

**Lexical Bundles in
Conversation across
English Varieties:
A Core-periphery
Approach**

Zeping Huang

Hong Kong Baptist University

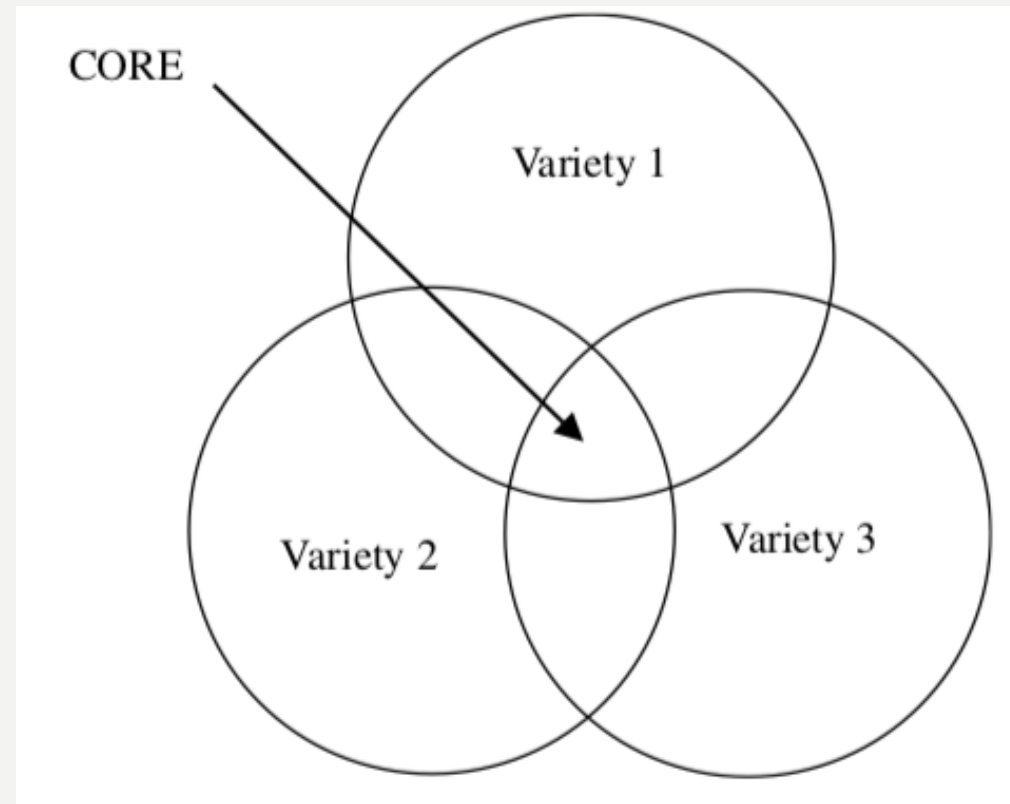
huangzeping@hkbu.edu.hk

1. METHODOLOGICAL FRAMEWORK

- English is widely used today as an international language.
- A **common core** must underline both speech and writing for all varieties of English to achieve intelligibility in communication (Quirk, et. al, 1985).

1. METHODOLOGICAL FRAMEWORK

- Nelson (2006, 2014) defines the core and periphery of world Englishes by depicting each variety of English as a circle. The **core** is the area where all the circles **overlap** with each other, while the **periphery** is any area where no overlap occurs.



1. INTRODUCTION

- Nelson (2006, 2014) explores the overlap and non-overlap of single word units among ten varieties of English.
- His two studies found that at the **morphological level**, the **core** is relatively small, but its frequency of use is greater than that of the **periphery**.
- However, the nature of the core and periphery at other linguistic levels, such as the **lexico-grammatical level** remains unexplored.

1. INTRODUCTION

Schneider (2007: 249):

- “[...the] indigenisation of language structure occurs mostly on a **lexico-grammatical level.**”
- “When words co-occur increasingly frequently, locally characteristic collocations and **‘lexical bundles’** (Biber et al. 1999:1036) will emerge, and in the long run this may result in the development of fixed expressions or idioms”.

