What Can the Corpus of Mid-20th Century Hong Kong Cantonese Tell Us About Hong Kong Society of Half a Century Ago?

Andy Chin
The Education University of Hong Kong

Abstract   This paper reports on a corpus-based sociolinguistic study of terms of address with a special focus on kinship terms found in *The Corpus of Mid-20th Century Hong Kong Cantonese*, which has a size of about one million Chinese character tokens. The corpus data was collected by transcribing the speech dialogues of 81 black-and-white movies produced in Hong Kong between 1940 and 1970. The kinship terms extracted from the corpus can tell us about the family structure and marital life of Hong Kong six decades ago.

Keywords   Corpus-based sociolinguistic study. Cantonese corpus. Early Hong Kong society. Terms of address. Family culture.

Summary   1 Introduction. – 2 The Corpus of Mid-20th Century Hong Kong Cantonese. – 3 Applications of HKCC: Tracking Changes of Society. – 4 Kinship Terms and Family Culture. – 5 Terms of Address in HKCC. – 5.1 Terms of Marriage. – 5.2 Terms of Kinship. – 5.3 Other Terms of Address for Family Members. – 6 Concluding Remarks.

1   Introduction[1]

Baker (2010) commented that cross-fertilisation between two seemingly unrelated disciplines, namely corpus linguistics and sociolinguistics, has been done very little although the two disciplines have established their traditions in the field of linguistics for a long time. Baker explained that this may be due to the fact that corpus linguistics sometimes gives the impression that it "has made only a relatively small impact on sociolinguistics" (2010, 1). In spite of this, Baker (2010, 8-9) showed that the two disciplines share a lot of common features: a) analysing naturally occurring and empirical language data; b) emphasising on language-in-use or social context; c) making use of quantitative methodologies; d) examining and comparing variations and changes; e) providing explanations for the findings. All these common features demonstrate that these two disciplines can produce cluster research. One notable example is Davies' study of "issues related to culture and society, either in terms of change over time or variation between [English] dialects" (2017, 19) by means of various gigantic English corpora.[2] For example, Davies (2017, 27) found that, with data from GloWbe, the word 'terrorism' appears more in the varieties of English spoken in South Asian countries, such as Pakistan and Sri Lanka, than in British English and American English. Furthermore, he found that Australian English has more word types with the suffix *-ies* than other varieties of English in the Inner Circle *à la* Braj Kachru's model of World Englishes.

One research area in sociolinguistics seeks to examine language variations and changes either in diachronic or synchronic dimensions. Adopting a corpus-based approach to study linguistic variations from a diachronic perspective entails that one has to look for

---

historical data or to construct a historical corpus. This is not an easy task when one wants to collect real-time language data produced from the past. As McEnery and Hardie put it,

> for these and other extinct languages there is a fixed "corpus" of surviving texts which will never grow any further, except in the rare circumstance that hitherto unknown texts are discovered. An electronic corpus composed of all of these surviving texts (or a sampled subset of them) is thus the ideal tool for taking into account as much data on these historical forms as possible in an analysis of how language has changed. (2012, 94-5)

A corpus-based study of the diachronic development of a language will become fruitful and illustrative only when we manage to collect and process language data produced in the period we want to examine. At the same time, we also need to ensure that the corpus data we collect is "representative", "balanced" and "comparable" (McEnery, Hardie 2012, 10), although it is always not easy to have a corpus that perfectly meets all these three attributes.

2 The Corpus of Mid-20th Century Hong Kong Cantonese

This paper introduces a corpus-based sociolinguistic study of kinship terms in Hong Kong Cantonese, a language spoken as a home language by nearly 90% of the population in Hong Kong.[3] The data comes from *The Corpus of Mid-20th Century Hong Kong Cantonese* (hereafter HKCC) developed at The Education University of Hong Kong since 2011.[4] The data of HKCC was collected by transcribing the speech dialogues of 81 black-and-white movies produced in Hong Kong between 1940 and 1970. There are two phases of corpus development, at different stages and with different sources of funding.[5] The two phases of HKCC have processed spoken Cantonese data with a size of nearly one million Chinese characters.[6] The transcribed data of both phases in HKCC was tokenised and assigned with Cantonese pronunciations. The data in the second phase of HKCC was also annotated with parts-of-speech.

---

[3] See table 3.12 of CSD 2016. For the sociolinguistic situation of Hong Kong, see Tsou 1997 and Bacon-Shone, Bolton, Luke 2015.

[4] The URL of HKCC is `http://hkcc.eduhk.hk`.

[5] Dialogues of 21 and 60 movies were transcribed in the first and second phases respectively. HKCC is now available online for searching.

[6] Dialogues of three genres of movies were transcribed in HKCC: a) melodramas with themes on family and romance; b) detective and suspense; c) comedy.

