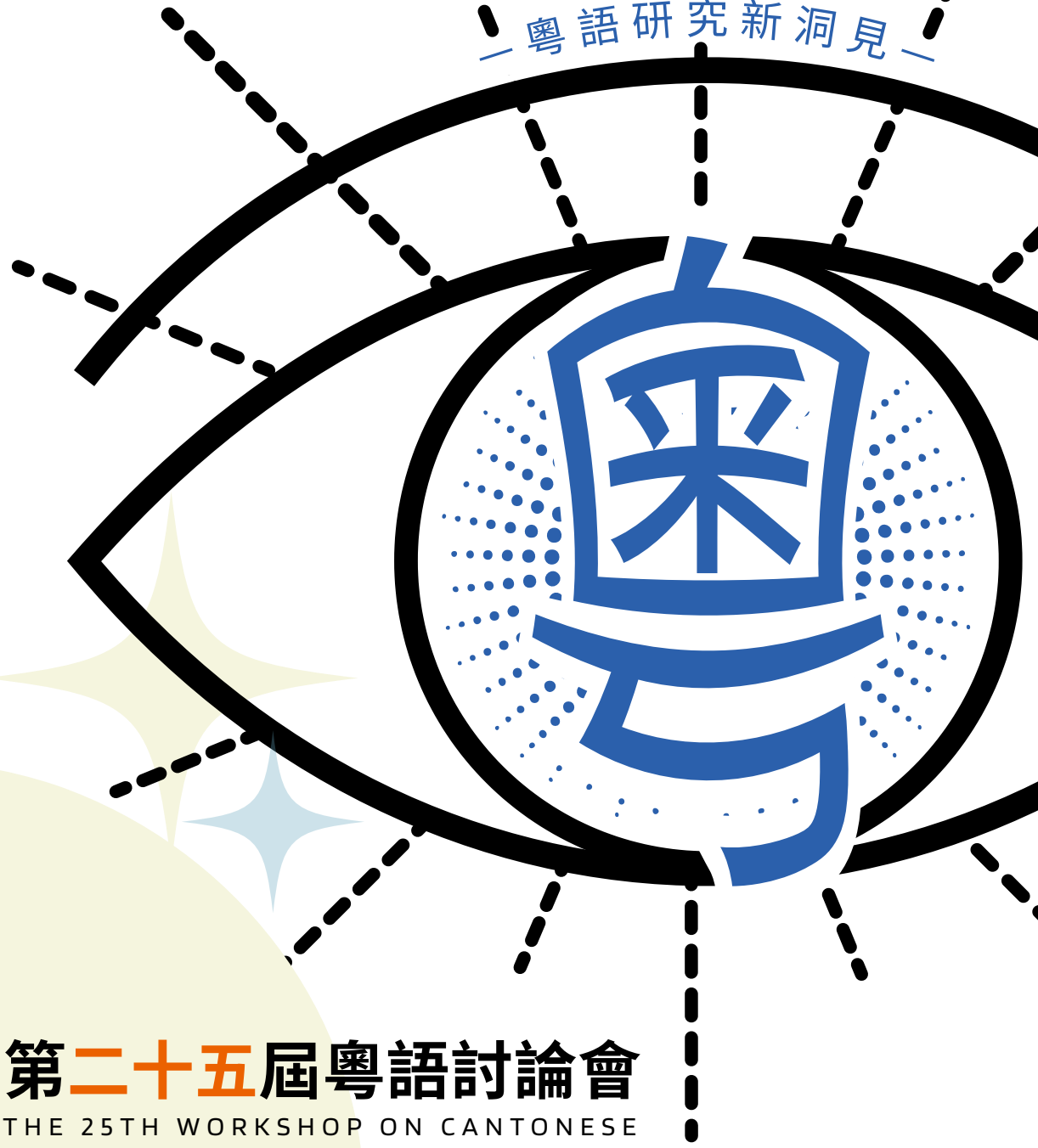
由細睇到大

粵語研究新洞見



WoC 25 第二十五屆粵語討論會

THE 25TH WORKSHOP ON CANTONESE

程序表及會議摘要

Programme and Abstracts

2025 年 **7** 月 **19** 日 (星期六)

19 JULY 2025 (SATURDAY)

香港教育大學大埔校園

The Education University of Hong Kong (EdUHK), Tai Po Campus

Centre for Research on Linguistics and Language Studies

第二十五屆粵語討論會 WOC-25 🧑

🧿 由細睇到大——粵語研究新洞見 ✨

FROM SEED TO SIGNIFICANCE: NEW INSIGHTS IN CANTONESE STUDIES

日期：2025年7月19日（星期六）
地點：香港教育大學大埔校園
主辦：香港語言學學會（LSHK）及 香港教育大學語言學及語言研究中心（CRLLS, EdUHK）
Registration 報到處： B2-LP-16 (Room 16, LP Floor, Block B2) @EdUHK

程序表 Programme (v1 20250715)