1. INTRODUCTION

What is lexical bundle ? (Biber 2009: 283)

- recurring sequences of **three or more** words
- bridge two **structural units**
- at one end is **a clause** or **phrase** boundary; at the other end is the **beginning element** of a **second grammatical** structure

e.g., I want to know, well that's what I

- discourse building blocks

1. INTRODUCTION

Hypothesis:

- English speakers of different varieties rely on a common **core** lexical bundles to construct meaning in communication, but at the same time, **variations** may **coexist**.

Purpose:

- The present study, **adopting the “core and periphery” approach**, aims to compare the use of **lexical bundles** in one specific genre – **conversation** – among four varieties of English:

British English (BrE), Canadian English (CanE), Singapore English (SgE), and Hong Kong English (HKE).

2. RESEARCH QUESTIONS

1. What is the **distribution** of the core and periphery bundles in conversation across the four varieties?
2. Can the core and periphery bundles reveal any **(dis)similarity** in linguistic patterns? If so, what are the **similarities and the differences?**

3. METHODOLOGY

Corpus

- The International Corpus of English (ICE) (Greenbaum 1996)
- 90 texts of face-to-face conversation;
- 2000-word per text

	ICE-GB	ICE-CA	ICE-SIN	ICE-HK
Words	182,730	191,797	210,972	184,603
Total	=770,108			

3. METHODOLOGY

Identifying lexical bundles

Similar to the previous research (e.g., Cortes 2004; Pan, Reppen and Biber 2016)

- **Frequency:** 25 per million words (raw freq. >5)
- **Dispersion:** occur in > 3 texts
- **Length:** 3-words

But excluding:

- **Acronyms:** e.g., *BBC, PhD, or NUS*
- **proper nouns:** e.g., *Hong Kong*
- **Sound fillers:** *um, uh* and *mmm*

They are content-based lexis rather than discourse building devices; while the sound fillers are not lexis but rather filled pauses.

5. RESULTS

5.1 DISTRIBUTION OF LB

Table 1. Percentage of 3-word bundles in each corpus

	Type of bundles	Frequency of bundles	Total tokens of each corpus	% of tokens of bundles in each corpus
ICE-GB	2,234	24,825	182,730	13.6%
ICE-CAN	2,393	27,147	191,797	14.2%
ICE-SIN	2,044	22,196	184,609	12.0%
ICE-HK	2,062	24,699	210,972	11.7%

Table 2. Top 20 bundles from each “overlap” list

4 overlap absolute core	3 overlap	2 overlap	Non-overlap absolute periphery
I don't know	yeah yeah yeah	ya ya ya	a sort of
it's a	yeah it's	ya it's	then after that
a lot of	and all that	ya that's	lah you know
it's not	yeah that's	you know ya	it's gonna
I think it	at the moment	've got to	the sort of
you know I	going to have	ya ya I	you're gonna
I don't think	and he's	ya I think	don't know lah
I'm not	yes it's	well that's	lah I mean
think it's	yeah I think	ya you know	have got to
I mean I	and I was	ya but I	yes I mean
no no no	a bit of	I'm gonna	this one is
I think I	he's got	yes I think	's all right
it's it	's what I	you see so	we're gonna
you have to	this kind of	's got a	it is like
's it's	there was a	well I'm	oh you mean
but it's	was going to	ya and then	I was gonna
that's right	most of the	going to go	mean I think
it's very	it is a	I mean there	's right yes
going to be	there is a	ya I mean	ya well I
you want to	well I mean	Yeah but I	is it ya
Type = 454 (9%)	519 (10%)	1,053 (20%)	3,255 (61%)

5. RESULT

Table 3. Frequency of 3-word bundles in the “overlap” list

	4 overlap	3 overlap	2 overlap	Non-overlap	Totals
Total freq. of all bundles	42,167	16,669	17,945	22,027	98,808
% of overall frequency of all bundles	43%	16%	18%	22%	100%
Total bundles as a % of the entire dataset	6.4%	2.4%	2.7%	3.3%	
Average freq. of each bundle	93	32	15	7	

This result suggests that at the lexico-grammatical level, language speakers use a relatively small number of common bundles, but the core bundles are highly frequently used.