Chin (2013; 2019a) provided detailed descriptions of the two phases of HKCC, including the data source and the rationales behind the construction of the corpus. The primary aim of HKCC is to provide real time language data for conducting diachronic studies on Cantonese and comparing the Cantonese language spoken in Hong Kong in the contemporary period and that of half a century ago. The HKCC data also bridges the gap of Cantonese linguistic research on early Cantonese (back to early 19th century) and contemporary Cantonese. Specifically, the mid-20th century is a transitional period in which some critical linguistic changes took place in Cantonese: the corpus data can thus provide authentic language data to examine the switchover from the old features to the new features.[7]

Another important feature of HKCC is that it can supply quantitative and qualitative information for examining the characteristics of the Cantonese language. HKCC can generate lists of segmented tokens according to their parts-of-speech and usage frequency, which can provide useful data for selecting items for compiling learning and teaching materials. Furthermore, the sample sentences based on the movie dialogues can allow users to have a better understanding of the use of language in context. Although one many argue that the data of HKCC comes from half a century ago and may be considered outdated and unsuitable for language teaching and learning, HKCC is still valuable because some of the usages and sentence patterns had not changed significantly since mid-20th century. This is especially the case for function words such as aspect markers, which have exceptionally high occurrences in HKCC. For example, the perfective aspect marker 咗 *zo2*[8] has a frequency of 3,300 in HKCC, which is far more than its occurrence (869 tokens) in HKCanCor.[9] To our best understanding, no existing learning and teaching resources can provide comparable amount of data and sample sentences for illustration. In addition, the search functions of the second phase of HKCC have been significantly enhanced so that users can incorporate flexible search criteria such as 'Numeral + Classifier + Noun' to retrieve more results for analysis and comparison.[10]

---

7 Some examples include the development of neutral questions (also known as Yes-No questions) and indirect object markers (also known as dative markers). For details, see, for example, Cheung 2001 and Chin 2011 respectively.

8 Cantonese examples are transcribed with the Jyutping Romanisation scheme developed by The Linguistic Society of Hong Kong. For details, see `https://www.lshk.org/jyutping`.

9 HKCanCor (*The Hong Kong Cantonese Corpus*) was developed by Professor Luke Kang Kwong at the University of Hong Kong in the late 1990s. The corpus has 869 occurrences of 咗 *zo2* out of 180,000 word tokens. The corpus data can be downloaded from `http://compling.hss.ntu.edu.sg/hkcancor`. For details of HKCanCor, see Luke, Wong 2015.

10 For the search functions in the second phase of HKCC, see Chin (forthcoming).

While there are Cantonese corpora developed in the past two decades, none of them is comparable to HKCC in terms of size and data source.[11] In spite of the availability of Cantonese corpora, linguistic research with Cantonese corpus data mainly focuses on the internal system such as syntax, lexicon, and phonology. This can be seen from a search of the keywords 'corpus' and 'Cantonese' in Google Scholar. Some of the research outputs include, for example, loanword truncation in Cantonese (Luke, Lau 2008), comparisons of temporal and tonal aspects in Mandarin and Cantonese (Peng 2006), the GIVE-construction in Mandarin and Cantonese (Wong 2009), the analysis of type and token frequencies of phonological units in Hong Kong Cantonese (Leung, Law, Fung 2004), the verbal suffix 着 *zoek6* (Lai, Chin 2018). These sample studies show how corpus data can enhance our understanding of the linguistic properties of Cantonese. However, they are still limited to language internal features. There are in fact many extra-linguistic issues that can be pursued with corpus data. One of the merits of HKCC is the dialogic and highly interactive nature of its data. It is thus useful for studying issues on discourse, pragmatics and sociolinguistics, which are relatively under-explored in Cantonese linguistic research. The author and his research team have conducted a number of studies on Cantonese discourse with data from HKCC. For example, Tse and Chin (2015) examined the features of co-referential noun phrases such as 你個衰人 *nei5 go3 seoi1jan4* 'you CLF bad guy, you the bad guy', that have the same surface structure as the possessive noun phrase with a classifier used as possessive marker, such as 你個公仔 *nei5 go3 gung1zai2* 'you CLF doll, your doll'. Chin (2018a) explored discourse markers including the tag questions 好唔好 *hou2 m4 hou2* 'is it alright' and sentence final particles. Chin (2018b) compared the two Cantonese prohibitive markers 唔好 *m4hou2* and 咪 *mai5*, which are usually treated as synonyms in Cantonese dictionaries and textbooks. The study examined the verbs these two prohibitive markers take, as well as the length of the verb phrases. It is interesting to see that each marker shows some distinct features which are not found in the other marker.

---

11 For details on the nature and data source of other Cantonese corpora, see Chin 2013; 2019a.

3 Applications of HKCC: Tracking Changes of Society

HKCC is important and useful for studying variations and development of Hong Kong Cantonese over time. There are lexical items and syntactic structures in HKCC which are no longer active in contemporary Cantonese. Examples include 霎氣 *saap3hei3* 'having an argument with someone', 蘇蝦 *sou1haa1* 'baby'. As for syntactic structures, we can find both old and new patterns co-existing in the same sentence, i.e. hybrid forms.[12] Besides linguistic analysis, we can also make use of the data from HKCC to examine sociocultural issues, because the content of the movies can reflect the popular and key social issues of Hong Kong society of the period concerned. Lui (1988) studied the housing issue of Hong Kong in the 1950s with reference to two melodrama movies, namely *In the Face of Demolition* (危樓春曉, 1953) and *The Kid* (細路祥, 1950).[13] Specifically, Lui argued that

> these films do provide corroborative evidence in understanding the decade of the 1950s. The feeling among Hong Kong people that the government should play a leading role in solving their housing problem grew only in the past ten to twenty years. (1988, 90)

In his study of Cantonese melodrama with the theme of familial relationships in the 1950s and 1960s, Law observed that the disappearance of Cantonese melodrama after the 1960s could be due to "rapid modernisation of Hong Kong" and "the spread of the nuclear family as the basic social unit and its accompanying individualism". These changes of social life and interpersonal relations "outstripped the development of the form and content of Cantonese melodrama" (Law 1986, 19).

The above two studies of Hong Kong society through early Cantonese movies show that movies can act as a telescope allowing us to look at some deeper issues of the community in which they are depicted. As language is argued to be the carrier of culture, we can thus observe, through the movie dialogues, what was being practised by people, as well as the characteristics of the social life and culture in the community concerned.