10:00	報到 Registration	
10:15	開幕致辭 Welcoming Remarks	
	Session 1A (B2-LP-13)	**Session 1B (B2-LP-14)**
10:30	**「一陣」與「陣間」： 從詞序差異看粵語時間副詞的句法分布與語義解釋** 吳啓鋒 澳門大學	**Heritage Cantonese learner linguistic identities: Navigating cultural preservation and language ideologies in multicultural Canada** Raymond Pai The University of British Columbia
11:00	**The absence of D-linking effects in Cantonese Wh-islands** Chi Hou Pau, Grant Goodall UCSD	**Tone Change of Disyllabic English Names in Hong Kong Cantonese** MAK Pui Ying (Jolene) The University of Hong Kong
11:30	**Is Cantonese Dou1 Changing?** Charles Lok The Chinese University of Hong Kong	**由細唱到大——香港小學粵語校歌的樂調與字調** 李佳妮，張凌 香港教育大學
12:00	午膳 Lunch	
	Session 2A (B2-LP-13)	**Session 2B (B2-LP-14)**
13:30	**The Structural Organisation of Cantonese Sentence-Final Particles: A Minimalist Perspective** Hoi Ming Chan University of Cambridge	**Emoji是一個詞嗎？—— 從傳統語法視角分析香港與內地社交平台的表情符號** 林雅珊，萬佳琪，萬佳瑤，陳健榮 香港樹仁大學
14:00	**粵語「喺度」的功能及身份探析** 陳子軒 香港中文大學	**《英華分韻撮要》量詞研究** 霍穎欣，陳以信 香港伍倫貢學院
14:30	**粵語否定詞「未」的語義** 劉海澄 香港中文大學	**廣東高校粵語播音主持專業學生粵語能力弱化成因研究** 曾岑 廣州大學新聞與傳播學院
15:00	茶歇 Tea Break B2-LP-16	
	Session 3A (B2-LP-13)	**Session 3B (B2-LP-14)**
15:30	**粵語句末助詞“啊aa3”和“喇laa3” 在話語情境裏所起的作用** 謝國強 Nanyang Polytechnic Singapore	**Towards Glyphon Frequencies in Written Cantonese** Jon Chui canto.hk
16:00	**粵語玉林話的兩個否定詞** 郁孫豪[1]，晨曦[2] 1 香港科技大學，2 廣西大學	**大型語言模型（LLM）粵語知識盲點分析** 劉擇明 香港教育大學
16:30	**閉幕 Closing** B2-LP-16	

「一陣」與「陣間」：從詞序差異看粵語時間副詞的句法分布與語義解釋

澳門大學 吳啓鋒
mc43620@um.edu.mo

摘要

粵語時間副詞「一陣」(jat1 zan6) 與「陣間」(zan6 gaan1) 均指涉短暫的時間，在口語中使用頻繁，但其句法行為與語義內涵的細微差別常為人忽略，現有研究對這兩者亦缺乏系統性的比較分析。本文旨在深入探討此二詞的詞彙構成、句法分布規律、語義解釋及語用功能，核心關注點在於分析詞序差異與其功能之間的聯繫。

首先，本研究採用描寫語言學與比較分析的方法，客觀描述「一陣」與「陣間」的語素構成：「一陣」由數詞「一」與核心語素「陣」構成，「陣間」由核心語素「陣」與黏著語素「間」構成。

其次，通過考察兩者在句首、動詞前、動詞後及句末等不同句法位置的出現情況，及其與體貌標記、句末助詞等其他句子成分的共現模式，系統比較其句法分布的異同。研究發現，「一陣」與「陣間」在句法分布上呈現顯著差異。「陣間」更常出現於句首位置；其亦可出現於動詞前。相比之下，「一陣」較少獨立出現於句首；其句法位置更為靈活，常見於動詞前；更典型地出現於動詞之後。研究認為，由於語素「間」強化了其時間區間的指涉功能，句首的「陣間」常發揮為整個句子設定時間背景的功能（框架性時間副詞）。分析表明，動詞後的「一陣」在功能上近似時量補語或界限體標記，表示動作的短暫持續或即時的短暫延遲。

再者，在語義內涵上，通過語境分析發現兩者存在細微分野。「一陣」可用於指涉緊隨說話時刻的未來、當下的短暫動作或過去某個動作的短暫持續。而「陣間」主要用於指向相對於說話時間稍後的未來時間點或區間。研究推測，「一陣」的核心語義側重於「即時性」和「極短的持續」。分析認為，「陣間」所指的未來時段相對於「一陣」所指的未來略有延遲；基於其語素「間」的明確標示，其所指的未來時段在概念上可能更為界定和獨立。

最後，在語用功能方面，基於語料觀察與分析，「陣間」更常用於提及已有所預期或初步計劃在稍後發生的事件；而「一陣」則更多用於即時的、非預謀的短暫行為或延遲請求。兩者在會話管理和禮貌表達等層面展現出不同的適用場景。研究認為，句首的「陣間」常承擔時間話題的功能，而動詞後的「一陣」則常構成焦點的一部分。分析探討了兩者在句法結構中的附着高度差異與其作用域不同的可能關係。此外，研究指出「一陣」在動詞後位置呈現類體貌化的趨勢，這為漢語方言詞彙語法化的研究提供了值得探討的例證。

本研究的結果有助於深化對粵語時間表達系統精細分工的理解，建立可延伸的分析框架，為日後其他粵語副詞如「一陣間」和「有陣時」等的研究奠下基礎，同時深入探討了詞序與語義功能之間的聯繫，並可為粵語教學及漢語方言比較研究提供有價值的參考。

關鍵詞 粵語、時間副詞、句法分布、詞序、語素分析

The absence of D-linking effects in Cantonese Wh-islands

Chi Hou Pau, Grant Goodall
University of California, San Diego
chpau@ucsd.edu

Background: There has been a debate on what the nature of island constraints is. D-linking effects in wh-islands provide an opportunity to examine this issue. It is observed that the use of D-linked wh-phrases ameliorates wh-island effects as demonstrated in the contrast between (1) and (2). Syntactic accounts ascribe the D-linking effects to Featural Relativized Minimality (fRM) [1] whereas processing accounts ascribe it to reduced similarity-based interference [2]. Alternatively, semantic accounts explain the D-linking effects in terms of semantic incompatibility [3]. While these accounts predict the presence of D-linking effects in wh-ex-situ languages, their predictions about D-linking effects in wh-in-situ languages diverge. For syntactic accounts, assuming a covert movement analysis [4], fRM predicts the presence of D-linking effects as D-linking reduces the degree of morphosyntactic feature overlap as in (3). For processing accounts, processing an in-situ wh-phrase involves a cue-based retrieval to search for its scope position [5] and the scope position of an embedded question induces similarity-based interference. Since D-linking does not alter the retrieval cues associated with the scope positions, it is irrelevant to the degree of similarity-based interference induced as illustrated in (4). Hence, processing accounts predict the absence of D-linking effects. Semantic accounts predict the presence of D-linking effects as the semantic interpretation of wh-phrases is the same regardless of being in-situ or not. The current study examines the D-linking effects in Cantonese wh-islands using an acceptability experiment.

