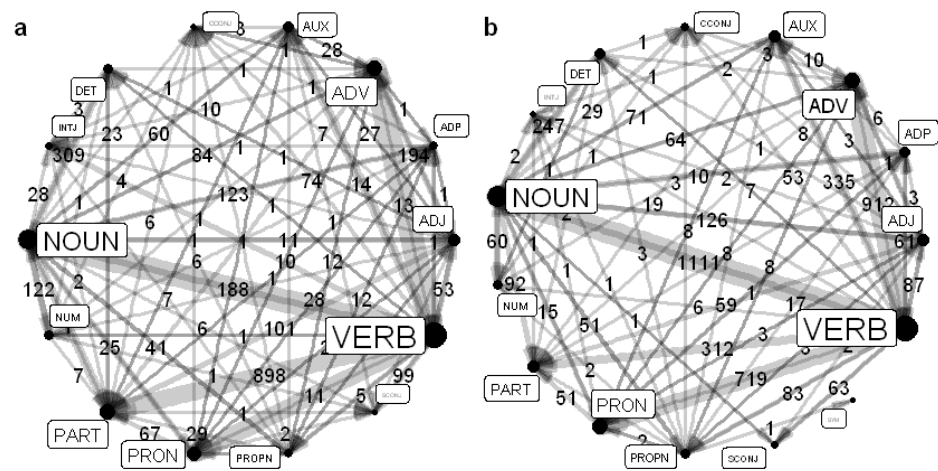


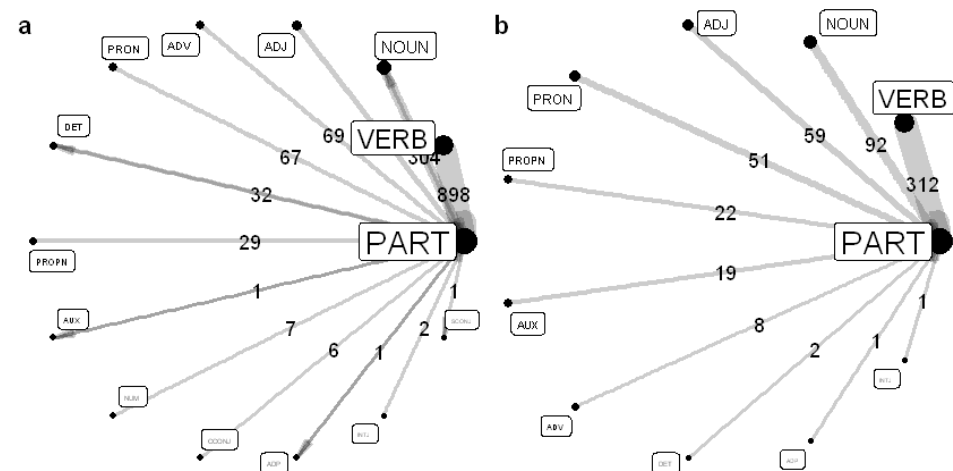
Sentence-final particles as a central node in the scale-free and small-world dependency network