Mid-20th century saw the booming of Hong Kong's movie industry. According to Chung (2004), more than 1,500 movies, literally known as 'Cantonese long movies' (粵語長片 *jyut6jyu5 coeng4pin2*), were produced between 1950 and 1960. The dialogues in these movies can be claimed to have faithfully recorded the Cantonese language

---

[12]   One example is neutral questions produced in the movies included in HKCC. For details, see Chin 2019b.

[13]   These two movies were also included in HKCC.

spoken in Hong Kong at that time. Some of these Cantonese movies have their stories centring on the social situation of Hong Kong of that time. Some of the themes include familial relationships, especially conflict of interest among family members, romance among young people, and tragedies arising from social issues such as poverty and humanity. We thus believe that the data from HKCC can serve as a good resource for conducting a corpus-based sociolinguistic study.

In the following, based on the data extracted from HKCC, we will examine the kinship terms and lexical items related to family and marriage with an aim to explore the family culture and family organisation in Hong Kong half a century ago.

4   Kinship Terms and Family Culture

Terms of address are lexical items used to address a person in conversations. For kinship terms which are used to refer to family members, the amount and complexity are highly correlated with the concepts of family structure in the respective speech community. There have been numerous studies comparing the kinship term systems between the Chinese language and other languages such as English. It is generally acknowledged that kinship terms in Chinese have a "finely grained semantic structure" (Qian, Piao 2009, 190), which can be associated with the complex family structure of Chinese society. For example, Chinese families reflect the patrilineal character (Wu 1927) and this is rendered in the kinship terms referring to grandparents. Kinship terms for maternal grandparents carry the prefix 外 *ngoi6*, literally 'external, outside', such as 外公 *ngoi6gung1* 'maternal grandfather' and 外婆 *ngoi6po4* 'maternal grandmother'. Furthermore, Chinese kinship terms make distinction in terms of age and gender, while English in some cases uses one single kinship term instead.[14] Typical examples are *uncle*, *aunt* and *cousin*. All these differences between kinship terms in Chinese and English can reflect the family structures of the two cultural traditions.

We can also have a look at the family structure of early Hong Kong by examining the kinship terms found in HKCC. As we discussed in § 2, the movies we selected to transcribe cover three genres, namely melodrama, detective and suspense, and comedy. Many of these movies have their stories and plots centring on family members. For example, in some suspense movies, the stories were about disputes among family members, such as brothers and sisters fighting for the

---

14   Taking all these attributes into consideration, kinship terms in Chinese (including its dialects) can be examined by means of componential analysis. See, for example, Chao 1956; McCoy 1970; Cheung 1990; Qian, Piao 2009.

property left by their parents. Sometimes members of extended families such as uncles and aunts were also involved in the story.

Furthermore, it is noted that "propositional synonyms" referring to "a single kinship concept" always exist (Qian, Piao 2009, 193). These are also interesting terms that we can examine as they may signify different styles or degrees of solidarity between the addresser and the addressee. This will be discussed in § 5.3.

Besides kinship terms, we will also examine words related to the concept of marriage. Kinship relationships are built upon marriage between a man and a woman although, in modern society, families with single-parent, single-child, same-sex couples or heterosexual cohabiting partners give rise to many new kinship terms, as illustrated by Qian and Piao (2009). In other words, the examination of kinship terms of different time periods can allow us to observe the development of society in terms of marital life and family organisation.

5 Terms of Address in HKCC

5.1 Terms of Marriage

Before examining the kinship terms in HKCC, let us start with the concept of *marriage*, which is the foundation for family organisation. Besides core terms like 婚姻 *fan1jan1* 'marriage' and 結婚 *git3fan1* 'getting married', we also searched for words describing different stages in the marital journey. These lexical items and their frequencies in HKCC are shown in table 1.[15]

Table 1 Lexical items related to the concept of 'marriage' in HKCC

Term related to marriage*	Meaning	Frequency in HKCC
相睇 *soeng1tai2*	blind date	8
婚姻 *fan1jan1*	marriage	47
拍拖 *paak3to1*	dating	29
求婚 *kau4fan1*	proposal	29
訂婚 *ding6fan1*	engagement	58
結婚 *git3fan1*	getting married	379
離婚 *lei4fan1*	divorce	44
嫁 *gaa3*	marry a man	376
娶 *ceoi2*	take a wife	206

---

15 Unless stated otherwise, the data of HKCC are based on the second phase, which has about 800,000 Chinese character tokens.

媒人 *mui4jan2*	matchmaker	30

<table>
<tr><td>* The frequency also includes items such as 離咗婚 lei4 zo2 fan1 'having divorced', which was segmented in HKCC into three tokens: 離, 咗, 婚.</td></tr>
</table>

Among the terms associated with marriage, 結婚 *git3fan1* 'get married' has the highest frequency, suggesting that this is one of the major events in movies with plots on romance and familial relationships.

In traditional Chinese families, children's marriage is always arranged by their parents, possibly through a matchmaker and blinddates. The relevant words 媒人 *mui4jan2* 'matchmaker' and 相睇 *soeng1tai2* 'blind date' appear 30 times and 8 times respectively in HKCC, as shown in table 1 above. This kind of marital arrangement received a lot of criticism as young people tended to bargain for more freedom and autonomy in their own marriage. In the following dialogues, we can see the pre-arrangement of marriage by senior family members.

1. *Your Infinitive Kindness* (恩義難忘, 1965)
    婚姻大事都係由老人家作主好啲嘅
    fan1jan1 daai6si6 dou1hai6 jau4 lou5jan4gaa1 zok3zyu2 hou2 di1 ge2
    'It is better for the elderly to decide on the marriage'.