Method: The experiment applies a 2x2x2 (*Dependency Length* x *Structure Type* x *Wh-phrase Type*) design as shown in (5). Question-answer pairs were presented to ensure the wh-question interpretation. There are four question/answer pairs per condition, resulting in 32 experimental items. Participants were asked to rate each of the question/answer pairs on a scale from 1 - 7.

Results: After filtering inattentive participants, the data collected (N=44) were z-transformed (plotted in (6)) and then fitted with a linear mixed model in (7). The interaction effect between *Dependency Length* and *Structure Type* is estimated to be -0.87992 ($p<2e^{16}$), confirming the presence of wh-island effects. Importantly, the estimation of the three-way interaction does not reach statistical significance (p=0.5096), suggesting the absence of D-linking effects in wh-islands. This finding is also supported by the result of a likelihood ratio test (Chisq=0.4321, p=0.511) comparing models with/without the three-way interaction shown in (8).

Discussion: The results are in favour of processing accounts and against syntactic accounts and semantic accounts. Accounting for wh-island effects in terms of similarity-based inference can explain why D-linking effects are present in wh-ex-situ language but absent in wh-in-situ language, supporting the idea that wh-island effects stemmed from constraints on the retrieval mechanism in working memory. Using wh-islands as an example, this study demonstrates different languages utilize the same mechanism to process wh-dependencies and certain types of island constraints can be reduced to more general constraints on language processing mechanism. This might help to explain why island constraints can be acquired even without negative evidence in the inputs.

(1) *What$_i$ does Tom wonder who repaired _$_i$?

(2) ?Which computer$_i$ does Tom wonder who repaired _$_i$?

[uWH][uN] [+WH][+N] [+WH]

(3) ___ go3 ji1sang1 soeng2zi1dou6 bin1go3 saan3bo3-zo2 bin1zek3 beng6duk6 aa3?
CL doctor wonder who spread-PFV which virus Q
"Which virus does the doctor wonder who spread?"

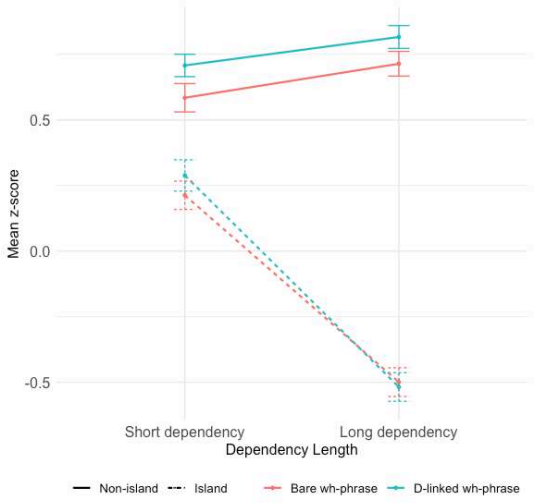
[+C][+Q] [+C][+Q]

(4) [$_{CP}$go3 ji1sang1 soeng2zi1dou6 [$_{CP}$bin1go3 saan3bo3-zo2 bin1zek3 beng6duk6] aa3]?
CL doctor wonder who spread-PFV which virus Q
"Which virus does the doctor wonder who spread?"

(5)

	Short Dependency	***Long Dependency***
Non-island	{bin1go3 ji1sang1/bin1go3} jing6wai6 siu2waa4 saan3bo3-zo2 zek3 {which doctor /who } think Siuwaa spread -PFV CL beng6duk6 aa3? virus Q "{Which doctor/Who} thinks Siuwaa spread the virus?"	go3 ji1sang1 jing6wai6 siu2waa4 saan3bo3-zo2 {bin1zek3 beng6duk6 / mat1je5} aa3? CL doctor think Siuwaa spread -PFV {which virus /what } Q "{Which virus/What} does the doctor think Siuwaa spread?"
Island	{bin1go3 ji1sang1/bin1go3} soeng2zi1dou3 bin1go3 saan3bo3-zo2 zek3 beng6duk6 aa3? {which doctor /who } wonder who spread -PFV CL virus Q "{Which doctor/Who} wonders who spread the virus?"	go3 ji1sang1 soeng2zi1dou3 bin1go3 saan3bo3-zo2 {bin1zek3 beng6duk6 / mat1je5} aa3? CL doctor wonder who spread -PFV {which virus / what } Q "{Which virus/What} does the doctor wonder who spread?"
Answer	ji1zi6 lau4gam2 go2 go3 ji1sang1 treat flu that CL doctor "The doctor who treats flu."	dou6zi3 tou5o1 go2 zek3 beng6duk6 cause diarrhea that CL virus "The virus which causes diarrhea."

(6)



(7) z-score ~ Wh-phrase Type * Structure Type * Dependency Length + (1 + Structure Type | Participant) + (1 + Dependency Length | Item)

(8) z-score ~ Dependency Length * Structure Type + Dependency Length * Wh-phrase Type + Structure Type * Wh-phrase Type + (1 + Structure Type | Participant) + (1 + Dependency Length | Item)

References

[1] Villata, S. (2017). *Intervention effects in sentence processing* [Ph.D.]. Univ. Genève.

[2] Atkinson, E., Apple, A., Rawlins, K., & Omaki, A. (2016). Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology*, *6*, 167105.

[3] Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics*, *1*(3), 235–284.

[4] Hunag, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Ph.D.]. Massachusetts Institute of Technology.

[5] Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies—The case of Mandarin wh-in-situ dependency. *Journal of Memory and Language*, *84*, 139–166.

Is Cantonese *Dou*1 Changing?

Charles Lok
The Chinese University of Hong Kong
cheukfunglok@cuhk.edu.hk

This pilot study investigates whether the use of *dou1* in Cantonese changed under Mandarin influence, e.g., whether younger Cantonese speakers start to use *dou1* as a universal quantifier (UQ) without any preceding indicator or context.