5. RESULT

5.2 VARIETAL COMPARISON

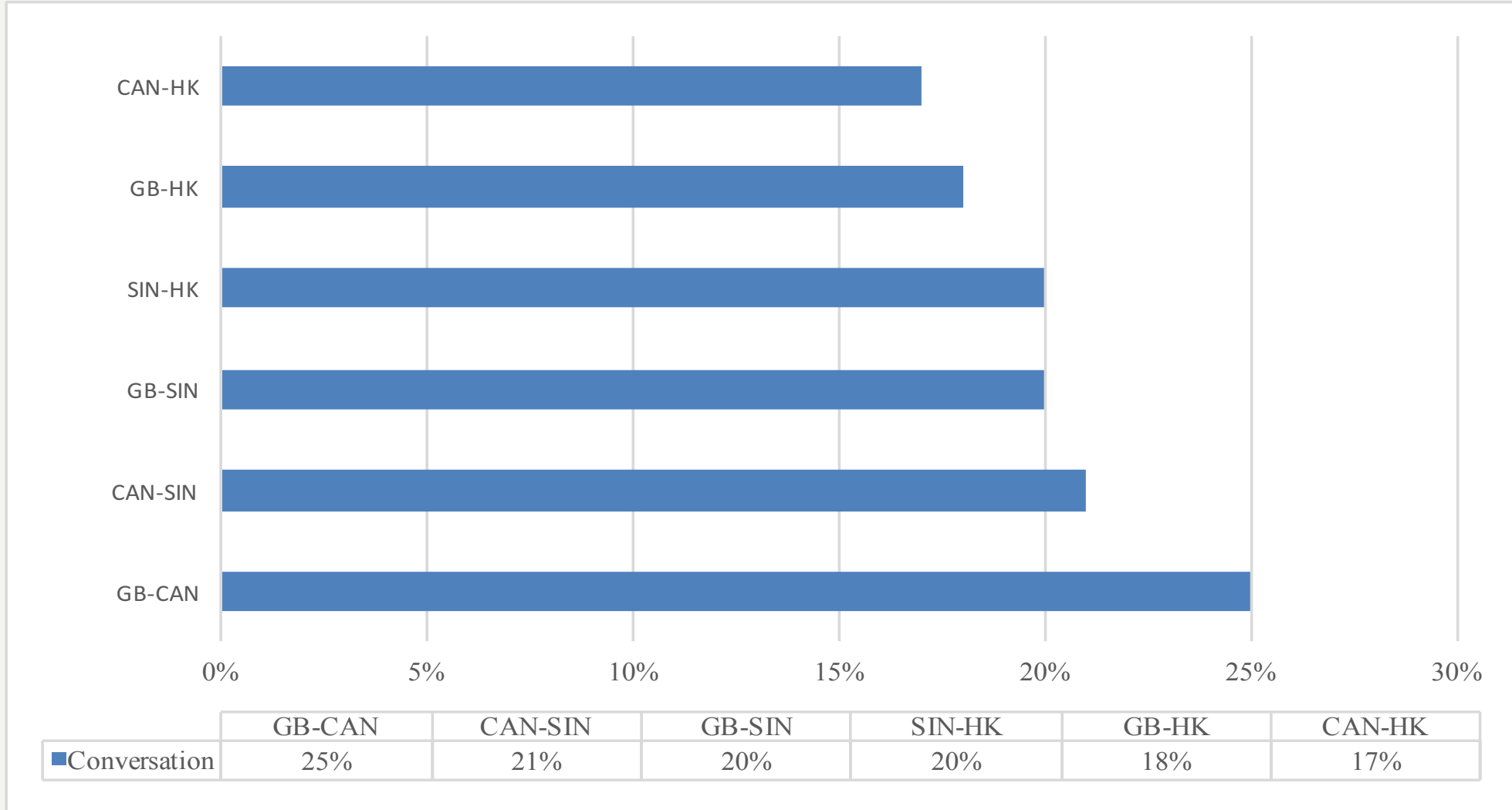


Figure 1. Percentage of overlap bundles in type between individual corpora

5. RESULT

Table 4. Spearman's rho correlations of the all overlap bundles (n=2,026) in **frequency** between the four varieties of English

	ICE-GB	ICE-CAN	ICE-SIN	ICE-HK
ICE-GB	1.00			
ICE-CAN	<u>.52</u> * ($p = .00$)	1.00		
ICE-SIN	.21* ($p = .00$)	.28* ($p = .00$)	1.00	
ICE-HK	.17* ($p = .00$)	<u>.08</u> ($p = .00$)	<u>.35</u> * ($p = .00$)	1.00