2. *Love Burst* (難為了嬌妻, 1966)
    婚姻大事係要聽父母之命媒酌之言
    fan1jan1 daai6si6 hai6 jiu3 teng1 fu6mou5 zi1 ming6 mui4zoek3 zi1 jin4
    'Marriage has to be based on parents' order and matchmaker's word'.

We also see how young people feel against the tradition of having marriage arranged by their parents or other senior members such as grandparents in the family. The following dialogue shows an argument between a father and his daughter.

3. *Foster-Daddy's Romantic Affairs* (契爺艷史, 1952)
    Father: 你嘅婚姻事爸爸會同你揸主意㗎。
    nei5 ge3 fan1jan1si6 baa4baa1 wui5 tung6 nei5 zaa1 zyu2ji3 gaa3
    'Daddy will take care of your marriage'.
    Daughter: 爸爸, 婚姻嘅事情我哋自己會理㗎啦。
    baa4baa1, fan1jan1 ge3 si6cing4 ngo5dei6 zi6gei2 wui5 lei5 gaa3laa3
    'Daddy, we can take care of our marriage'.

The following dialogue illustrates how young people feel dissatisfied toward pre-arranged marriage and ask for freedom on the decision of their marriage.

4. *Stubborn Love* (癡兒女, 1943)
   取消呢種封建嘅婚姻制度。
   ceoi2siu1 ni1 zung2 fung1gin3 ge3 fan1jan1 zai3dou6
   'We need to abolish this kind of feudal style of marriage system'.
   而且婚姻要自由呀。
   ji4ce2 fan1jan1 jiu3 zi6jau4 aa3
   'Furthermore, we need to have freedom in marriage'.
   阿媽點都唔能夠強迫我婚姻自由。
   aa3maa1 dim2 dou1 m4 nang4gau3 koeng4bik1 ngo5 fan1jan1 zi6jau4
   'Mother cannot take away my freedom of marriage'.

It is also common for parents (especially those of a daughter) to have business partners as their potential in-laws. There is one proverb in Chinese, namely 門當戶對 *mun4dong1wu6deoi3* 'families of equal rank', advocating for marriage between people with similar backgrounds. In spite of this old-fashioned mindset, there were sometimes parents who were open-minded and willing to allow their children to choose their lifelong partners. Dialogue (5) below is an utterance made by a mother to her daughter, whose marriage was arranged by her father.

5. *When Girls are in Love* (女生外向, 1965)
   Mother: 我時時都唔贊成你爸爸將佢嘅生意
   ngo5 si4si4 dou1 m4 zaan3sing4 nei5 baa4baa1 zoeng1 keoi5 ge3 saang1ji3
   同埋你嘅婚姻拉埋一齊。
   tung4maai4 nei5 ge3 fan1jan1 laai1maai4 jat1cai4
   'I have never agreed with your father in linking his business with your marriage'.

What the above dialogues extracted from HKCC show is that marriage in the old days was not necessarily built upon love and could be arranged by parents without the consent of the children. In a survey conducted by Podmore and Chaney with 1,123 respondents aged between 15 and 30 in the 1970s, 91% indicated that "love was the appropriate basis for marriage" (1974, 403), while 94% of the respondents were "against the idea of arranged marriage" (404). In this connection, it is relevant to examine the verb 娶 *ceoi2* 'to marry a woman' as it can take two different objects: 老婆 *lou5po4* 'wife' and 新抱 *san1pou5* 'daughter-in-law'. The two verb-object phrases capture different perspectives on 'marrying a woman'.[16] The former takes the perspective of the son, while the latter that of the parents. In HKCC, the two phrases have 83 and 14 occurrences in HKCC respectively. Interestingly, among the 83 phrases of 娶老婆 *ceoi2 lou5po4* 'taking

---

[16] It is interesting to note that the verb 嫁 *gaa3* 'to marry a man' does not have such a dual usage. This verb can only be used to mean 'marrying a man to be his wife'.

a wife', 28 contain a prepositional phrase headed by 同 *tung4* 'for', carrying the meaning of *for*. Two examples are given below.

6. *She's so Neat* (彩鳳引金龍, 1957)
而家同你娶老婆囉
ji4gaa1 tung4 nei5 ceoi2 lou5po4 bo3
'We are now going to take a wife for you'.

7. *Standard Husband* (標準丈夫, 1965)
你快啲話畀佢聽同佢娶老婆嘅咪得囉
nei5 faai3di1 waa6 bei2 keoi5 teng1 tung4 keoi5 ceoi2 lou5po4 gam2 mai6 dak1 lo1
'You'd better tell him that we are going to take a wife for him'.

The adjunct phrase headed by 同 *tung4* 'for' shows that the act of taking a wife is not necessarily initiated by the son himself, but by someone in his family, such as parents or even grandparents. For the verb phrase 娶新抱 *ceoi2 san1pou5*, the subject is always the parents, and we do not find the adjunct phrase headed by 同 *tung4* (see the three examples below), which re-affirms that the act of marrying a woman as one's wife could be done sometimes by the family. From example (10), we can even see that in some families, getting a daughter-in-law (i.e. 娶新抱 *ceoi2 san1pou5*) is more important than marrying off the daughter (i.e. 嫁女 *gaa3 neoi5*).

8. *Lovesick* (為情顛倒, 1952)
我阿媽成日都想娶新抱。
ngo5 aa3maa1 seng4jat6 dou1 soeng2 ceoi2 san1pou5
'My mother always wants to get a daughter-in-law'.