While *dou* can act as a quantifier alone in Mandarin, many studies have focused on its relationship with the preceding D-quantifiers, e.g., *mei-ge*. Yip (2021) argues that *mei-ge* bears an interpretable but unvalued feature such that it requires agreement with *dou* in order to be valued. In Cantonese, as pointed out by Hsu & Lee (2020), *dou1* is generally interpreted as 'also'. It will have a UQ reading only when preceded/followed by a quantifier (e.g., *mui5go3*, *cyun4bou6*, doubled Cl, *saai3*), a wh-word, or given a context. Therefore, UQ *dou1* generally requires a quantifier as an indicator in Cantonese.

However, in recent years, it is observed that there is a trend of using bare *dou1* as UQ in Hong Kong Cantonese. (1) is an example from an online forum:

(1) Ngo5dei6 dou1 ng6wui6-zo2 mei5si1
We DOU misunderstand-ASP Messi
'We all misunderstood Messi.' (LIHKG, 2024)

A survey is conducted to investigate whether Cantonese speakers start to interpret *dou1* without any indicator (hereafter bare *dou1*) as a UQ. Assuming younger Cantonese speakers have more exposure to Mandarin at a younger age, it is predicted that more young Cantonese speakers will interpret bare *dou1* as UQ.

148 respondents completed the questionnaire. Their ages range from 12 to 71. Of all respondents, only 3.38% interpreted bare *dou1* as UQ when a bare noun subject is quantified. When a preceding collective quantifier *di1* is added before the subject, 19.6% have a UQ reading. 18.2% interpreted bare *dou1* as UQ when the (topicalized) object is quantified. When a preceding *di1* is added, 34.5% have a UQ reading. All respondents who have a UQ reading of bare *dou1* are exposed to Mandarin before age 6, supporting our hypothesis that this effect is a result of Mandarin influence. However, the linear logistic regression model shows that a significant effect of age is not found in the interpretation of bare *dou1* (e.g., in topicalized object construction) (β = -.005, *SE* = .01, *z* = -.34, *p* = .73).

To conclude, although the results support that interpreting bare *dou1* as UQ is a result of early exposure to Mandarin, it is not found that younger Cantonese speakers tend to interpret bare *dou1* as UQ more. Most respondents still interpret bare *dou1* as 'also', regardless of age. This could, of course, imply that *dou1* has not changed as expected in young Cantonese speakers. Nevertheless, this result can also be explained by various reasons regarding the limitations of this pilot study, e.g., design of questionnaire, stimuli used, sampling of participants etc. For example, a considerable number of respondents reported that they could guess the purpose of the questionnaire, which might lead to response bias. Moreover, Mandarin influence on the reading of *dou1* might be shown in constructions not used in the current questionnaire. Future studies can use more indirect methods and other sentence structures to test if *dou1* is undergoing a change in Cantonese.

Heritage Cantonese learner linguistic identities: Navigating cultural preservation and language ideologies in multicultural Canada

Raymond Pai raymond.pai@ubc.ca The University of British Columbia

In Canada's multicultural linguistic landscape, Cantonese, a historically significant Chinese language variety, faces increasing pressures in an environment lacking learning resources and institutional support (Xiao, 1998; Yu, 2013; Yu & Chan, 2017). This study investigates how heritage Cantonese speakers negotiate their linguistic and cultural identities amid prevalent societal language ideologies and the challenges of navigating a predominantly English-speaking environment. Focusing on heritage Cantonese adult learners, the research explores the strategies they employ to foster linguistic continuity within their families, through academic institutions, and in various communities as well as how these efforts intersect with broader cultural and linguistic expectations.

Using qualitative data from a larger research project, this study explores the learning trajectories of adult Cantonese-speaking participants across Canada recruited from large-scale surveys followed by semi-structured interviews to gain comprehensive insights into language learning and maintenance practices. Surveys reveal general trends in language use, while in-depth focus interviews with ten students and self-learners illuminate personal perspectives on language identity and the nuances of cultural preservation.

Preliminary findings through reflective thematic data analysis highlight the proactive strategies heritage Cantonese learners in Canada use to enable bilingualism, including enrollment in accredited university programs or heritage language schools, the self-establishment of Cantonese-speaking environments, and the integration of Cantonese media. However, their efforts face challenges, such as the dominance of English in schools, peer influences, and the societal devaluation of heritage languages, which can lead to internalized negative attitudes within the diasporic communities.

This research underscores the complex balancing act heritage Cantonese learners face between cultural preservation and assimilation. It emphasizes the importance of community support and targeted policy efforts to sustain multilingualism and cultural diversity. The findings contribute to discussions on language preservation, identity, and multiculturalism, providing valuable insights for educators, policymakers, and community organizations striving for a more inclusive linguistic landscape in Canada as well as Cantonese pedagogy in language education.

Tone Change of Disyllabic English Names in Hong Kong Cantonese
MAK Pui Ying (Jolene), M.A. dissertation

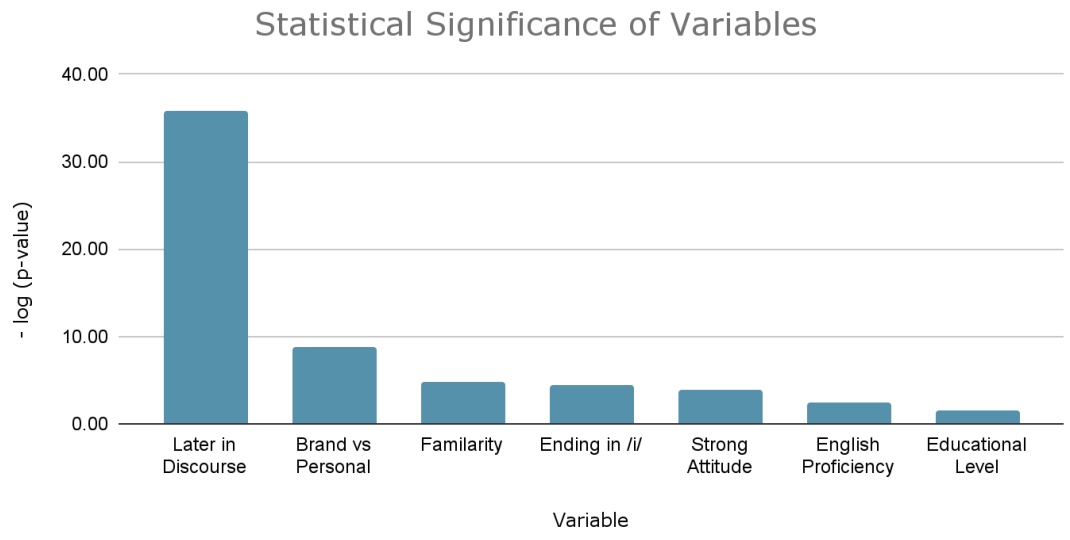
Abstract
Tone change is a common derivative process in Cantonese. Morphologically, it can transform the grammatical category of the word or indicate diminution. Sociolinguistic studies have also shown it is more prevalent in casual speech. Nonetheless, existing research on tone change in Cantonese has mostly focused on formulating phonological theories which govern either native Cantonese words, or orthographically adapted loanwords from English.

