new lexical item of a pidgin/creole while the substrate language(s) contribute(s) to the grammar of that lexical item can account for the formation of long in CPE. However, I show here that the emergence of long does not seem to support such a neat division of labor. I argue along the lines of Mühlhäusler (1982) and Kihm (1989) and propose that the form and grammar of long is a result of conflating the phonology, syntax and semantics of Cantonese and English.

Mühlhäusler (1982) is among the earliest studies that proposes the idea of conflating in pidgins and creoles and challenges the traditional assumption of a single etymology for lexical items in pidgins and creoles. Examining the lexicon of Tok Pisin, he points out that a significant portion of the vocabulary may have multiple origins; the words in (40) are some examples (Mühlhäusler, 1982: 103).
An attempt to define confl ation in contact languages is given in Kihm (1989) where confl ation is described as a process whereby:

… confusing in one’s mind linguistic objects that in reality have no connection whatsoever, only because they sound alike and seem to fill identical functions. It follows that, when ascertaining cases of confl ation, we certainly do not need and do not expect to find the precise sound correspondences required by traditional diachronic studies. Superficial resemblances, regardless of phonemic or morphological constituency, are quite enough for our purpose, just as they apparently were quite enough for the people who drew on them during the process that led to the creation of a new, creolized language. (Kihm, 1989: 355-356)

Kihm sees confl ation as a process of approximating the sounds of different input languages in determining the functions and meanings of lexical items in pidgins and creoles. He examines the phenomenon of confl ation in the grammatical words in Kriyol, a Portuguese-based creole spoken in Guiné-Bissau. In Portuguese, the form *não* is used for negating sentence and predicate as seen in (41).

(41) *não, não  o  vi* (Kihm, 1989: 356)

no neg him I-saw

'No, I haven’t seen him.’

Kriyol, on the other hand, has two negation markers: *naw* for sentential negation and *ka* for negating verb as in (42).

(42) *naw, n ka  oja-l* (Kihm, 1989: 356)

no I neg see him/her/it

'No, I haven’t seen him/her/it.’

The etymon of *naw* is Portuguese *não*. The source of *ka* is generally thought to be Portuguese *nunca* ‘never’. However, Kihm argues that one problem in accepting this etymology is that while *ka* is always stressed, stress is placed on the first syllable of *nunca*. Therefore, Kihm questions the rationale for Kriyol speakers to retain the untressed syllable in *nunca* for predicate negation. The form *ka* is phonetically similar to a number of substrate forms: *dika* (43) and *kats* (44) in Manjaku.

(43) *man  dika  ro  wul* (Kihm, 1989: 357)

I neg-u do it

'I won’t do it.’
In Manjaku, *dika* functions as the unaccomplished negation and *kats*, meaning 'no longer', is a negative auxiliary. Other substrate languages also possess a range of negation markers, for example Mankanya has *nkö* and *kö*, Balanta has *ke* and *-ggə [kə]* and Mandinka has *buka*, *kana* and *kaka*. What is in common among these languages is that these forms can only be used for negating predicates. Also, they share a phonetic segment [kV], in which the V has the feature [-high]. Based on this evidence, Kihm argues that the Kriyol negation marker *ka* is reinforced by a phonetic resemblance with the substrate languages and the separate negation markers for sentential negation and predicate negation is also triggered by substrate grammar.

### 7.1 Phonological Conflation

Lefebvre (1998, 2004) argues that the phonological form of a new lexical item is derived from the lexifier language only. However, in a recent study of Papiamentu *ku*, Lefebvre and Therrien (2007) admit that the phonology of the substrate languages can also play a role. As has been discussed in the lexical origins of *long*, although the orthography suggests a closer connection with English *along*, the corresponding substrate morpheme *tung4* may also act as a reinforcing factor in the creation of *long*.

As argued in Kihm (1989), the creation of new lexical items in contact languages does not require exact correspondence between the contributing languages; superficial similarity seems to be quite sufficient for triggering phonological conflation. There is evidence from traditional Chinese phonology that people do not perceive sounds of words as indivisible parts. In Chinese phonological theory, *fănqiè* (反切), literally ‘reverse-cut’, is a technique for deriving the pronunciation of words. In using this method, the pronunciation of a Chinese character can be derived from two other characters. The first of the two characters provides the onset while the second of the two characters the rime. For example the pronunciation of 東 *dōng* in Mandarin is derived from two other characters: 德 *dé* and 紅 *hōng*. The way *fănqiè* works is like this: the pronunciation of 東 *dōng* is represented by taking the onset of 德 *dé* and combines it with the rime of 紅 *hōng*. In this way, another sound *dōng* is obtained. That *fănqiè* is a viable way for Chinese to learn new pronunciation is also supported by Tong’s employment of this method to teach English pronunciation in the *Instructor* where there was a section on *fănqiè*. 
Now let us see how fänqiè provides a way of explaining the creation of long in CPE. Mapping of English along [əlɒŋ] or [əlɔŋ] and Cantonese tung4 [tʰoŋ] might proceed in the following steps:

(i) [ə] deletion. Since the schwa [ə] in along is an unstressed syllable, it is susceptible to being lost during the formation process. The phonetic form of the Cantonese counterpart tung4 could also reinforce its disappearance.

(ii) Onset. The onset in along is [l] and the onset in Cantonese tung4 is [tʰ]. Mapping is difficult since the two consonants have quite different phonetic properties: [l] is a lateral while [tʰ] is a dental stop, but note that both are alveolar.

(iii) Rime. The rime in along ([ɒŋ], [ɔŋ]) could be slightly different depending on the dialects of the English speakers. The rime in Cantonese tung4 is [oŋ] (Bauer & Benedict, 1997). Therefore the rimes in English along and Cantonese tung4 might be perceived as more alike than the orthography might suggest.

Following the principles of fänqiè, it can be seen that notwithstanding the contrast in the onset, the rimes of along and tung4 demonstrate sufficient similarity to trigger association of the two words, especially given their shared semantics. So it is quite plausible that the creation of long may be partially attributed to the substrate language. The examples from Kihm (1989), Lefebvre and Therrien (2007) and CPE long show that the phonological forms of lexical items in contact languages could be influenced by both the lexifier and substrate forms.

7.2 Syntactic Conflation

In section 6.3 I showed the influence of Cantonese PP placement on CPE long. Although the substrate account is quite satisfactory in accounting for the preverbal word order of long-PP, it doesn’t explain the whole story because Table 1 shows that preverbal and postverbal uses of prepositional long are fairly evenly distributed. Besides phonological conflation, I argue that syntactic conflation, that is retention of lexifier syntactic features and influence of substrate structure, is also manifested in long.

Comitative long is attested in both Chinese and English sources but these sources show different preferences for the syntactic positions of long. Preverbal and postverbal positions are equally frequent in the Chinese source, whereas postverbal position is the dominant order in English sources (for figures see
Table 1). With verbs of motion such as go and come, while comitative PPs in English sources are attested as following the verbs as in (45)-(46), both preverbal and postverbal positions are attested in Chinese sources as seen in (47) and (48).

(45) belong have got one piecee flend, my go along he (-1860; Anon, 1860: 100)
   ‘I have a friend, I can go with him.’
(46) S’pose you no wantche look see, mi wantche you come along mi catchee samshoo
    (1859; Lindley, 1866: 13)
   ‘If you want to see, I want you to come with me and get some samshoo.’
(47) Monday you long he come (-1862; Tong, 1862: VI.52)
   ‘Bring him here on Monday’
(48) my wantchee you go long me (-1862; Tong, 1862: VI.42)
   ‘I want you to go with me.’

Examples (49-50) again show the strong preference for postverbal PPs in English sources:

(49) I like werry much, do littee pidgeon long you. (1836-37; Downing, 1838: I.279)
   ‘I would like very much to do some small business with you.’
(50) What thing Dr. Parker talkee ’long you? (1867; Coffin, 1869: 310)
   ‘What did Dr Parker say to you?’

With other verb classes, the comitative phrases also favor postverbal position in English sources; whereas in Chinese sources, long is often attested as occurring before the verb as in (51).

(51) he long one gentleman talkee (-1862; Tong, 1862: VI.39)
   ‘He is talking with a gentleman.’

Word order variation found in comitative long in Chinese and English sources may be attributed to the syntax of prepositional phrases in English and Cantonese. English is a language that consistently places prepositional phrases after the verb (52), whereas Cantonese typically places prepositional phrases before the verb (53).

(52) Sally is playing [pp with her little brother].
(53) Joe [pp tung4 keoi5 sai3lou2] waan2 gan2.
    Joe with 3sg little.brother play asp
   ‘Joe is playing with her little brother.’

With respect to the comitative role of long, there seems to be a clear correlation between syntactic positions and types of sources. The overwhelming majority of occurrences of long are placed in postverbal position in English sources, which indicates an apparent influence from English syntax. However, one should not neglect the fact that postverbal comitative long is equally
frequent in the Chinese source. This is one of the aspects of long where we find retention of lexifier syntax and thus no clear division of labor between substrate and lexifier.