9. *The Merry Matrimony* (喜結良緣, 1966)
阿強媽想快啲娶新抱呀嘛。
aa3 koeng4 maa1 soeng2 faai3di1 ceoi2 san1pou5 aa1maa3
'Ah Keung's mother wants to get a daughter-in-law as soon as possible'.

10. *Foster-Daddy's Romantic Affairs* (契爺艷史, 1952)
嗽呀梗係娶咗新抱先至嫁女㗎囉。
gam2 aa6 gang2hai6 ceoi2 zo2 san1pou5 sin1zi3 gaa3 neoi5 go3bo3
'Then, we certainly take a daughter-in-law before we marry off the daughter'.

The above HKCC dialogues containing words related to 'marriage' show the family structure and the arrangement of marriage in mid-20th century Hong Kong. Generally speaking, it was considered a normal practice for someone to get married when they become adults. If the children did not have any intention to form their own families,

their parents would do that for them by all means. In other words, the concept of family is somewhat important in the old days of Hong Kong, as the majority of the population in Hong Kong were Chinese who follow the tradition that men and women form their own families through marriage (Wu 1927; Baker 1979). In the next section, we will examine the kinship terms found in HKCC.

5.2 Terms of Kinship

Since the data in HKCC was only tagged with parts-of-speech, it is not easy to extract kinship terms as a semantic notion directly from HKCC. However, as Qian and Piao (2009) show, there are some unique morphemes referring to kinship. We thus compiled a list of Cantonese kinship morphemes, plotted on a simplified family tree according to the generations they belong to in a traditional Cantonese family [fig. 1].

G+2	公 *gung1* 'maternal grandfather', 婆 *po4* 'maternal grandmother'	爺 *je4* 'paternal grandfather', 嫲 *maa4* 'paternal grandmother'	
G+1	媽 *maa1* / 母 *mou5* / 娘 *noeng4* 'mother'	爸 *baa1* / 父 *fu6* / 爹 *de1* 'father'	
G	堂 *tong4*, 表 *biu2* (prefixes for cousins) 哥 *go1* / 兄 *hing1* 'elder brother', 姐 *ze2* / 姊 *zi2* 'elder sister' 嫂 *sou2* 'sister-in-law', 夫 *fu1* 'husband'	EGO	堂 *tong4*, 表 *biu2* (prefixes for cousins) 弟 *dai6* 'younger brother', 妹 *mui6* 'younger sister' 婦 *fu5* 'wife', 夫 *fu1* 'husband'
G-1		仔 *zai2* 'son', 女 *neoi5* 'daughter', 婿 *sai3* 'son-in-law', 新抱 *san1pou5* 'daughter-in-law', 甥 *sang1* 'child of sister', 姪 *zat6* 'child of brother'	
G-2		孫 'grandchild' *syun1*	

Figure 1 Cantonese kinship morphemes

The above is not an exhaustive list but these morphemes cover the basic kinship that a traditional Hong Kong family might have.[17] With these kinship morphemes, we were able to retrieve about 100 kinship terms from HKCC. Among these 100 items, some are core and common kinship terms such as *father*, *mother*, *brother*, and *sister*, which are listed in table 2.

In addition, there are a few items referring to members of extended families in the grandparents' generation: 叔公 *suk1gung1* 'the younger brother of the paternal grandfather' (i.e. father's paternal uncle); 姑婆 *gu1po4* 'the sister of one's paternal or maternal grandfather' (i.e. father or mother's paternal aunt); 姨婆 *ji4po4* 'the sister of the maternal grandmother' (i.e. mother's maternal aunt). There are also terms that are used by a wife to address the relatives of her husband: 姑奶奶 *gu1naai4naai2* and 舅老爺 *kau5lou5je4*.[18] The former is used to refer to the husband's paternal aunt, while the latter to the husband's maternal uncle. These kinship terms of grandparents' generation demonstrate the scale of the family of old Hong Kong.

Table 2 Kinship terms of core family members and their frequencies in HKCC

Father		Mother		Elder sister		Elder brother	
爸爸 *baa4baa1*	1551	(阿)媽 *aa3maa1*	1080	家姐 *gaa1ze1*	349	大哥 *daai6go1*	175
老豆 *lou5dau6*	122	媽媽 *maa4maa1*	142	(阿)姐 *(aa3)ze1*	52	(阿)哥 *aa3go1*	213
(阿)爹 *(aa3)de1*	95	媽咪 *maa1mi4*	83	大家姐 *daai6gaa1ze1*	33	哥哥 *go4go1*	35
(阿)爸 *(aa3)baa4*	19	老母 *lou5mou5*	48	姐姐 *ze4ze1*	25	二哥 *ji6go1*	17
家父 *gaa1fu6*	10	母親 *mou5can1*	8	大姊 *daai6zi2*	2	大佬 *daai6lou2*	15
父親 *fu5can1*	7	家母 *gaa1mou5*	7	姊姊 *zi2zi2*	1		

		Younger sister		Younger brother	
		(阿)妹 *aa3mui2*	92	細佬 *sai3lou2*	53
		妹妹 *mui6mui2*	24		
		三妹 *saam1mui2*	19		
		二妹 *ji6mui2*	5		
		細妹 *sai3mui2*	2		

---

17 The tree only provides the general meaning of the kinship morphemes. Some of these morphemes can have more than one meaning depending on the kinship terms they form. For example, the morpheme 公 *gung1* is usually understood as 'maternal grandfather', as in the kinship term 公公 *gung1gung1* or 外公 *ngoi6gung1*. However, 公 *gung1* can also appear in the term 老公 *lou5gung1*, meaning 'husband'.

