Sentence-final particles cause different structures in Mandarin Chinese and Cantonese: Evidence from the Universal Dependencies treebank

CHEN Yixi YAN Jianwei
Department of Linguistics, Zhejiang University, Hangzhou, China
E-mail: canjikhei1030@outlook.com

Introduction

Attached SFP	Sentence	Description of the SFP
喎 (wo5)	佢喺度搵嘢喎。 'It is said that they are looking for something.'	Certainty about the stated fact
啩 (gwaa3)	佢喺度搵嘢啩。 'It is probable that they are looking for something.'	Uncertainty about the stated fact
咗 (ze1)	佢喺度搵嘢咗。 '(Nothing.) They are just looking for something.'	Understatement of the stated fact

Compared to spoken Cantonese, spoken Mandarin Chinese, a language syntactically similar to Cantonese, equally uses SFP to express moods, but SFP therein is more "general in meaning and broad. **The richness of SFP in spoken Cantonese thus seems unnecessary and taxes extra efforts in language processing.**

If the pervasiveness and richness of SFP in spoken Cantonese are not trivial, SFP's influence on the mood should not be solely contextual, but also structural. In other words, **Cantonese SFP should have a solid capacity to connect other language components, which renders Cantonese structurally different from other languages.**

Data and Methods

Universal Dependencies (UD) treebank

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)				
nei5	*hai2-dou6*	*wan2*	*mat1-je5*	*aa3*
you	PROG	look for	WH	SFP
你在找些什麽? (UD-CHIN)				
nǐ	*zài*	*zhǎo*	*xiē*	*shén-me*
you	PROG	look for	CLF	WH
'What are you looking for?'				

Dependency grammar

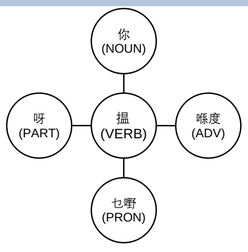
Under the framework of dependency grammar, dependencies refer to the binary asymmetric syntactic relation between two words. Both treebanks were manually annotated in terms of POS tags and dependency relations and coded in CoNLL-U files, which are commonly used in dependency annotation. Therein, SFP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

POS-based syntax network

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

Notably, we converted the word-based dependencies to **POS-based dependencies to explicitly examine the dependencies between SFP and other categories of words.** We calculated the following metrics to quantify the features of two constructed networks:



Degree k of VERB = 4

Metric		Description
Mean degree $\bar{k}$	$\bar{k} = \frac{1}{n}\sum_{i=1}^{n} k_i$	Overall capability of one node connecting with others in a network (n, the number of nodes; m, the number of edges)
Network density ρ	$\rho = \frac{m}{n(n-1)}$	
Degree centralization NC	$NC = \frac{n}{n-1}\left(\frac{k_{max}}{n-1} - \rho\right)$	Importance of hub nodes for the whole network
Scale-free property $P(k)$	$P(k) \sim k^{a}$	Only a small number of nodes have high degrees in the network (~ the principle of least effort)

Conclusion

- **Communication efficiency**: POS-based dependency networks of both spoken Cantonese and Mandarin Chinese manifest centralization and scale-free properties. These two properties enable efficiency in communication.
- **SFP as a hub node**: SFP, represented by PART, serves as a hub 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.