In contrast, the present study takes a descriptive and empirical approach to a specific tone change pattern in disyllabic English proper names, as spoken in Hong Kong Cantonese. To illustrate, 'Winnie' (/ˈwɪni/) has a falling pitch contour in English, but may be realised as win^{1}ni^{2} in Cantonese (with a high-rising tone on the last syllable), rather than win^{1}ni^{4} (with a final low-falling tone).

This study examines what sociolinguistic and discourse factors contribute to this change, or the lack thereof. Audiotaped interviews were conducted with 28 Cantonese-English bilinguals to elicit their pronunciations of English brand and personal names, and their commentary on each name. Statistical analyses of the recordings revealed that tone change is more likely for brand names compared to personal names. Names ending in /i/ also undergo tone change more readily. The results support previous views that tone change and the tonal adaptation of loanwords in Cantonese are processes not uniformly applied to all words.

Pragmatic and sociolinguistic variables also influence tone change. For instance, a speaker's familiarity with a name, as well as a strongly positive or negative attitude towards it, leads to a higher likelihood of tone change. Furthermore, tone change becomes much more likely as discourse progresses. Meanwhile, higher education levels and English proficiency correlate with lower likelihoods of tone change.

The following chart summarises the factors influencing this tone change phenomenon. A larger value on the vertical axis representing a more statistically significant effect.



由細唱到大——香港小學粵語校歌的樂調與字調

李佳妮（s1137847@s.eduhk.hk）、張凌*（zhangl@eduhk.hk）
香港教育大學

摘要

校歌是傳承校園文化的重要載體，是學生們從小唱到大的青春記憶。本文立足於語言學角度，從細處分析香港小學粵語校歌的粵語字調和校歌樂調的配合情況，從大處可見語言學的字調（lexical tone）和音樂樂調（musical tune）配合的類型格局。

本文通過窮盡式的網絡搜索，從全港 548 所小學中，搜集和整理出 243 首中文校歌，再經過仔細甄別，只對有真人演唱的版本分析，整理出 132 首粵語校歌。其中有相當一部分校歌不合字調。不合調的校歌，聲調的分佈是散亂無規律的，因此暫不作探討。值得注意的是，在基本合字調的粵語校歌裡，還有一些歌詞字眼是不合字調的，本文進一步研究，探討不合調的情況是否和粵語字調的類型特征有關。

過往的研究認為一般粵語歌需要合調，六個字調在填詞時從低到高可分為“0243”四個音高域。本文以演唱時歌詞基本合字調的 54 首粵語校歌為研究對象，統計香港小學粵語校歌演唱中不合字調的音節，初步整理出「香港小學粵語校歌不合調字表」。再對該字表進行深入分析，詳細總結出不合調音節的具體調值變化狀況，以及它們在句中所處位置（句首/句中/句末），並歸納其變調的語音規律。

從研究結果可以看出，在粵語校歌的變調情況中，「陽平調變陰平調」是出現頻率最高的聲調改變組合，出現了 62 次，佔總變調組合的 29.38%，其次是「陰去調變陰平調」、「陽去調變陰平調」。多個聲調易在唱時變為陰平，可能因為陰平調的調型為高平，容易與樂調高昂時的音高表達一致。另外，研究發現「陽平調」在校歌中也是最容易被改變的原始聲調。究其原因，陽平調音高最低，在聽感上呈現出明顯的低降特徵，當其不符合校歌高昂情緒表達時就會改為比它音高更高的其它聲調。此外，變調情況多發生在樂句末尾的音節位置，也因校歌旋律多傾向在句末使用較高昂的音高。

因此，整體而言，比起字調樂調的協音，校歌更注重歌詞所蘊含的意義與其所帶來的正面影響。此外，校歌作為特殊的音樂形式，篇幅一般較短，往往會譜寫音調較高的旋律來塑造更明亮的形象，因此會犧牲低沉的陽平調來確保高昂情緒的表達，也偏好使用高昂舒展的陰平調來替代其它聲調，兩者皆反映出校歌這一音樂體裁在創作上的功能性考量。

The Structural Organisation of Cantonese Sentence-Final Particles: A Minimalist Perspective

Hoi Ming CHAN
University of Cambridge
ernestchanhm@gmail.com

Previous accounts of Cantonese sentence-final particles (SFPs) often overlook how the Final-Over-Final Condition (FOFC) (Sheehan et al., 2017) fits into their analysis; this is particularly important as FOFC is considered to be a strong tendency or even a universal constraint under the Minimalist framework. Since SFPs are often assumed to be realisations of head-final C or T in the literature (Cheng, 2023; Law, 2004), given that they dominate head-initial vPs in their sentence-final position, it may result in a final-over-initial configuration and, hence, a potential violation of FOFC.

Through considering the pragmatic properties of a list of seven different SFPs (including *tim1*, *ge3*, *laa3*, *wo3*, etc.), I propose that there are four structural positions that host SFPs, namely the specifiers of (i) vP, (ii) the low Speech-Act Projection (SAP) in the vP domain (cf. Speas & Tenny, 2003; see also Biberauer, 2018), (iii) CP, and (iv) the high SAP in the CP domain.