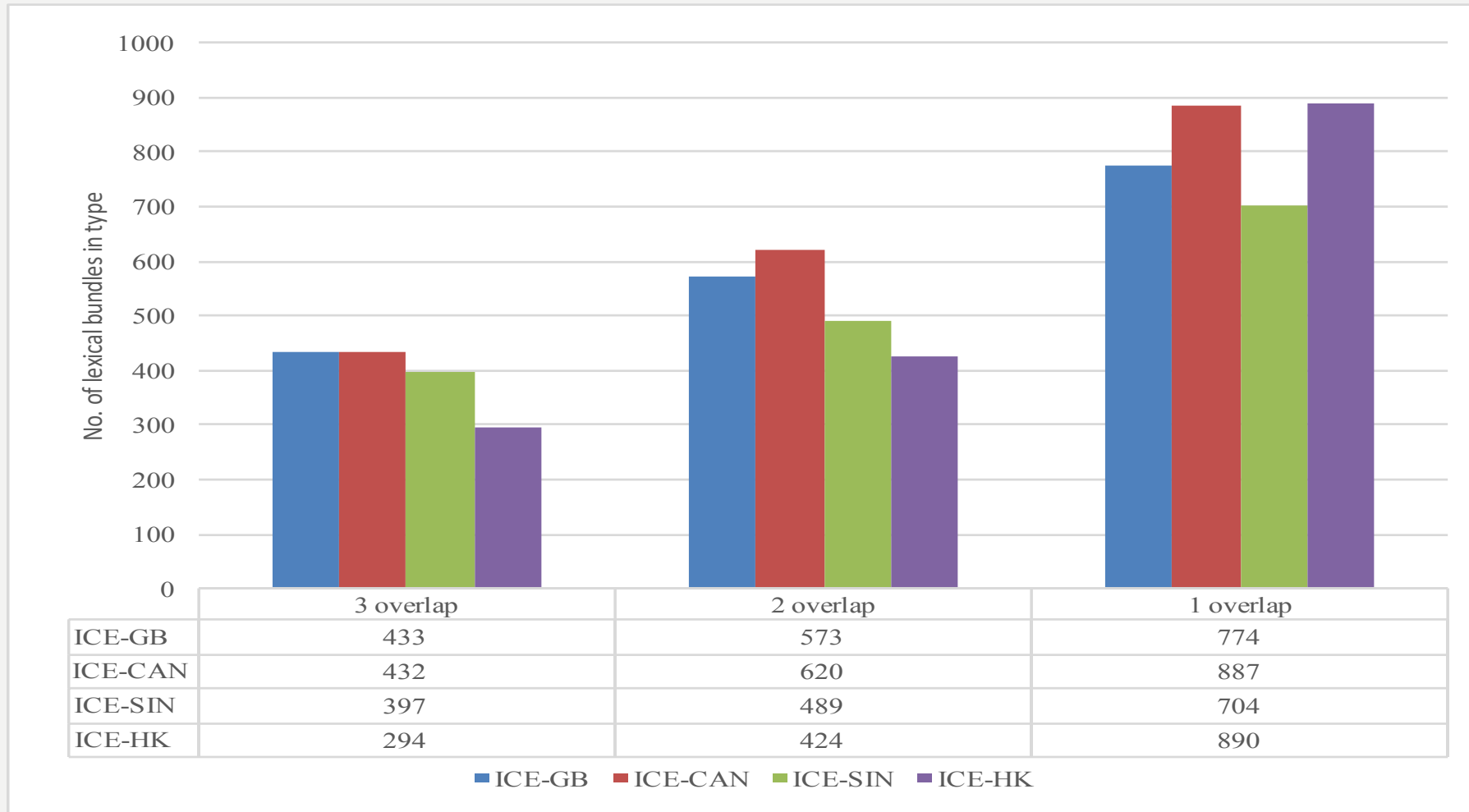
* Correlation is significant at the 0.01 level (2-tailed)

The Spearman result further proves that in terms of frequency,

- **ICE-GB** and **ICE-CAN** have the **strongest** correlation ($r = .52$)
- **ICE-HK** and **ICE-SIN** have the second strongest correlation ($r = .35$)
- **ICE-HK** has the **least** correlation with **ICE-CAN** ($r = .08$) and **ICE-GB**. ($r = .17$)

