

# Compiling *The Diachronic Corpus of Hong Kong English* (DC-HKE): motivation, progress and challenges

Carolin Biewer (University of Zurich), Tobias Bernaisch, Mike Berger (University of Giessen), Benedikt Heller (KU Leuven)

## 1. Introduction: the colonial history of Hong Kong

Before 1842 Guangzhou (Canton) as “a centre for the tea trade” draws British merchants to this part of the world (Setter et al. 2010: 4)  
 1842 Hong Kong Island comes under British rule, then Kowloon Peninsula (1860) and Lantau Island (1898)  
 30 June 1997 THE HANDOVER: Hong Kong is returned to China

The handover marks the beginning of socio-political changes in Hong Kong that also affect the status and function of English in Hong Kong (for the effect of new language policies see for instance Bolton (2012)).

## 2. Current research on English in Hong Kong

Hong Kong English shows remarkable structural differences from Standard British and American English on all levels of language organisation (see e.g. Bolton 2002a, Bolton 2003, Setter et al. 2010), for instance (example sentences from Setter et al. 2010):

- borrowings from Cantonese and Mandarin Chinese (*taipan*, *dim-sum*, *tai chi* ...)
- subjectless sentences (*My father can only take care of two of us, cannot like go with many young kids.*)
- different use of articles (*You're doing receptionist job.*)
- the use of *will* as past habitual marker (*we will go for we would go*)
- subject-verb disagreement (... *the church organise* ...)
- omission of prepositions (*We must care the old people.*)
- Hong Kong Chinese use *actually* more frequently than non-Hong Kongers and focus on different pragmatic functions of *actually* than non-Hong Kongers (Cheng & Warren 2001)

Its overall sociolinguistic developmental status, however, remains largely unresolved to date, as does the possibility of a political influence on the structure of Hong Kong English

### Open questions

- Schneider (2007: 137) claims that Hong Kong English has progressed from the phase of exonymic stabilisation into the phase of nativization. What hard evidence do we have for that?
- With the socio-political changes since 1997 are we now experiencing a transition period in Hong Kong in which English is developing from a second-language variety to a foreign language? (for opposite views see e.g. Bolton (2002b) versus Görlach (2002))
- How have the political changes in Hong Kong influenced the structure and use of English? Can we see changes in the structure of Hong Kong English beyond the introduction of abbreviations like SAR (Special Administrative Region) in political discourse?

We need diachronic data and new sociolinguistic surveys!

## 3. The project

### 3.1. The idea

- compilation of *The Diachronic Corpus of Hong Kong English*
- to be representative of English in Hong Kong of the 1930s/1960s/1990s/the period of 2009–14
- to be comparable to subcorpora of the BLOB/LOB/FLOB/CLOB family (British English of the 1930s/60s/90s and 2009)
- to be comparable to subcorpora of *The Diachronic Corpus of Singapore English* (project Prof. Sebastian Hoffmann and Prof. Andrea Sand, University of Trier)

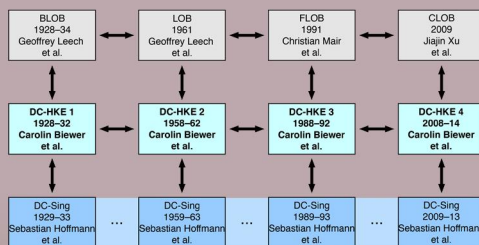


Figure 1: Comparability of the DC-HKE to other corpora

Table 1: Text categories of DC-HKE (matching the LOB family)

	1928–32	1958–62	1988–92	2008–14
A Press: reportage	100,000	100,000	100,000	100,000
B Press: editorials	50,000	50,000	50,000	50,000
C Press: reviews	30,000	30,000	30,000	30,000
E Non-academic writing	70,000	70,000	70,000	70,000
J Academic writing	150,000	150,000	150,000	150,000
K Fiction	100,000	100,000	100,000	100,000
<b>Total</b>	<b>500,000</b>	<b>500,000</b>	<b>500,000</b>	<b>500,000</b>