Comparison between Mandarin Chinese and Cantonese



陳翊晞 閻建瑋

2024-07-06



Outline

● Introduction

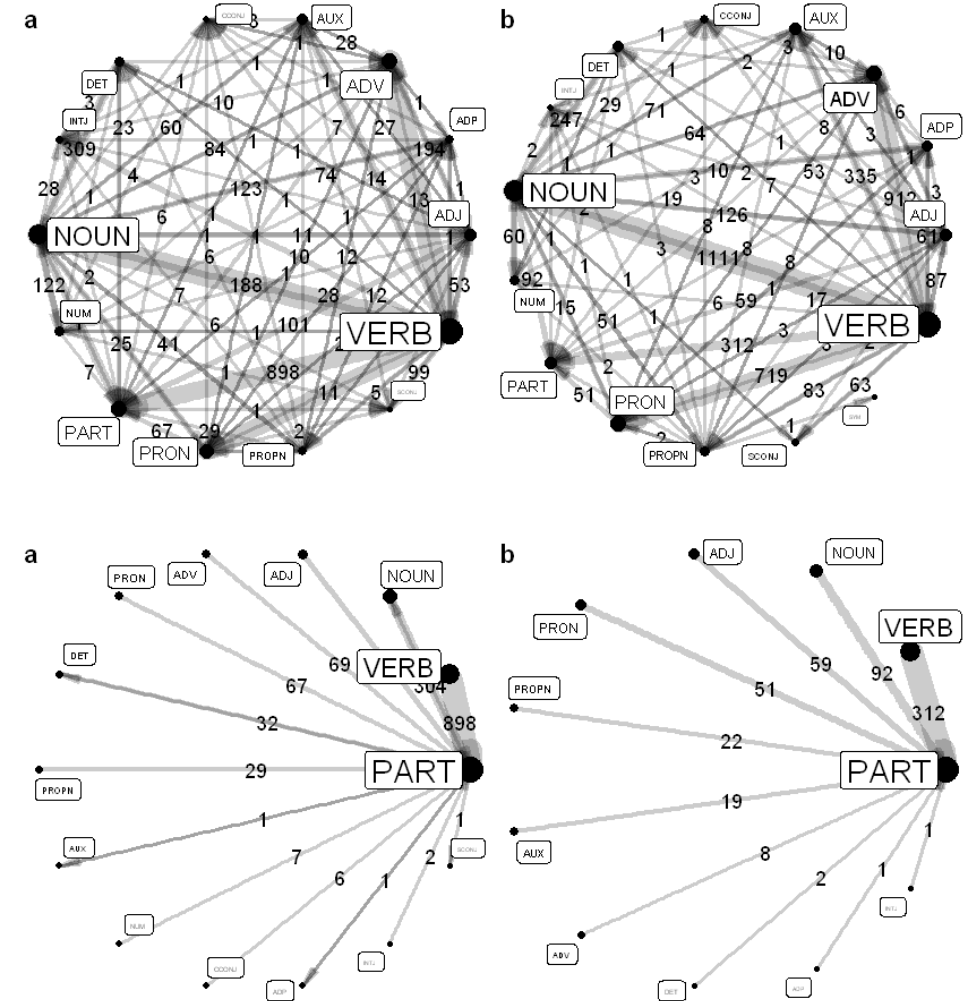
● Data and methods

- Data
- Methods

● Results and discussion

- Scale-free and small-world
- Hubs and centrality
- Ego-centric analysis

● Conclusion



Introduction

Attached SFP	Sentence	Description of the SFP
噏 (wo5)	佢哋度搵嘢 <u>噏</u> 。 'It is <u>said that</u> they are looking for something.'	Certainty about the stated fact
㗎 (gwaa3)	佢哋度搵嘢 <u>㗎</u> 。 'It is <u>probable that</u> they are looking for something.'	Uncertainty about the stated fact
㗎 (ze1)	佢哋度搵嘢 <u>㗎</u> 。 '(Nothing.) They are <u>just</u> looking for something.'	Understatement of the stated fact

- SFP in spoken Cantonese is more pervasive and diverse than it is in Mandarin Chinese.
 - Mood expression and speakers' attitude

Is the rich inventory of SFP in spoken Cantonese unnecessary, or even redundant?

- If yes, the richness of Cantonese SFP may tax extra efforts for speakers. It is not efficient to use so frequently and pervasively.
- If not, Cantonese SFP should make some difference in the structure of linguistic system.

Introduction

🤔 **Is the rich inventory of SFP in spoken Cantonese unnecessary, or even redundant?**

- If yes, the richness of Cantonese SFP may tax extra efforts for speakers. It is not efficient to use so frequently and pervasively.
- If not, Cantonese SFP should make some difference in the structure of linguistic system.

🤔 **What should “the structure of linguistic system” look like?**

- “Language is a network of concepts which in turn is part of the general cognitive network of the mind.” (Hudson, 2006)
- “Syntax is the component of language that has been the most controversial in the discussions about whether languages are optimized for efficient communication.” (Fedorenko et al., 2024)

💡 **Language can be a network, where linguistic components are mutually connected by syntactic relations.**

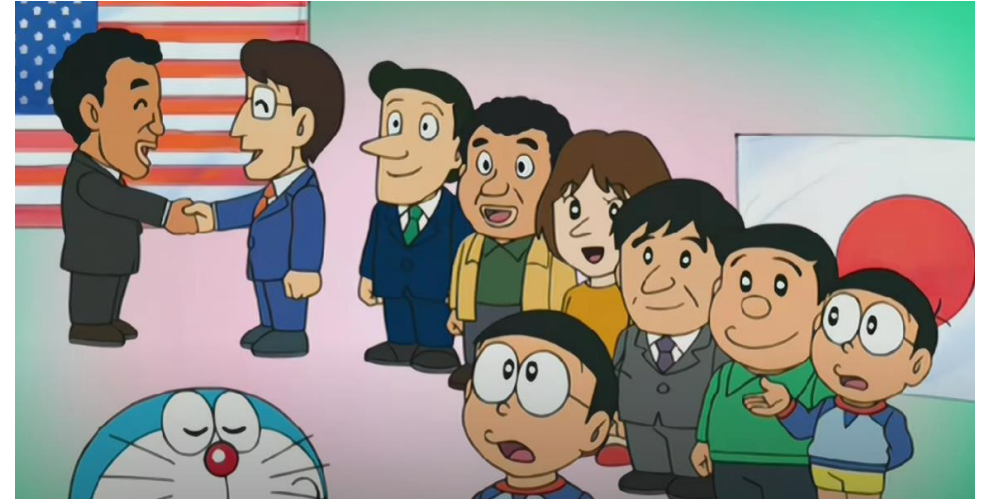
Hudson, R. (2006). *Language Networks: The New Word Grammar*. Oxford University Press.

Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, 630(8017), 575–586.

Introduction

Previous studies found that a language network has some common features shared by real world networks (e.g., social network and internet).

- **Scale-free (無標度)**: only a very small number of nodes bear a overwhelming number of connections
- **Small-world (小世界)**: a node can be connected to any other node in the network within a few steps



Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, 11(4), 598–618.


Corominas-Murtra, B., Valverde, S., & Solé, R. (2009). The ontogeny of scale-free syntax networks: Phase transitions in early language acquisition. *Advances in Complex Systems*, 12(03), 371–392.

Ferrer-i-Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3), 788–791.

Ferrer-i-Cancho, R., Solé, R. V., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, 69(5), 051915.

Liu, H. (2008). The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and Its Applications*, 387(12), 3048–3058.

Introduction

 Is the rich inventory of SFP in spoken Cantonese unnecessary, or even redundant?

Are spoken Cantonese and Mandarin Chinese efficient as indicated by previous studies?

(scale-free and small-world)

Does SFP cause structural differences between these two languages? If yes, what are the differences?

(hubs and centrality)

Data and Methods

Data in the current study consists of two paralleled treebanks retrieved from the database of Universal Dependencies (UD):

- UD Cantonese HK (**UD-CANT**, *N* of sentences = 1,004, *N* of tokens = 13,918)
- UD Chinese HK (**UD-CHIN**, *N* of sentences = 1,004, *N* of tokens = 9,874)

* UD treebanks are open-sourced and available on <https://universaldependencies.org/>. The version used in the current study is 2.12, released on May 15, 2023.

你喺度搵乜嘢呀? (UD-CANT)

<i>nei5</i>	<i>hai2-dou6</i>	<i>wan2</i>	<i>mat1-je5</i>	<i>aa3</i>
you	PROG	look for	WH	SFP

你在找些什麼? (UD-CHIN)

<i>nǐ</i>	<i>zài</i>	<i>zhǎo</i>	<i>xiē</i>	<i>shén-me</i>
you	PROG	look for	CLF	WH

‘What are you looking for?’

Data and Methods

Under the framework of dependency grammar, dependencies refer to the binary asymmetric syntactic relation between two words. Both treebanks were manually annotated in terms of POS tags and dependency relations and coded in CoNLL-U files, which are commonly used in dependency annotation.

💡 Therein, SFP is annotated as “PART”.

Token	Token Order	Token POS	Head Order	Head POS	DEP
你	1	NOUN	3	VERB	nsubj
喺度	2	ADV	3	VERB	advmod
搵	3	ROOT	0		root
乜嘢	4	PRON	3	VERB	obj
呀	5	PART	3	VERB	discourse

Data and Methods

We built two syntax networks of spoken Cantonese and spoken Mandarin Chinese based on dependency grammar.

💡 We converted the word-based dependencies to **POS-based dependencies to explicitly examine the dependencies between SFP and other categories of words.**

Liu, H. (2008). The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and Its Applications*, 387(12), 3048–3058.

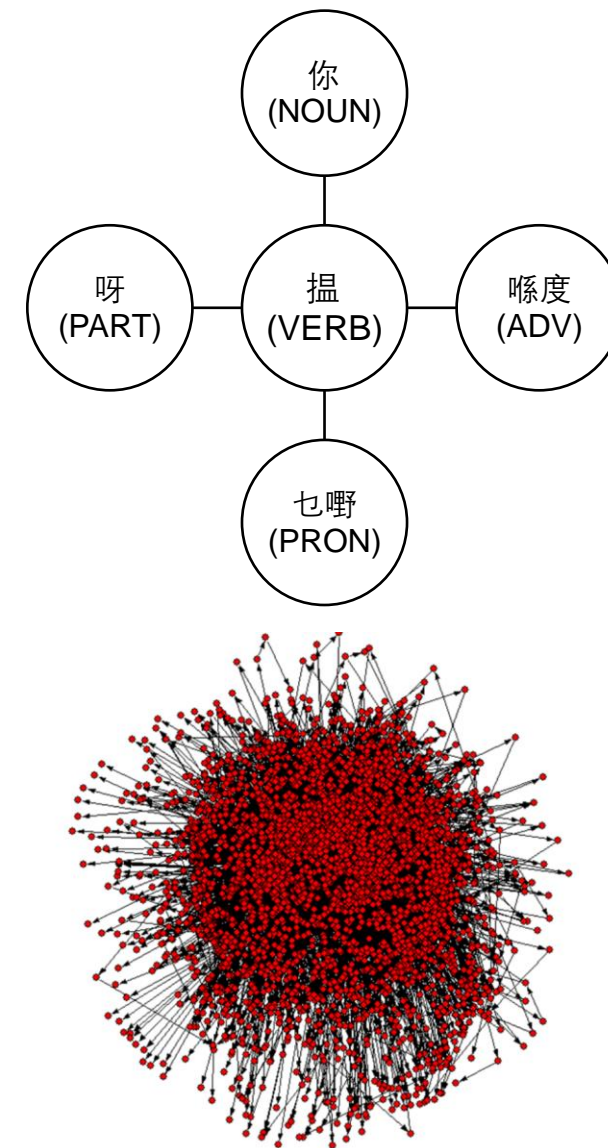


Fig. 4. Syntactic network built on xwlb treebank.