While comitative long shows both preverbal and postverbal positions, other meanings of long seem to favor one syntactic position only: benefactive and source long are exclusively found in preverbal position in the Chinese source only, and malefactive long in postverbal position in English sources only. For benefactive and source long which have an apparent counterpart tung4 in Cantonese, their syntactic position conforms to that of Cantonese. For the malefactive role, there is doubt whether tung4 is the source of this meaning. In section 6.1 I gave (28), repeated here as (54), as the corresponding meaning of malefactive long:

(54) ngo5 tung4 keoi5 zai3zou6 zo2 hou2do1 maa4faan4
    1sg for 3sg make ASP many trouble
    ‘I have created many problems for her.’

However, a more natural way of rendering (54) is (55).

(55) ngo5 zai3zou6 zo2 hou2do1 maa4faan4 bei2 keoi5
    1sg make ASP many trouble3g
    ‘I have created many problems for her.’

If malefactive long is modeled on bei2 ‘give/for’ instead of tung4, this could explain why this particular meaning of long occurs in postverbal position. The word order discrepancy also leads one to suspect that there might be lectal differences between Chinese- and English-speaking CPE users. Since this requires more detailed comparison, I leave this speculation for further research.

According to Lefebvre’s (1998, 2004) relexification hypothesis, the syntax of a pidgin/creole lexical item is derived from the syntax of substrate languages. This predicts that long should show Cantonese syntax, which is true to a certain extent but does not explain the whole story because what has just been shown is variation between the lexifier and substrate word order. Lefebvre argues that word order variation may exist as a result of different syntactic patterns found in substrate languages and leveling eventually settles on one syntactic structure. While this account may work for Haitian Creole which has a number of substrate languages, it cannot satisfactorily explain the syntactic variation manifested in CPE long. CPE originated and flourished in Canton, where Cantonese was undoubtedly the major substrate language. Even when CPE was later spread to other treaty ports along the China Coast in the second half of nineteenth century, hard evidence is lacking to suggest that other dialects such as Shanghainese or Min played a crucial role in the development of
CPE. Besides, Yue, Min and Wu dialects behave homogeneously with respect to the syntactic structure of comitative, i.e. comitative markers in all these dialects occur in preverbal position (Qian, 1997; Chappell, 2000). Therefore, the difference in word order attested in comitative long should be attributed to a contrast in the syntax of comitative PP between Cantonese and English.

7.3 Semantic Conflation

Lefebvre’s relexification hypothesis also asserts that the semantics of the creole lexical items is derived from the substrate language. The case of CPE long shows that the lexifier also makes contribution to the semantics of the new lexical item. While Lefebvre’s representation of relexification hypothesis does not attribute the semantics of the creole lexical items to the lexifier, semantic conflation is not incompatible with relexification. As Lefebvre claims, since relexification is a semantically driven process, in order to create a new lexical item in pidgins and creoles there must be some semantic overlap between the lexifier and the substrate languages. Muysken (1981: 62) also expresses a similar proposal: “For relexification to occur, the semantic representations of source and target language entries must partially overlap; otherwise, the two entries would never be associated with each other. Other features of the two entries may, but need not, be associated with each other.” This notion of partial semantic overlap is important to our understanding of the emergence of long. Cantonese tung4 ‘with’ and English along (with) both express the meaning of ‘accompaniment’ or ‘togetherness’, so possibly comitative long originates from the complex form along with which is then simplified to (a)long under the reinforcement of a phonological, syntactic and semantic similar form tung4 in Cantonese. Hence, comitative long in CPE can be considered as jointly derived from Cantonese and English.

8. Conclusions

In tracing the origins and grammar of long, it is striking to see the interplay of the lexifier and substrate in the process of its creation. Conventional approaches such as Lefebvre’s (1998, 2004) relexification hypothesis often ascribe a simple division of labor with respect to the contribution of the lexifier and substrate languages, i.e. the phonetic representation comes from the lexifier and the grammar is derived from the substrate, to the formation of new lexical items in pidgins and creoles. Since Lefebvre (1998, 2004) claims that relexification is a mental process, psychological reality should be demonstrated to show that
the demarcation of source language contributions is mentally viable. In adopting relexification as one of the processes for pidgin and creole formation, Lefebvre (1998: 7) argues that “In this approach, each module is independent from the others. Hence, phonological representations may be treated independently from the semantic and syntactic properties that define the functions of particular lexical entries.” However, results from linguistic and psycholinguistic studies do not support such a dissociation of grammar from lexicon (Goldberg, 1995; Bresnan, 2001; Jackendoff, 2002). Using data from child language development, language development in normal children and children with language impairment, aphasic patients and real-time processing, Bates and Goodman (1997) conclude that results from these studies show that lexicon and grammar are inseparable and that a lexicalist approach to language is to be preferred to a modular account.