Results and Discussion

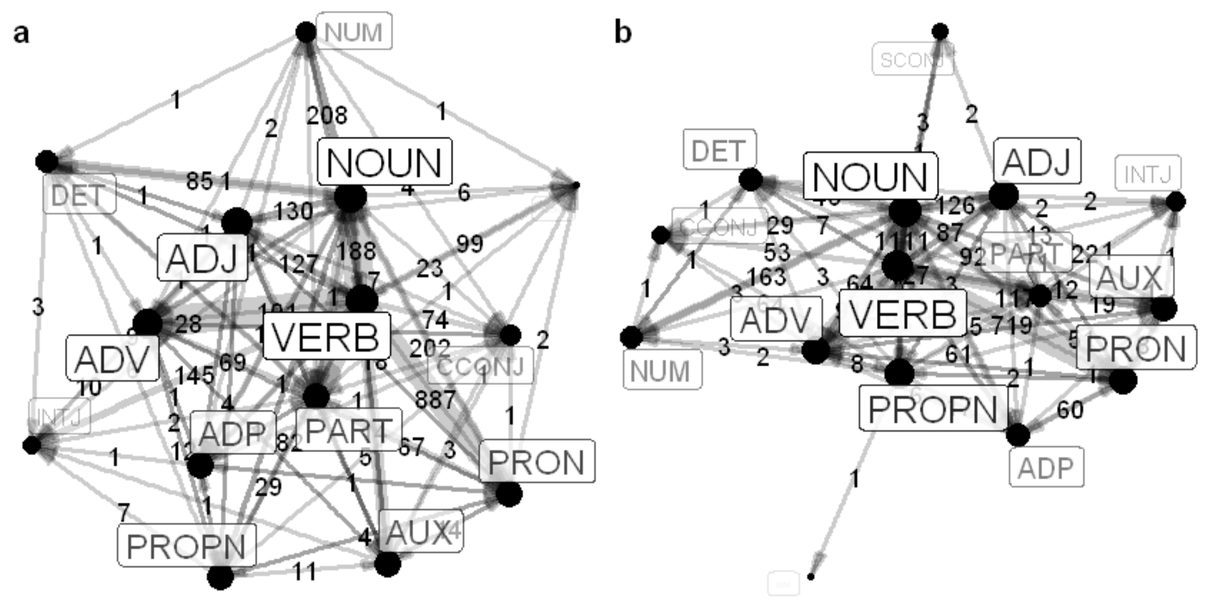
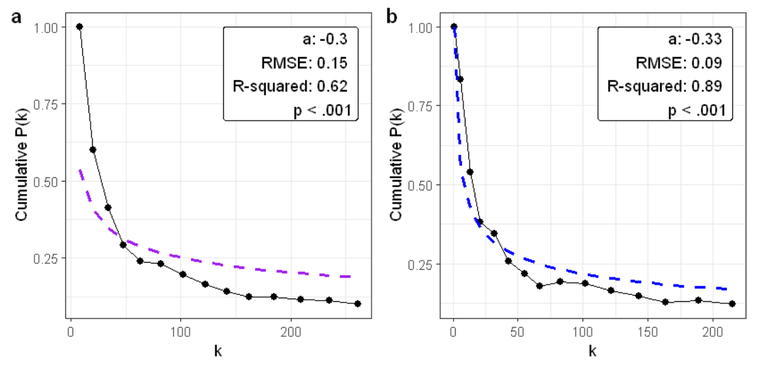


Fig. 1 POS-based dependency networks of UD-CANT (a) and UD-CHIN (b)

The size and alpha of each label in each figure represent the degree of the POS, and the width of each edge is related to the frequency of the dependency, which is also annotated on the edge.

These POSs also serve as the hub nodes, bearing rich connections with other nodes in the dependency network: NOUN (noun) , VERB (verb), ADV (adverbial), ADJ (adjective), PROPN (proper noun) and PRON (pronoun).

One noteworthy point is that **PART (particle), representing SFP, serves as the hub node only in spoken Cantonese.** The prominence of PART aforementioned is equally captured by its greater degree in spoken Cantonese (kPART = 20) than that in spoken Mandarin (kPART = 11). It supports the richness of SFP as a peculiarity in spoken Cantonese.



	UD-CANT	UD-CHIN
n	14	15
m	124	101
$\bar{k}$	18.64	14.33
ρ	.86	.48
NC	1.42	1.47

Fig. 2 Cumulative distributions of degrees based on UD-CANT (a) and UD-CHIN (b)

Greater ρ indicates that **the network of spoken Cantonese manifests stronger centralization properties.** In other words, the nodes therein have a stronger capability of connecting with others than their counterparts in spoken Mandarin Chinese.

The scale-free property is validated by the goodness-of-fit of the degree distribution. **only a small number of hub nodes are structurally important and connected to a large number of nodes** in the two dependency networks. These hub nodes allow "shortcuts" in the network to **ensure communication efficiency**.

(UD-CANT)	(UD-CHIN)	POS DEP (Freq.)
你如果係選嘅話，一陣半個鐘頭之內，十分鐘之內添，佢就已經係主席㗎嘞。 ("even"; seeking for mutual agreement)	如果要進行選舉的話，有人將於稍後半小時，或十分鐘後便會當選主席。	PART <- NOUN (189 | 17)
中間唔要有空白呀，知唔知呀? (softening the negation and lessening the face threat)	中間不要有空白，明白嗎?	PART <- ADV (58 | 0)
係呀，好煩㗎，佢。 (expressing determination and assertion)	對，她有點煩人。	PART <- ADJ (57 | 25)