Results and discussion

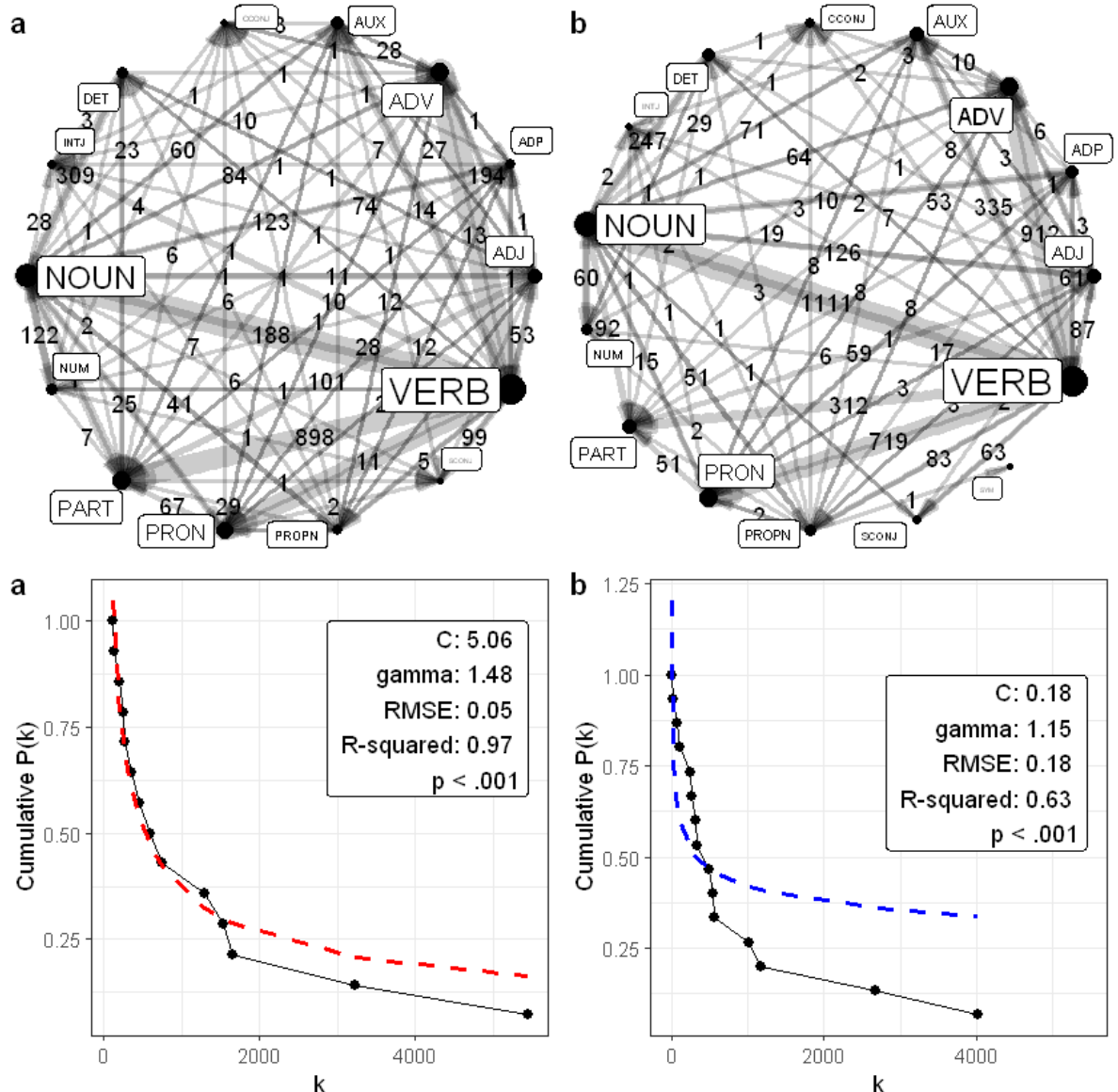
Scale-free and small-world

Directed and weighted network

- Node: POS
($n = 14$; $n = 15$)
- Edge: dependency
($m = 124$; $m = 101$)
- Weight: freq. of dependency
($m_c = 5446$; $m_c = 4023$)