18 The above five kinship terms 叔公 *suk1gung1*, 姑婆 *gu1po4*, 姨婆 *ji4po4*, 姑奶奶 *gu1naai4naai2* and 舅老爺 *kau5lou5je4* appear 2 times, 4 times, 3 times, 10 times, and 4 times respectively in HKCC.

In addition, there are a few items referring to members of extended families in the grandparents' generation: 叔公 *suk1gung1* 'younger brother of paternal grandfather' (i.e. father's paternal uncle); 姑婆 *gu1po4* 'sister of one's paternal or maternal grandfather' (i.e. father or mother's paternal aunt); 姨婆 *ji4po4* 'sister of maternal grandmother' (i.e. mother's maternal aunt). There are also terms that are used by a wife to address the relatives of her husband: 姑奶奶 *gu1naai4naai2* and 舅老爺 *kau5lou5je4*.[19] The former is used to refer to the husband's paternal aunt while the latter the husband's maternal uncle. These kinship terms of grandparents' generation demonstrate the scale of the family of old Hong Kong.

5.3 Other Terms of Address for Family Members

It is common to have more than one item addressing the same person, as shown in table 2 above. Sometimes, the choice among the different items depends on extra-linguistic factors such as solidarity and politeness (Wardhaugh 1992; Gu 1990). Some of these terms are used to show the respect of the addresser towards the addressee, and these terms are usually called honorific terms. In HKCC, there are a number of honorific terms referring to the core family members of the addressee. These honorific forms carry the prefix 令 *ling6*. Interestingly, the kinship terms following the prefix are not the same as the common forms.[20] Table 3 lists the honorific terms and their frequencies in HKCC.

Table 3 Honorific terms in HKCC

Meaning	Honorific term	Frequency in HKCC
your father	令尊翁 *ling6zyun1jung1*	6
	令尊 *ling6zyun1*	15
your mother	令壽堂 *ling6sau6tong2*	6
	令堂 *ling6tong2*	1
your brother	令兄 *ling6hing1*	1
your sister	令妹 *ling6mui2*	1
your son	令郎 *ling6long2*	7
your daughter	令千金 *ling6cin1gam1*	14

---

19   The above five kinship terms, 叔公 *suk1gung1*, 姑婆 *gu1po4*, 姨婆 *ji4po4*, 姑奶奶 *gu1naai4naai2*, and 舅老爺 *kau5lou5je4*, appear 2 times, 4 times, 3 times, 10 times, and 4 times respectively in HKCC.

20   For example, the honorific form for 'your father' is 令尊 *ling6zyun1* or 令尊翁 *ling6zyun1jung1*, but not 令爸 *ling6baa4*.

These terms are seldom used in modern Cantonese, and only in some very traditional settings.[21]

Another feature of the family structure of mid-20th century Hong Kong society is polygamy. It was quite common for men to take more than one wife, especially when the first wife could not bring any children to the family. There are several terms found in HKCC addressing the concubine or second wife of a man, and the stepmothers.

Table 4 Terms for concubines and stepmothers

Term	Meaning	Frequency
妾侍 *cip3si6*	Concubine	16
細姐 *sai3ze2*	Vocative for father's concubine	51
細婆 *sai3po4*	Vocative for grandfather's concubine	4
後底嫲 *hau6dai2naa2*	Stepmother	2
填房 *tin4fong4*	Stepmother	10

The practice of polygamy ended in 1971 as a result of the changes in the marriage law (Liu 1999; Sullivan 2005; Ip 2014). Therefore, we can see that terms addressing second wives and stepmothers were still quite common in mid-20th century movies.

Many families keep house workers, generally known as servants or maids. As Watson stated, maids were "purchased" (1991, 240), suggesting that the masters were usually wealthy and in the higher socioeconomic class. As for those maids who were bought to the family when they were very young, they were referred to as 妹仔 *mui1zai2* 'little maid'. There were also some servants who helped the mistresses of the family to take care of the children in activities such as breast-feeding. They were called 奶媽 *naai5maa1* 'wet nurse'. Below are some dialogues containing these terms. In dialogue (11), we can see that maids and servants were usually badly treated by the master and his family members.

11. *A Ready Lover* (十月芥菜, 1952)
    阿爸爸呀, 你唔好因佢係妹仔睇低佢喎!
    *aa3 baa4baa1 aa3, nei5 m4hou2 jan1 keoi5 hai6 **mui1zai2** tai2dai1 keoi5 wo3*
    'Daddy, you should not look down on her just because she is a little maid'.

---

12. *The Joyful Matrimony* (龍鳳合歡花, 1960)
    邊個養大個女嚟做妹仔服侍佢呀, 吓?
    bin1go3 joeng5 daai6 go3 neoi5 lai4 zou6 **mui1zai2** *fuk6si6 keoi5 aa3, haa2*
    'Who is willing to raise a daughter to be a little maid to serve him?'

13. *Midnight Werewolf* (夜半人狼, 1963)
    㗎, 肚餓叫奶媽攞嘢食啦!
    je1, tou5ngo6 giu3 **naai5maa1** *lo5 je5 sik6 laa1*
    'Yeah, if you are hungry, ask wet-nurse for food'.

14. *The Millionaire's Daughter* (千金之女, 1963)
    TOMMY, 奶媽頭先話佢唔精神。
    Tommy, **naai5maa1** *tau4sin1 waa6 keoi5 m4 zing1san4*
    'Tommy, wet nurse just said she did not feel well'.