Even though it is true that the majority of a pidgin or creole’s lexicon is derived from a single language, in most cases the lexifier, this study on CPE long and previous studies on confl ation show that multiple etymologies in pidgin/creole lexical items may not simply be regarded as incidental forms—functions convergence. From the point of view of learning, it could be a strategy that learners use to facilitate the creation and learning processes of new lexical items. The case of long demonstrates that a straight division of labor between lexifier and substrate is perhaps too simplistic an account to explain how pidgin/creole speakers recombine available features in their native languages. CPE long clearly shows that an analysis involving multiple etymologies provides a more adequate answer to the origins and grammar of this pidgin lexical item.

The present study also stimulates further research on the relationship between CPE and Melanesian Pidgin. Mühlhäusler (1986: 160) notes that most pidgins of the Pacific region are characterized by having a one-preposition or two-preposition system with long being the only or the most common preposition. For example, long in Tok Pisin can refer to a range of spatial and temporal relations usually translated as English prepositions at, in, on, to, by, from, with, etc. In CPE, long is certainly the most important preposition. The functions of long in Melanesian Pidgin have been well documented (Verhaar, 1995; Crowley, 2004). This study not only presents a thorough description of the functions of long in CPE, but also provides significant implications for reconsidering the historical connections between Pidgin English in the Pacific. Based on his corpus, Baker (1987) rejects the idea that long in Pacific Pidgin English is diffused from CPE long which occurs later than Pidgin English in the Pacific. However, new findings point to the contrary of Baker’s argument. CPE long was attested as a comitative preposition as early as 1819,
earlier than its appearance in New South Wales in 1827. This suggests a new possibility for re-examining the historical relationships between CPE and Pacific pidgins. However, unlike in Melanesian Pidgin, CPE long is not an all-purpose preposition, so this function is probably an innovation of Melanesian Pidgin as argued in Clark (1979). As for the ultimate origin of long, Baker and Mühlhäusler (1986: 557) claim that “In the absence of any pre-1826 Pacific texts in which (a)long(a) has only the sense of ‘with’, there are no strong grounds for supposing that the Australian diffused feature has an Atlantic origin. However, CPE along may well have been introduced by sailors from the Atlantic region”. Whether CPE long originates from the Atlantic varieties and then diffused to Pacific pidgin awaits more investigations. Baker (1987: 181) inclines to treat the forms and meanings of (a)long in CPE, Pacific Pidgin English and Atlantic English lexifier pidgins and creoles as the result of independent development.11

The present study not only contributes to an understanding of the multifunctionality of long in CPE, it also invites new thoughts on the emergence of pidgin grammar and the interrelationship between different varieties of Pidgin English. Given that long is regarded as a worldwide feature in pidgins/creole (Clark, 1979; Baker & Huber, 2001), it is important to address the question of why this form or its variants is selected in different contact languages and whether the form and functions of long in Atlantic and Pacific Pidgin English are independent developments or are connected in some ways during their formation periods. Future research on CPE, one of the oldest pidgins, may shed light on these questions.

Acknowledgments

I would like to thank two anonymous reviewers for their valuable comments and suggestions. I am also grateful to Stephen Matthews and Umberto Ansaldo for reading and correcting earlier versions of this paper.

11 A reviewer suggested that since trade with China was monopolized by the East India Company until 1833, there was little prospect for direct trade between China and the Pacific. Thus, long was unlikely to be a feature diffused to Pacific Pidgin English from CPE. However, European ships carrying sandalwood from Fiji to China were already conducting a flourishing trade in the early nineteenth century (Dodge, 1976). Although Chinese and Pacific islanders may not have direct contact, a possible route for pidgin features to be diffused to other varieties of pidgin English is via European traders and sailors. The connections and contact situations between Chinese, Europeans and Pacific islanders require further research.
References


**Primary texts**


Feng, Xue-xing (馮雪卿) et al. 1860. *Yinghua Zhujie《英話註解》*. Shanghai.


Tilden, Bryant P. 1833-34. Journal. Fifth voyage to China, in brig Nabob, via Batavia 1833-34.