Weighted degree distribution

$$P(k) \sim k^{-(\gamma-1)}$$
$$P(k) = \frac{C}{\gamma-1} k^{-(\gamma-1)}$$




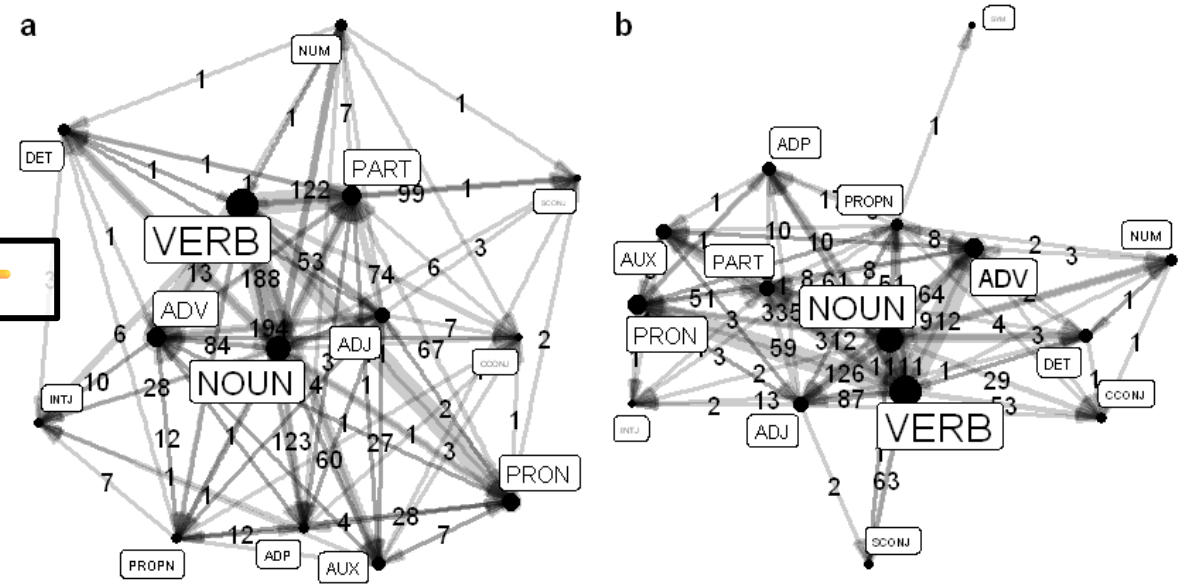
Results and discussion

Scale-free and small-world

Directed and weighted network

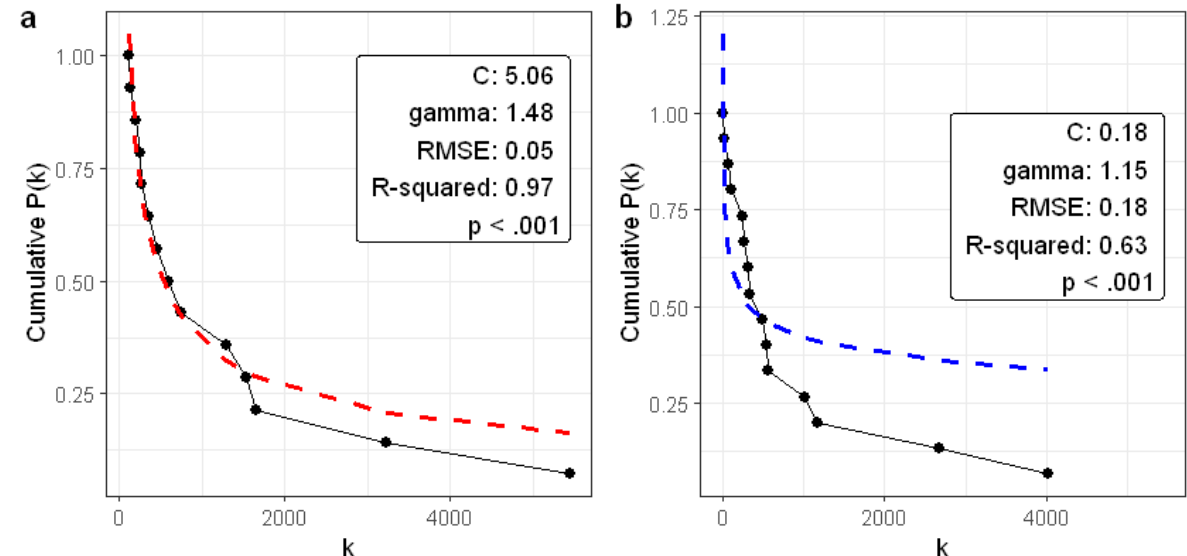
- Node: POS
($n = 14$; $n = 15$)
- Edge: dependency
($m = 124$; $m = 101$)
- Weight: freq. of dependency
($m_c = 5446$; $m_c = 4023$)

(Different layout) 