### 3.2. Cooperation

The DC-HKE will be compiled in cooperation with Prof. Dr. Kingsley Bolton, *Nanyang Technological University*, Singapore and Prof. Dr. John Bacon-Shone, *University of Hong Kong*

### 3.3. Data for the DC-HKE collected so far

- A–C: Press Press news articles from *The South China Morning Post* (SCMP) from 2011 and 2013 (3 million words)
- J: Academic writing over 900 PhD and MA Theses from *The City University of Hong Kong* in PDF format, partly OCREd (67 from the 1960s, 247 from the 1990s and 625 from 2008–2014)

### 3.4. Methodological challenges

- The **availability of suitable data**: it will be difficult to get enough data for all selected text categories the further we go back in time (in particular, no real tradition of English literature in Hong Kong).
- The **quality of the data**, e.g. for newspapers: often the authors of the articles are unknown; in earlier editions of the SCMP no distinctions are made between columns as they are made in later editions => how to assess text categories?
- Speaker selection**: in earlier newspaper editions many authors are Indians or expatriate journalists, in later newspaper editions authors are mostly Chinese; a focus on Hong Kong Chinese is not enough to prove or disprove Schneider's claim, but don't we at least need to know the ethnic background of the author to test his hypothesis?
- The **data processing**: the data are available in different formats (PDF, printed on paper, on microfiche); we have to find a way to digitize the data that is not too expensive and time-consuming.

## 4. Possible areas of research to be pursued with the DC-HKE

### 4.1. Interesting structural features in Hong Kong English today ...

... the development of which would be interesting to investigate with the DC-HKE (results based on 3 million words of press news articles from the SCMP from 2011 and 2013)

#### 4.1.1. Verb complementation: LOVE to do something vs. LOVE doing something

- Cheng, 25, who loves singing karaoke, was encouraged by his friends to enter TVB's competition, *The Voice*.  
 [DC-HKE4-NP-Wuhsu\_star\_finds\_voice]
- Sheung-choi loves to watch Ronaldo, ...  
 [DC-HKE4-NP-The\_next\_Ronaldo\_-\_from\_Kwal\_Tsing']

Table 2: LOVE Ving vs. LOVE to V

	LOVE Ving	LOVE to V	Total	results $\chi^2$ :
DC-HKE press	47.52 % (n = 48)	52.48 % (n = 53)	100.00 % (N = 101)	HKE vs. BrE: p < 0.01
SAVE-IND	33.94 % (n = 37)	66.06 % (n = 72)	100.00 % (N = 109)	HKE vs. IndE: p > 0.01
BNC news	29.13 % (n = 74)	70.87 % (n = 180)	100.00 % (N = 254)	BrE vs. IndE: p > 0.01

(The results for IndE and BrE are reproduced from Bernaisch (2013). The SAVE-IND data stem from the Indian component of the *South Asian Varieties of English* (SAVE) Corpus)

The quantitative findings ( $\chi^2 \approx 10.86$ ,  $df = 2$ ,  $p < 0.01$ ) show that *-ing* complements occur significantly more often in the HKE than in the BrE data. This may be an indication that LOVE in HKE is further advanced in terms of the Great Complement Shift (cf. Vosberg 2006: 17) towards *-ing* complementation than LOVE in BrE.

#### 4.1.2. Verb complementation: PROVIDE so. with sth. vs. PROVIDE sth. to so.

- The Hong Kong Observatory provides us with weather forecasts and ...  
 [DC-HKE4-NP-Hong\_Kong\_Observatory\_needs\_calm\_heads\_in\_stormy\_weather]
- ... the service provides food to people in need ...  
 [DC-HKE4-NP-Keeping\_them\_off\_the\_streets]

	PROVIDE Oi with Od	PROVIDE Od to Oi	Total	Table 3: PROVIDE with vs. PROVIDE to
DC-HKE press	38.76 % (n = 50)	61.24 % (n = 79)	100.00 % (N = 129)	
BNC news	64.98 % (n = 269)	35.02 % (n = 145)	100.00 % (N = 414)	

There are statistically significant differences ( $\chi^2 \approx 27.89$ ,  $df = 1$ ,  $p < 0.01$ ) in the complementation of PROVIDE between the HKE data and the BrE data. In HKE press news articles the construction PROVIDE Od to Oi is preferred to the construction PROVIDE Oi with Od, whereas in the BNC news section the latter is preferred. According to Schneider (2007: 170) this preference is also found in Indian English, and it is possible that it also applies to similar verbs such as SUPPLY and FURNISH.