5. RESULTS

5.2 VARIETAL COMPARISON



OF the 3 overlaps

ICE-GB: 83.4%

ICE-CAN: 83.2%

ICE-SIN: 76.5%

ICE-HK: 56.6%

Figure 1. Distribution of bundles in type overlapping in 3 corpora, 2 corpora and 1 corpus (total no. of 3-overlap bundles =519; total no. of 2-overlap =1,053; total no. of non-overlap =3,255)

Table 5. Top 20 core bundles in the 4-overlap list based on LL

	ICE-GB	ICE-CAN	ICE-SIN	ICE-HK	
	Freq.	Freq.	Freq.	Freq.	Log-likelihood
going to be	101	133	37	9	173.71
i 've got	144	37	28	21	163.00
you 've got	105	30	16	9	140.13
i want to	34	24	61	147	110.10
think it 's	64	54	68	202	107.41
well it 's	77	44	7	16	91.90
it 's very	59	24	89	142	87.94
i think it	89	102	95	245	84.99
and you know	33	100	24	21	83.41
that 's why	26	37	140	67	80.90
've got a	57	15	8	5	77.76
that 's right	111	106	61	31	73.97
that 's what	60	89	38	15	70.91
it 's like	38	97	96	30	70.62
but i mean	59	49	10	11	70.55
know how to	5	7	16	60	70.30
that 's a	59	67	19	13	67.33
i have to	23	64	37	113	62.73
kind of thing	15	16	56	6	62.39
and it was	33	54	11	6	61.83
i mean it	99	60	44	25	61.39
but i think	36	32	56	120	60.30
's going to	69	50	36	11	57.94
yes that 's	49	7	6	21	56.44

454 core bundles that occur in all the four varieties

Table 5. Distinctive bundles in each overlap list (relative freq. per million words > 200; LL>50 for the 4 overlaps)

	ICE-GB	ICE-CAN	ICE-SIN	ICE-HK
4-overlap	going to be, I've got you've got that's right I mean it	going to be that's right it's like	that's why kind of thing it's like	I want to I think it but I think I have to
3-overlap	at the moment yeah that's a bit of yeah it's was going to yeah I mean going to have that sort of yes it's he's got sort of thing 'a bit; it was going; and he's	and I was and he's going to have it was just there was a it was like it was really I'd like that would be and she's 's what I; he's got	and all that it'll be it is not	yeah yeah yeah yeah it's yeah I think at that time most of the yeah that's this kind of yeah it's there's a do you like it is a; I go to; think it is
2-overlap	have you got well that's I mean there 's got a well I'm	ya it's ya that's you know ya I'm gonna ya you know well that's ya I think ya it was ya I know; 'd have to	ya ya ya ya it's ya that's ya you know ya I think ya but I	yes I think yeah but I because I have yeah yeah so it's difficult
non-overlap	a sort of that sort of	it's gonna	then after that lah you know	how to say there are many how about you

5. RESULT

Semi-modal bundles: *going to be; 's going to*

- This is in line with the findings of previous studies that quasi-modal *be going to* is thriving in speech in the IC varieties (see Collins 2005, 2009; Collins and Yao 2012). However, *be going to* still remains comparatively underused in HKE and SgE despite its rising popularity in the two IC varieties.

Response token bundle: *that's right, right that's*

- BrE and CanE speakers prefer to use these formulaic chunks in conversation apart from employing other response devices starting with *yes or yeah*.
- Comparably, the low frequency of *that's right / right that's* in the two Asian English varieties indicates that HKE and SgE speakers do not employ the same response tokens to interact with interlocutors as frequently as the CanE and BrE speakers do. Instead, they tend to prefer other chunks such as repetitive bundles *yeah yeah yeah* in HKE or *ya ya ya* in SgE.

ICE-SIN:

Quotative *like* bundles (*it's like*)

- Undergone a process of lexicalization and tends to be processed as a fixed formulaic chunk by SgE speakers.

(1) <ICE-SIN: SIA-042#X295:I: B>

Ya I I think those days when I looked back at it **it's like** I I I **looked** at it like from you know

<ICE-SIN: SIA-042#X296:I:B> It's as though you know it's a nightmare

(2) <ICE-SIN: SIA-031#156:I:A> What happened

<ICE-SIN: SIA-031#157:I:C> **It's like** I I went to the Repro room Reprographic room you see

(3) <ICE-SIN: SIA-046#241-243: B> I just can't be but I didn't say things like that <.>

And then he said oh is it because of puppy love <.> **It's like** I was so

6. RESULTS

ICE-SIN:

Vague bundles *kind of thing* and *and all that*.