Weighted degree distribution

$$P(k) \sim k^{-(\gamma-1)}$$
$$P(k) = \frac{C}{\gamma-1} k^{-(\gamma-1)}$$

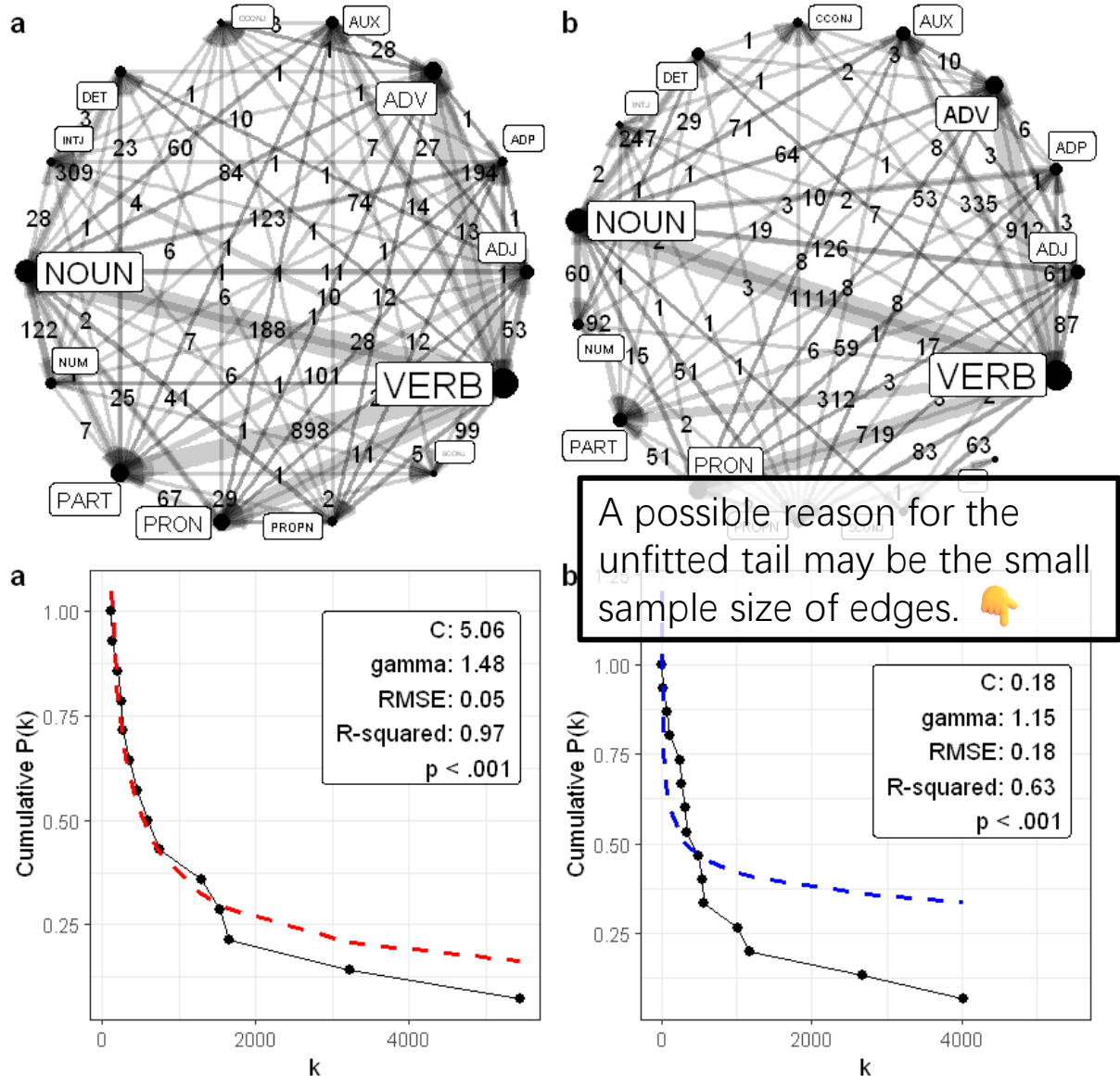


Results and discussion

Scale-free and small-world

Directed and weighted network Weighted degree distribution

- These POSs also serve as the hub nodes, bearing rich connections with other nodes in the dependency network: NOUN (noun) , VERB (verb), PRON (pronoun) and ADJ (adjective).
- (Express meaning with a few categories of words)



Results and discussion

Scale-free and small-world

Small-world: lower L, higher C and higher E_g

(Compared to simulated random and scale-free networks, times = 500)

D			L		C		E_g	
	Diameter		Average path length		Clustering coefficient		Global efficiency	
UD-CANT	23		5.88		0.84		0.34	
NX_rand	405.69	***	43.34	***	0.85	***	0.19	***
NX_sf	114.17	***	12.87	***	0.67	***	0.24	***

D			L		C		E_g	
	Diameter		Average path length		Clustering coefficient		Global efficiency	
UD-CHIN	17		5.33		0.78		0.24	
NX_rand	668.95	***	61.61	***	0.69	***	0.17	***
NX_sf	200.08	***	19.5	***	0.58	***	0.21	***

Results and discussion

Scale-free and small-world

- Both UD-CANT and UD-CHIN are small-world networks, which means information therein can be efficiently transmitted.
- UD-CANT seems to be more globally efficient (more adjacent dependencies; possibly more colloquial and less formal).