#### 4.1.3. Distribution of lexical bundles: spoken vs. written patterns



- Examples: *for the first time* (Preposition + NP fragment), *Hong Kong is a* (Pronoun/NP + copula BE)
- The first category (Preposition + NP fragment) is typical of academic language, the second is typical of a conversational style (cf. Biber et al. 1999: Chapter 13)
- In terms of lexical bundles, the SCMP seems more formal than the Guardian.

#### 4.2. How politics may influence current language use in the news

Different expressions of an author's stance/different strategies of text composition in the SCMP when commenting on the 16th anniversary of the handover on 1 July 2013:

**Example 1: "Protest in Hong Kong shows time for dialogue on vote" [DC-HKE4-NP-Protest\_in\_HK]; author is pro democracy**

- Implicit/ambiguous attribution of stance
  - no use of the first personal pronoun, instead:
  - passive constructions: *it is to be hoped that ...*
  - constructions with *it* as an empty subject: *it will be in the interest of Beijing; it remains unclear; it will be a significant step forward ...*

- Evaluative word choice  
 The author talks about the *city's thirst for democracy* that should be greeted with an *... accommodating response from Beijing*

(c) The name of the author is withheld

**Example 2: "Democracy is not a panacea" [DC-HKE4-NP-Democracy\_not\_panacea]; author is contra full democracy**

- Explicit attribution of stance  
 Use of the first person pronoun: *As much as I support their goal, I don't share their optimism; I will address the widespread belief ...*

- Evaluative word choice  
 The author talks about a *zest for full democracy* as an overflowing excitement that clouds people's judgment and a *belief that democracy will soften deep social divisions in Hong Kong*, not a truth

- The name and the email address of the author are given  
 With the DC-HKE it will be possible to find out whether there are statistically significant differences in terms of text composition and the attribution of stance before/after 1997.

## 5. Next steps

- a trip to Hong Kong in June 2014
- an application for funding in September
- pilot studies on structural features

## Bibliography

- Bernaisch, Tobias. 2013. *Lexis and Lexicogrammar of Sri Lankan English: A Corpus-Based Study of the Nativization of a Link Language*. Unpublished PhD Thesis, Giessen: University of Giessen.
- Biber, Douglas, Sif, Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Longman.
- Bolton, Kingsley. 2002a. *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press.
- Bolton, Kingsley. 2002b. *The sociolinguistics of Hong Kong and the space for Hong Kong English*. In: Kingsley Bolton (ed.), *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 29–55.
- Bolton, Kingsley. 2003. *22nd Annual English Language Conference*. CUP.
- Bolton, Kingsley. 2012. *Language policy and planning in Hong Kong: The historical context and current realities*. In: Eui-Ling Low & Azhar Ibrahim (eds.), *English in Southeast Asia: Features, Policy and Language in Use*. Amsterdam: John Benjamins, 221–238.
- Cheng, Winnie and Warren, 2001. *The historical context and current realities*. In: *International Journal of Corpus Linguistics* 6(2), 257–280.
- Görlach, Manfred. 2002. *English in Singapore, Malaysia, Hong Kong, Indonesia, the Philippines ... a second or a foreign language?*. In: Manfred Görlach (ed.), *South Asian Varieties of English*. Amsterdam: John Benjamins, 98–117.
- Hoffmann, Sebastian. 2009. *Diachronic English: Varieties Around the World*. Cambridge: CUP.
- Schneider, Edgar W., 2007. *World Englishes*. Edinburgh: Edinburgh University Press.
- Setter, John, Cathy S. Setzler & ... (eds.). 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.
- Vosberg, Ulrike. 2006. *English as a Lingua Franca: Evidence and Implications for English as a Second Language*. Tübingen: Narr.