Considering Sybesma and Li's (2007) dissection of SFPs, I argue that most SFPs have a base form and other tonal variants, which show different scope-related properties. It hence suggests that their interpretations may differ based on the syntactic positions they are in, and are hence multifunctional items. In line with Duffield's (2013) analysis of Vietnamese particles, I argue that they are semantically underspecified and are acategorial in nature. Due to their lack of a category label, they are therefore adjuncts and not heads; following Sheehan et al. (2017), they do not constitute the Extended Projection and FOFC is not violated.

As phrases with the [+V] feature are presumably specified for head-initiality, SFPs are spelled out finally in an order mirroring the head-complement order, to the right of the phrases specified for [+V] in the Extended Projection. This proposal also attempts to account for the clustering restrictions of SFPs, through considering a semantic hierarchy of features: particles with more semantic features are spelled out before ones with less, and particles in complementary distribution have an identical set of features. I argue for this to be a necessary postulation for sentences to make semantic sense.

Analysing SFPs with respect to FOFC shows that FOFC should not be treated as an exception in Cantonese (*pace* Cheng & Tang, 2022). Taking also into account the rich SFP inventory of Cantonese, the present study serves as preliminary evidence that SFPs in any given language are FOFC-compliant.

References

Biberauer, T. (2018, June 21). *Peripheral significance: a phasal perspective on the grammaticalisation of speaker perspective* [Conference presentation]. Diachronic Generative Syntax 20, York.

Cheng, S.-P. (2023). Cantonese Sentence-final Particles Pertaining to the TP Domain. *Current Research in Chinese Linguistics*, *102*(1), 81–100.

Cheng, S.-P., & Tang, S.-W. (2022). Syntax of Sentence-final Particles in Chinese. In C.-R. Huang, Y.-H. Lin, I.-H. Chen, & Y.-Y. Hsu (Eds.), *The Cambridge Handbook of Chinese Linguistics* (pp. 578–596). Cambridge: Cambridge University Press.

Duffield, N. (2013, May 11–12). *Minimalism and Semantic Syntax: Interpreting Multifunctionality in Vietnamese* [Talk]. Hanoi.

Law, Y. K. A. (2004). *Sentence-final focus particles in Cantonese* (Publication No. U643064) [Doctoral dissertation, University College London]. UCL Discovery.

Sheehan, M., Biberauer, T., Roberts, I. & Holmberg, A. (2017). *The Final-Over-Final Condition*. MIT Press.

Speas, P. & Tenny, C. L. (2003). Configurational properties of point of view roles. In A. M. Di Sciullo (Ed.), *Asymmetry in Grammar: Volume 1: Syntax and semantics* (pp. 315–344). John Benjamins.

Sybesma, R., & Li, B. (2007). The dissection and structural mapping of Cantonese sentence final particles. *Lingua*, *117*(10), 1739–1783.

粵語進行體「喺度」的功能及身份探析

陳子軒
香港中文大學
1155176404@link.cuhk.edu.hk

香港粵語（下稱「粵語」）通常以動詞後綴表達句子體貌，但進行體除「緊」外亦能在動詞前加上「喺度」表示。過往文獻雖然有留意到動詞前「喺度」標示進行體的功能，惟目前對其討論不多。本文將從句法及語義的角度分析此「喺度」於粵語中的功能及身份。

本文發現，「喺度」進行體用法存在獨有的限制及特性，如不能與完結類動詞搭配（見（1）），不能被焦點選擇（見（2））。又，「喺度」能接納動作的起點（見（3））、以及重覆事件（見（4））等。這些限制與特性皆不見於粵語中另一個用以表達進行體的後綴「緊」。

（1） a. *佢喺度食個蘋果。
　　 b. 佢食緊個蘋果。
（2） a. 佢喺度見學生㗎。（「喺度」只能解作處所）
　　 b. 佢見緊學生㗎。
（3） a. 佢今朝畀老師鬧完之後就喺度喊。
　　 b. *佢今朝畀老師鬧完之後就喊緊。
（4） a. 佢係嘅喺度跑。
　　 b. *佢係嘅跑緊。

至於進行體「喺度」的身份問題，有文獻主張將其定性為體標記（Lam 2009；彭小川 2010）。這個做法下的「喺度」性質屬副詞，但基於「喺度」在不同測試下的表現，本文認為所有情況下的「喺度」其實皆為一由「喺」與「度」組成的處所短語，其基礎語義仍然為表達處所。粵語乃至漢語存在不少「框式結構」，其中有一類為「框式介詞」，由介詞與方位詞組成（劉丹青 2002）。本文推測，粵語中的「喺度」是由框式介詞「喺……度」通過刪減（omission）中心語的策略語法化而成。同時，人類語言中「空間」與「時間」兩個概念之關係緊密，處所表達本身就蘊含（entail）時間特性，能用在習慣句（habitual sentence）中。「喺度」語法化後變成一直指（deictic）短語，以說話者為中心點而將時間預設為「現在」，因此與無界（atelic）的動詞組合時才會展現進行體的語義。最後，本文會引入「萊辛巴赫-克萊因框架」（Reichenbach-Klein framework）及 Kracht（2002）對位置格短語的語義分析展示「喺度」語義的轉變過程。

本文期望能增進我們對「喺度」的認識，補足文獻中此方面的空白。又，漢語「在」、「在這兒」和其他方言中亦有不少表達與「喺度」相似，表現相近，說明處所短語或為標示時間手段之一。

參考資料（部分）

彭小川。2010。《廣州話助詞研究》。廣州：暨南大學出版社
劉丹青。2002。漢語中的框式介詞。《當代語言學》第 4 卷第 4 期：頁 241-253。
Lam, C. F. (2009). A comparison between the Cantonese aspect marker gan and haidou. Rivista di Gammatica Generativa 34, 87-113. Retrieved from: https://lingbuzz.com/j/rgg/2009/lam_rivista_2009.pdf
Kracht, M. (2002). On the semantics of locatives. Linguistics and Philosophy 25, 157-232.