Periphery bundle: *then after that*

- a process of grammaticalisation and routinisation
- developed into a **temporal marker** in SgE.

e.g.

(1) <ICE-SIN: SIA-048#I36: I: B> Ya then later you worry about them girlfriend boyfriend<,> **then after that** their work<.>Throughout your life you're worrying<,>

(2) <ICE-SIN: SIA-059#93: I: B> Oh and then what <?> You went upstairs to the MA room

< ICE-SIN: SIA-059#95: I:A> no **then after that** I met Chris <.> I mean he happened to walk walk by Dr Goatly's room and we went for lunch

5. RESULT

ICE-HK:

Prefer **epistemic bundles** containing *I think*, **semi-modals** *have to*, *want to*, **and causative connector** *because*.

Periphery bundle **how to say**

Used as **a filler** to help the speakers gain more time for language processing when they encounter difficulties in formulating online thoughts.

e.g.,

(1) <ICE-HK: SIA-075#24: I:A> Most of the job that is base on uhm <,> **how to say**_that uhm <,> it's temporary

(2) <ICE-HK: SIA-052#681: I:A>

It's an elementary course and if I can uh manage this course I'm going to learn to join the <,> the second class<,> I don't know **how to say** <,> it's <,> immediate

<ICE-HK: SIA-052#X686: I:Z> Intermediate

5. RESULT

44% of the 3 overlap bundles (225 out of 519) are **NOT** found in ICE-HK

The “missing” lexical bundles **NOT** detected in ICE-HK:

- bundles with a main verb in **past tense (24%)**: e.g., *and I thought, it was really, we used to , how did you...*
- semi-modal bundle: *be supposed to*
- Vague bundles: *and all that, stuff like that, a bit of, a bit more, sort of like, sort of thing*

6. DISCUSSION

- The overview of **the overlap and non-overlap bundles** in HKE demonstrates remarkable features of interlanguage arising from language acquisition factors.

e.g.,

- Whether or not to use **past tense** could be affected by the content related to the conversational topics. However, it is more likely that it **could be a result of LI interference** (Paquot 2013). In Chinese, the **finite / non-finite** or **tensed/ non-tensed distinction** does not exist. The distinction is often blurred in HKE, resulting in main verbs not being marked for tense (Hung 2012).

6. DISCUSSION

- Compared with HKE, SgE is a more developed dialect.
- Its more advanced status can be reflected by the larger number of bundles shared with BrE and CanE.

Difference:

- A variety of vague bundles are detected in ICE-SIN.
- SgE speakers prefer organizing discourse bundles (*and*) *then after that, that's why* and *it's like* to maintain the flow of conversation.

6. DISCUSSION



- The proportions of bundles shared between each pair of varieties is consistent with Schneider's (2007) Dynamic Model of New Englishes:
- Of the three new varieties, **CanE**, having fully reached the final phase (**Phase 5**) *differentiation* in the Dynamic Model, **shares the greatest number of bundles** with the source variety BrE.
- Following CanE, **SgE** has 20 percent of bundles overlapping with BrE, correlating with its evolutionary stage Phase **4** *endonormative stabilisation*
- At the other extreme of the evolutionary cycle, **HKE** representing Phase **3** *nativisation*, despite showing some traces of **Phase 2** *exonormative stabilisation*, has the smallest number of bundles in common with BrE.

The **more advanced a variety** is in the evolutionary cycle, the **more** its use of lexical bundles **are shared** with the source variety BrE in conversation.

7. CONCLUSION

- The **core-periphery approach** has helped reveal some interesting findings which has been overlooked in previous studies.
- All speakers of English varieties **employ the common prefabricated devices** to organize discourse, interact with interlocutors, sustain conversational routine, and maintain the flow of speech.
- However, **divergence** also coexists despite the **convergence** – different speech communities also has its **preferred patterns** in constructing utterances in conversation.

ACKNOWLEDGEMENT

I would like to thank my PhD supervisor Prof. General Nelson for his guidance and feedback on this research project.

Thank you very much for your attention.

Q & A