D			L		C		E _g	
	Diameter		Average path length		Clustering coefficient		Global efficiency	
UD-CANT	23		5.88		0.84		0.34	
NX_rand	405.69	***	43.34	***	0.85	***	0.19	***
NX_sf	114.17		12.87		0.67	***	0.24	***

D			L		C		E _g	
	Diameter		Average path length		Clustering coefficient		Global efficiency	
UD-CHIN	17		5.33		0.78		0.24	
NX_rand	668.95	***	61.61	***	0.69	***	0.17	***
NX_sf	200.08	***	19.5	***	0.58	***	0.21	***

Results and discussion

Hubs and centrality

POS	hub	eig	clo	rank_hub	rank_eig	rank_clo	Weighted degree
VERB	1.0000	1.0000	1.0667	1	1	13	5446
NOUN	0.1412	0.7028	0.8371	2	2	14	3232
ADV	0.0184	0.5306	4.9333	8	3	6	1652
PART	0.0004	0.4719	8.3333	12.5	4	1	1532
PRON	0.0235	0.4346	4.6190	6	5	7	1395
ADJ	0.0909	0.1876	3.7024	3	7	9	748
POS	hub	eig	clo	rank_hub	rank_eig	rank_clo	Weighted degree
VERB	1.0000	1.0000	1.4436	1	1	11	4623
NOUN	0.0856	0.7830	2.1640	2	2	10	2673
ADV	0.0041	0.5060	5.5000	10	3	4	1168
PRON	0.0171	0.4323	4.3762	7	4	8	1011
PART	0.0000	0.2152	0.0000	14	5	14	567
ADJ	0.0731	0.1759	5.7500	3	7	2	545

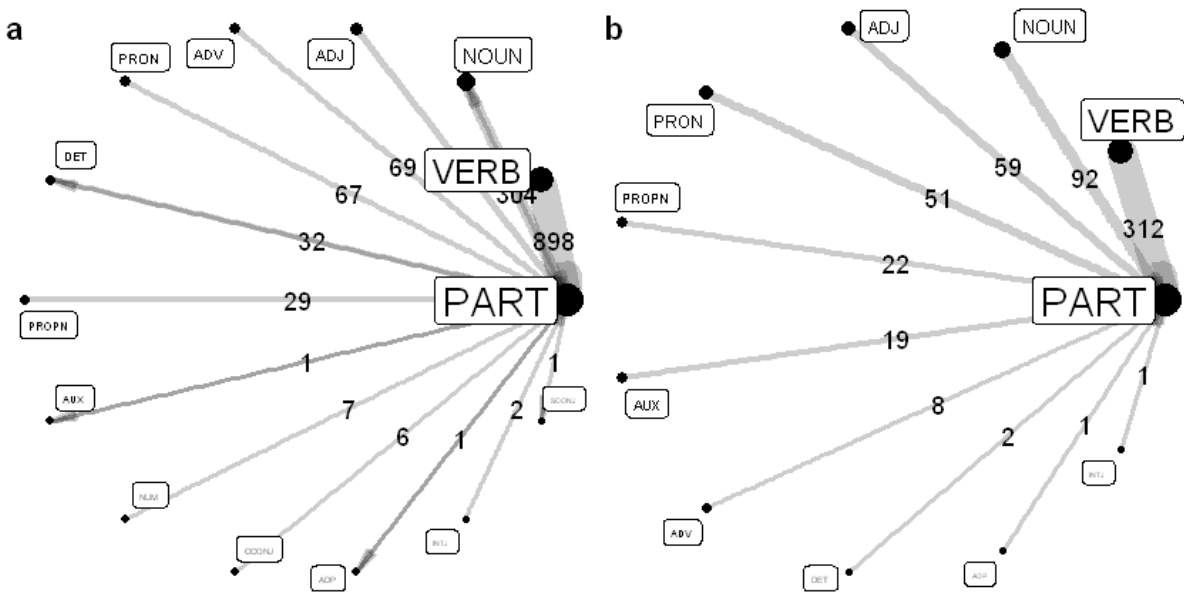
VERB and NOUN are hubs, but PART is more closely connected to other POSs in UD-CANT although it is not a hub.

Results and discussion

Ego-centric analysis

Ego-centric analysis only focuses on PART and its neighboring nodes.

- Majority of the connections are directed from (or to) NOUN, ADJ and VERB.
- PART has more diverse neighboring nodes in UD-CANT.