粵語否定詞「未」的語義

劉海澄

香港中文大學

1155176668@link.cuhk.edu.hk

在香港粵語（下稱「粵語」）的否定詞中，現有研究對「未」的分析集中於句法，對其語義解釋大多僅以普通話或英語作對應。因此，本文旨在討論粵語謂語前否定標記「未」的語義，並跟「冇」及「唔」作比較。

首先，本文描述了「未」的用法特點，整合了「未」跟對動詞、形容詞、賓語和動詞後綴的選擇，及跟時間副詞共現的限制。動詞方面，按 Z.Vendler（1967）提出的四個事件情狀類別，除了個人層面的狀態類謂語不可跟「未」組合外，其他謂語事件類型均可跟「未」組合（例子見（1）-（4））。形容詞方面，僅有具備閉端的等級形容詞可跟「未」組合。（例子見（5））賓語方面，「未」只可選擇有定賓語。（例子見（6））在表體貌的動詞後綴方面，跟「未」組合的謂詞僅可選擇經歷體後綴「過」。（例子見（7））

（1） 我未食嗰碗飯。 （完結類謂語）
（2） 我未食飯。 （活動類謂語）
（3） 我未到山頂。 （達成類謂語）
（4） 小明未需要做功課。 （階段層面狀態類謂語）
（5） 小明未安全
（6） 小明未飲呢三支啤酒。
（7） 小明未做過功課。

語義方面，本文通過引入萊辛巴赫-克萊因框架（Reichenbach-Klein framework），「未」僅可否定帶高體貌視角的句子，會引入完成時間段（perfect time span）或預期時間段（prospective time span），並否定時間段上謂語事件的發生，有別於另一粵語否定詞「冇」否定話題時間上謂語事件的出現或發生。此外，「未」帶有語義假設，假設所否定的事件在話題時間後極有可能發生，惟此假設當「未」跟經歷體謂語組合時會被取消。

參考文獻（節錄）

Comrie, B. （1976）. *Aspect : an introduction to the study of verbal aspect and related problems*. Cambridge University Press.

Kearns, K. （2011）. *Semantics* （2nd ed.）. Palgrave Macmillan.

Matthews, S., & Yip, V. （2011）. *Cantonese : a comprehensive grammar* （2nd ed.）. Routledge.

鄧思穎:《粵語語法講義》，香港：商務印書館，2015 年。

Emoji 是一個詞嗎？—— 從傳統語法視角分析香港與內地社交平台的表情符號

林雅珊　　萬佳琪　　萬佳瑤　　陳健榮

香港樹仁大學

電郵：kwchan@hksyu.edu

摘要

本研究探討 Emoji 是否已逐漸成為中文書寫系統中的一部分，並提出將 Emoji 視為一種「新字」進行語法分析的可能性。理論上，本研究以傳統漢語語法為分析框架，探討 Emoji 是否具備「新字」的語法功能。Belcher (2023) 對數位語類的討論則提供了背景脈絡，說明在數位溝通日益多模態、互動性與流動性的語境下，重新檢視 Emoji 語法地位的必要性與時代意義。鑒於數位語類打破了傳統語言形式與功能的界線，Emoji 不再僅是附屬於語言的情感標記，而可能已具備詞語層級，以下舉例。

（1）「在女神眼中我真是一個🤡。」｜（2）「*在女神眼中我真是一個。」

句（1）成立。因為「🤡」在此充當表語，代替名詞「小丑」，填補量詞「個」後必需的語法位置。

句（2）不成立。因為刪除「🤡」後，量詞「個」後缺少名詞性賓語，導致句子結構殘缺、語義不明。對比可見，在例（1）中，「🤡」是句子不可或缺的語法成分，承擔明確語義功能，而不僅是裝飾符號。

（3）「我明天要考試了😰。」｜（4）「我明天要考試了。」

例（4）中，刪除「😰」後句子語法完整、核心語義不變。「😰」僅附加表達說話者的緊張情緒，作用類似語氣詞，未承擔核心句法成分。因此，此處的「😰」是可選的、非結構性的情緒附加符號，不屬於中文句法必需部分。

本文研究對象為例（1），而非例（3）。因此，為檢視 Emoji 在當代中文語境中的語法地位，我們蒐集了兩地社群媒體中的語料：香港部分透過連登於 2025 年 4 月 1 日至 5 月 14 日期間，隨機選取三則 like/unlike 超過 700 次的貼文留言，共收錄 36,251 則留言，其中帶 Emoji 的留言有 20785 則，使用了 499 種 Emoji；內地部分則透過抖音、小紅書、B 站與微信於 2025 年 5 月 14 日至 6 月 9 日間收集，共計 1,308 則含 Emoji 的語例。分析標準為：Emoji 若能取代漢字、其刪除會使句子語法結構不合法者，方納入討論。語法分析從三個層面展開：其一，從語素角度檢視 Emoji 是否具定位/不定位、自由/不自由特性；其二，從詞類角度探討 Emoji 是否能充當實詞（動、名、形、副）或虛詞（介、連）；其三，分析其是否能構成複合詞或短語結構。

經初步分析，Emoji 在內地的語法化現象顯著高於香港。雖香港語料總量龐大，但實際符合語法分析條件的 Emoj 僅 5 種，語例 5 條，總字數 107 字；而內地則收集到 159 種 Emoji，語例 1,308 條，總字數 11,429 字。這顯示 Emoji 在內地中文中的語法地位更加活躍、自由與創新，已超出情感輔助工具的功能。

本研究揭示 Emoji 正在介入中文語法架構，並在不同社會語境中展現出功能與用法的分化。Emoji 已具備部分「新字」的語法條件。這一現象挑戰了傳統語類與語言單位的界線，也為數位語言變異研究提供了新的切入點。

關鍵字：Emoji、數位語類、語法分析、比較研究、社交媒體

參考文獻（節選）

1. Belcher, D. D. (2023). Digital genres: What they are, what they do, and why we need to better understand them [J].*English for Specific Purposes*, (*70*). *pp*.33–43. DOI: https://doi.org/10.1016/j.esp.2022.11.003.