6 Concluding Remarks

In this paper, we made use of the data from *The Corpus of Mid-20th Century Hong Kong Cantonese* to examine how Hong Kong society looked like half a century ago. Our focus was on kinship terms and terms related to marriage. Through these terms, we were able to see the family structure of the old Hong Kong, which was significantly different from contemporary Hong Kong. This could be due to changes in the concept of family and also in the lifestyle, such as working habits. Since the 1970s, Hong Kong people were strongly advised to have serious family planning and many families had only one or two children; this subsequently reduced the size of families.[22] There were no more 'big families' (大家族 *daai6gaa1zuk6*), which led to the reduced use of many kinship terms.[23]

This paper also demonstrates how HKCC can be used to conduct corpus-based sociolinguistic studies in Cantonese which had not been extensively and systematically explored. The corpus data is highly relevant in terms of time (i.e. mid-20th century) and nature (movies with their themes on daily life situations). It is hoped that more corpus-based sociolinguistic studies can be carried out in future with the development of more Cantonese corpora covering a broader variety of language data.

---

22 Wong discussed how the family planning campaign of Hong Kong in the 1970s challenged "traditional Chinese values in the areas of family size and gender dominance [...] that reshaped society in Hong Kong" (2018, 123).

23 There are some kinship terms showing the traditional big family structure. For example, 舅父仔 *kau5fu2zai2* 'little maternal uncle' is used to refer to the maternal uncle whose age is close or even smaller than the addresser. Other terms include 七妹 *cat1mui2* 'the seventh sister' and 四姨 *sei3ji1* 'the fourth maternal aunt'.

Bibliography

Bacon-Shone, J.; Bolton, K.; Luke, K.K. (2015). *Language Use, Proficiency and Attitudes in Hong Kong*. Hong Kong: Social Sciences Research Centre; University of Hong Kong.

Baker, H.D.R. (1979). *Chinese Family and Kinship*. New York: Columbia University Press.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Chao, Y.R. (1956). "Chinese Terms of Address". *Language*, 32(1), 217-41. `https://doi.org/10.2307/410666`.

Cheung, S.H. (1990). "Terms of Address in Cantonese". *Journal of Chinese Linguistics*, 18(1), 1-43. `https://www.jstor.org/stable/23767129`.

Cheung, S.H. (2001). "The Interrogative Construction. (Re)constructing Early Cantonese Grammar". Chappell, H. (ed.), *Sinitic Grammar. Synchronic and Diachronic Perspectives*. Oxford: Oxford University Press, 191-231.

Chin, A.C. (2011). "Grammaticalization of the Cantonese Double Object Verb [pei35] 畀 in Typological and Areal Perspectives". *Language and Linguistics*, 12(3), 529-63. `http://www.ling.sinica.edu.tw/files/publication/j2011_3_02_0726.pdf`.

Chin A.C. 錢志安 (2013). "Yueyu yanjiu xin ziyuan. Xianggang ershi shiji zhongqi yueyu yuliaoku" 粵語研究新資源——《香港二十世紀中期粵語語料庫》 (New Resources for Cantonese Language Studies. A Linguistic Corpus of Mid-20th Century Hong Kong Cantonese). *Zhongguo yuwen tongxun*, 92(1), 7-16.

Chin, A.C. (2018a). "Discourse Markers in Cantonese". Paper presented at *The 30th North American Conference on Chinese Linguistics* (Columbus, Ohio, 9-11 March 2018). Ohio State University.

Chin, A.C. (2018b). "唔好客氣 vs. 咪走寶. A Corpus-Based Study of Cantonese Prohibitive Markers". Paper presented at *The 18th Workshop on Cantonese* (Hong Kong, 21 April 2018). The Chinese University of Hong Kong.

Chin, A.C. (2019a). "Initiatives of Digital Humanities in Cantonese Studies. A Corpus of Mid-Twentieth-Century Hong Kong Cantonese". Tso, W.B.A. (ed.), *Digital Humanities and New Ways of Teaching*. Singapore: Springer, 71-88.

Chin A.C. 錢志安 (2019b). "Yanbian zhong de yuyan. Yi yueyu zhongxing wenju weili" 演變中的語言——以粵語中性問句為例 (Linguistic Change in Progress. A Case Study of Cantonese Neutral Questions). Paper presented at 海外珍藏漢語文獻與漢語研究高端論壇 *Symposium of Chinese Texts from Overseas and Chinese Linguistic Studies* (Guangzhou, 29 October 2019). San Yat-sen University.

Chin A.C. 錢志安 (forthcoming). "Hanyu fangyan yuliaoku de jiangou he yingyong. Yi Xianggang ershi shiji zhongqi yueyu yuliaoku weili" 漢語方言語料庫的建構和應用——以《香港二十世紀中期粵語語料庫》為例 (The Construction and Application of the Corpus of Chinese Dialects. A Case of the Corpus of Mid-20th Century Hong Kong Cantonese). *Hanyu yuyanxue jikan*.

Chung P.Y. 鍾寶賢 (2004). *Xianggang yingye bai nian* 香港影業百年 (The Movie and Television Industry of Hong Kong over the Past Hundred Years). Hong Kong: Joint Publishing Limited.

CSD. Census and Statistics Department (2016). *Summary Results of 2016 Population By-census*. `https://www.bycensus2016.gov.hk`.

Davies, M. (2017). "Using Large Online Corpora to Examine Lexical, Semantic, and Cultural Variation in Different Dialects and Time Periods". Friginal, E. (ed.), *Studies in Corpus-Based Sociolinguistics*. London: Routledge, 19-82.