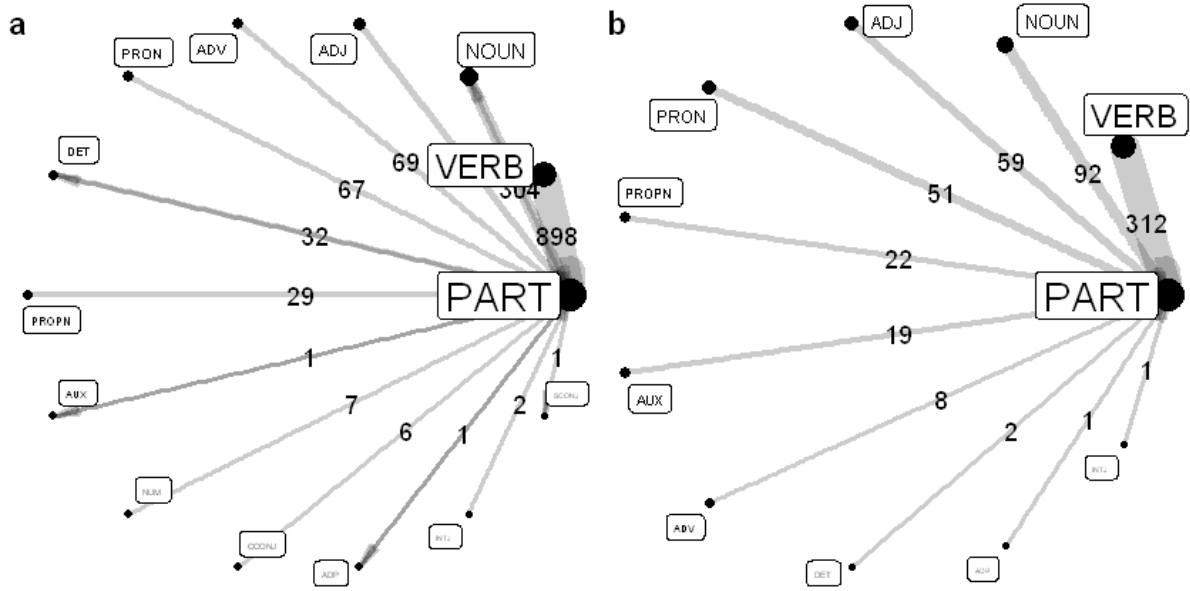
(UD-CANT)	(UD-CHIN)	POS DEP (Freq.)
你如果係選嘅話，一陣半個鐘頭之內， <u>十分鐘</u> 之內 <u>添</u> ，佢就已經係 <u>主席</u> 㗎嘞。 ("even"; seeking for mutual agreement)	如果要進行選舉的話，有人將於稍後半小時，或 <u>十分鐘</u> 後便會當選 <u>主席</u> 。	PART <- NOUN (189 17)
中間唔要有空白呀，知 <u>唔</u> 知 <u>呀</u> ？ (softening the negation and lessening the face threat)	中間不要有空白， <u>明白</u> 嗎？	PART <- ADV (58 0)
係呀，好 <u>煩</u> 㗎，佢。 (expressing determination and assertion)	對，她有點 <u>煩</u> 人。	PART <- ADJ (57 25)

Results and discussion

Ego-centric analysis

Dependencies of PART is used differently in UD-CANT and UD-CHIN.

- PART-NOUN (discourse:sp), PART-ADV and PART-ADJ
- Featured use to express moods



(UD-CANT)	(UD-CHIN)	POS DEP (Freq.)
你如果係選嘅話，一陣半個鐘頭之內， <u>十分鐘</u> 之內 <u>添</u> ，佢就已經係 <u>主席</u> 㗎㗎。 ("even"; seeking for mutual agreement)	如果要進行選舉的話，有人將於稍後半小時，或 <u>十分鐘</u> 後便會當選 <u>主席</u> 。	PART <- NOUN (189 17)
中間唔要有空白呀，知 <u>唔</u> 知 <u>呀</u> ？ (softening the negation and lessening the face threat)	中間不要有空白， <u>明白</u> 嗎？	PART <- ADV (58 0)
係呀，好 <u>煩</u> 㗎，佢。 (expressing determination and assertion)	對，她有點 <u>煩</u> 人。	PART <- ADJ (57 25)

Conclusion

- **Communication efficiency:** POS-based dependency networks of both spoken Cantonese and Mandarin Chinese (still) manifest scale-free and small-world properties. These two properties enable efficiency in communication.
- **SFP as a central node:** SFP, represented by PART, serves as a central node in spoken Cantonese but not in Mandarin Chinese. It suggests that the richness of SFP in spoken Cantonese stems from its strong capacity to connect with other POSs and enables “shortcuts” in the syntactic structure.
- **Mood expression:** Cantonese SFPs displayed more extensive mood expression devices, notably differing in PART-NOUN (discourse:sp), PART-ADV and PART-ADJ dependencies from Mandarin.

Thank you for listening

Email: canjikhei1030@outlook.com

Tel: +86 136 3209 9006