Gu, Y. (1990). "Politeness Phenomena in Modern Chinese". *Journal of Pragmatics*, 14, 237-57. https://doi.org/10.1016/0378-2166(90)90082-o.

Ip, K.Y. (2014). *The Abolishment of Concubinage in Hong Kong. An Analysis of Its Process and Opinions on the Issue* (*1948-1971*) [PhD Dissertation]. Hong Kong: The Chinese University of Hong Kong.

Lai Y.P. 黎奕葆; Chin A.C. 錢志安. (2018). "Yueyu de dongci houzhui 'zhe'" 粵語的動詞後綴'着' (The Cantonese Verbal Suffix *zoek6*). Ho, D.-A. et al. (eds), *Hanyu yu Hanzangyu qianyan yanjiu. Ding Bangxin xiansheng badie shouqing lunwenji*《漢語與漢藏語前沿研究——丁邦新先生八秩壽慶論文集》 (Frontiers in Sinitic and Sino-Tibetan Linguistics: Studies in the Languages of China: Festschrift in Honour of Professor Ting Pang-Hsin on His 80th Birthday). Beijing: Social Sciences Academic Press, 697-710.

Law, K. (1986). "Archetype and Variations". *Cantonese Melodrama 1950-1969*. Hong Kong: The Urban Council, 10-20.

Leung, M.; Law, S.; Fung, S. (2004). "Type and Token Frequencies of Phonological Units in Hong Kong Cantonese". *Behavior Research Methods, Instruments, & Computers*, 36(3), 500-5. https://doi.org/10.3758/bf03195596.

Liu, A.N.C. (1999). *Family Law for the Hong Kong SAR*. Hong Kong: Hong Kong University Press.

Lui, T.L. (1988). "Home at Hongkong". *Changes in Hong Kong Society through Cinema*. Hong Kong: The Urban Council of Hong Kong, 83-92.

Luke, K.K.; Lau, C. (2008). "On Loanword Truncation in Cantonese". *Journal of East Asian Linguistics*, 17, 347-62. https://doi.org/10.1007/s10831-008-9032-x.

Luke, K.K.; Wong, M.L.Y. (2015). "The Hong Kong Cantonese Corpus. Design and Uses", in Tsou, B.; Kwong, O.O. (eds), "Linguistic Corpus and Corpus Linguistics in the Chinese Context", monogr. no., *Journal of Chinese Linguistics*, 25, 312-33. https://www.jstor.org/stable/26455290.

McCoy, J. (1970). "Chinese Kin Terms of Reference and Address". Freedman, M. (ed.), *Family and Kinship in Chinese Society*. Stanford: Stanford University Press, 209-26.

McEnery, T.; Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.

Peng, G. (2006). "Temporal and Tonal Aspects of Chinese Syllables. A Corpus-Based Comparative Study of Mandarin and Cantonese". *Journal of Chinese Linguistics*, 34(1), 134-54. https://www.jstor.org/stable/23754151.

Podmore, D.; Chaney, D. (1974). "Family Norms in a Rapidly Industrializing Society. Hong Kong". *Journal of Marriage and Family*, 36(2), 400-7. https://doi.org/10.2307/351167.

Qian, Y.; Piao, S. (2009). "The Development of a Semantic Annotation Scheme for Chinese Kinship". *Corpora*, 4(2), 189-208. https://doi.org/10.3366/e1749503209000306.

Sullivan, P.L. (2005). "Culture, Divorce, and Family Mediation in Hong Kong". *Family Court Review*, 43(1), 109-23. https://doi.org/10.1111/j.1744-1617.2005.00011.x.

Tse C.M. 謝明燊; Chin, A.C. 錢志安 (2015). "Yueyu 'ming-liang-ming' jiegou de tongzhi yongfa" 粵語「名-量-名」結構的同指用法 (The Co-Referential Usage of "Noun-Classifier-Noun" in Cantonese). Paper presented at *The Fifteenth*

LSHK Workshop on Cantonese (Hong Kong, 11 April 2015). The University of Hong Kong. `https://www.jstor.org/stable/23756692`.

Tsou B. 鄒嘉彥 (1997). "San yan, liang yu shuo Xianggang" 三言兩語說香港 (Three Spoken Languages and Two Written Languages in Hong Kong). *Journal of Chinese Linguistics*, 25(2), 290-307.

Wardhaugh, R. (1992). *An Introduction to Sociolinguistics*. Oxford: Blackwell.

Watson, R. (1991). "Wives, Concubines, and Maids. Servitude and Kinship in the Hong Kong Region, 1900-1940". Watson, R.S.; Ebrey, P.B. (eds), *Marriage and Inequality in Chinese Society*. Berkeley; Los Angeles: University of California Press, 231-55.

Wong, M.L.Y. (2009). "*Gei* Constructions in Mandarin Chinese and *Bei* Constructions in Cantonese. A Corpus-Driven Contrastive Study". *International Journal of Corpus Linguistics*, 14(1), 60-80. `https://doi.org/10.1075/ijcl.14.1.04won`.

Wong, W.S. (2018). "Reconfiguring a New Tradition of Ideal Family Size. A Case Study of the Family Planning Association of Hong Kong, 1977-1982". Wong, W.S. (ed.), *The Disappearance of Hong Kong in Comics. Advertising and Graphic Design*. Cham: Palgrave Macmillan, 123-40. `https://doi.org/10.1007/978-3-319-92096-2_6`.

Wu, C.-C. (1927). "The Chinese Family. Organization, Names, and Kinship Terms". *American Anthropologist*, 29(3), 316-25. `https://doi.org/10.1525/aa.1927.29.3.02a00100`.