《英華分韻撮要》量詞研究

霍穎欣、陳以信
香港伍倫貢學院
yhuo@uow.edu.au, abrahamc@uow.edu.au

《英華分韻撮要》（A Tonic Dictionary of the Chinese Language in the Canton Dialect）是 19 世紀著名傳教士、漢學家衛三畏（Samuel Wells Williams）編撰的第一部粵語字典。該字典以《江湖尺牘分韻撮要合集》為底本，參照 Sir William Jones 提出的拼音方案，按字序編排，收錄漢字 7,800 餘個、詞例 29,000 餘條。

該字典收錄的量詞十分豐富，初步統計有 200 多個，涉及 600 餘條詞例。當中不乏現今已經不常用或消逝的量詞，例如「一吥泥」、「六刀紙」、「住埋一啱」等。部分有特定的修飾對象，如「叠」多用於植物（如「一叠樹葉」、「一叠花」、「一叠蕉」）。量詞與名詞之間，也呈現靈活的雙向對應關係：如「段」既見於「一段地」，也見於「一段事」，同一量詞可搭配多重對象；而修飾「樹」時，則可用「𣍐」、「根」、「條」、「篦」、「坎」、「碌」、「株」等不同量詞，呈現數個量詞配對一個名詞的現象。後者的搭配往往存在語義分工，如「一𣍐樹」似無特定的性質或類別限制，但「一坎樹」則用於描述受損害而生長遲緩、矮小的樹，具有形象意義的語用特點。

與現代粵語比較，可發現部分特定量詞已經逐漸簡化，例如字典中的「粗扇二持」、「一隻梳」、「一箍柴」、「一啱米」等，現代粵語均可用通用量詞「把」來表達。此外，部分量詞的語義及句法功能也產生變化。以「的」字爲例，在字典中可同時作為不定量詞（如「個的係乜人」、「呢的果生」）和表複數意義的助詞（如「快的閘埋」），但在現代粵語裏多已由「啲」所承擔，即「個啲係乜人」、「呢啲果生」、「快啲閘埋」。「的」和「啲」的關係也甚爲複雜，同時涉及語音演變和詞彙更替。

本文旨在從類別、搭配、句法、語義等多維角度，系統梳理該字典的量詞體系，整理現今已不常用或消逝的量詞，探討其源流、語義與使用情況，嘗試呈現十九世紀粵語量詞較完整的面貌。

參考文獻

Williams, S. Wells (1856). *A Tonic Dictionary of the Chinese Language in the Canton Dialect.* Canton: Office of The Chinese Repository.

廣東高校粵語播音主持專業學生粵語能力弱化成因研究

曾岑 廣州大學新聞與傳播學院

摘要：播音與主持藝術專業是一個融合新聞傳播學、戲劇影視學、語言學等多學科的應用型專業，是中國內地特有的、歸屬於藝術學一級學科的專業。1963年由北京廣播學院（現中國傳媒大學）開辦，距今已有超過60年的辦學歷程。

在廣東，粵語是粵港澳三地的通用語言和內部黏合劑（鄭可彤，林小榆，2021），亦以58.79%的使用率，成為粵港澳大灣區電臺播音語言使用最多的語言（王海蘭等，2020），在政治、文化和經濟多重因素驅動下，2005年，廣州大學首次在播音與主持藝術專業下設粵語播音方向，招收粵語播音專業學生並成建製培養（曾岑，2018）。20年來，累計為廣東輸送了500多名粵語播音專業畢業生。筆者2016年入職廣州大學，在一線工作中發現，近10年，廣東高校粵語播音專業學生的粵語能力出現了明顯的弱化現象，包括粵語聲調範疇混淆（Tonal Category Confusion）、粵語方言特征弱化（Dialectal Feature Erosion）、粵語文化詞匯缺失（Deficit in Culture-Specific Vocabulary）等，且日益顯著。

為深入解析這一現象的複雜成因機制，本研究採用混合研究方法，以廣東省內五所開設粵語播音專業高校的在校生及畢業生為研究對象，研究設計包含三部分：

1.教育民族志觀察：研究者持續嵌入研究對象的學習與生活環境（如課堂、實驗室、宿舍、社團活動場地等），進行為期4個月的非參與式觀察，系統記錄生生互動、語言實踐及校園語言生態，尤其關注粵語使用的真實場景、頻率及規範性。2.問卷調查：針對目標群體發放問卷，測量其家庭語言環境、粵語習得年齡與途徑、粵/普使用頻率與場景、語言認同與態度、對自身粵語聲韻調、辭彙、表達流利度及文化理解的自評、媒體接觸習慣等進行調查。3.焦點小組座談：依據初步觀察和問卷結果，組織若干組（每組6-8人）不同年級、背景（粵語母語者/非母語者）的粵語播音專業學生進行半結構化訪談，圍繞粵語能力弱化的個人體驗、歸因（生源、教學、環境）、學習困難、對大灣區粵語媒體生態的看法等展開討論，挖掘共有經驗和深層觀念。

數據分析採用三角驗證法，整合量化統計與質性資料，揭示粵語播音專業學生粵語能力弱化現象背後的多維成因：生源層面，粵語母語者占比從90%驟

降至不足一半；中小學階段過度側重普通話能力訓練，使用粵語的場景、機會大幅減少；教學層面，粵語師資不足導致核心課程粵語授課比例過低，且聲韻調訓練不足總課時5%；生態層面，校內推普政策壓縮粵語日均使用至0.7小時，大灣區廣播電視粵語內容占比僅42.83%（徐靜，2023），且粵語發音不標準現象越來越多，使學生缺少學習對象，缺失正音語境。

粵語播音專業學生原本應該是粵語發音準確性最高、粵語詞匯口語表達能力最強的群體，但現實告訴我們並非如此，由此也可窺視廣東地區青年群體的粵語能力衰變問題的嚴峻程度。如何通過教育幹預，為粵語保護和傳承提供可行的教學改革路徑，對全球式微方言或瀕危語言的傳承具有方法論啟示。

作者：曾岑，廣州大學新聞與傳播學院播音與主持藝術系教師，
zengcen@gzhu.edu.cn

參考文獻：

[1]鄭可彤　& 林小榆.(2021).內地粵播對提升港澳青年大灣區文化認同作用研究.新聞研究導刊,12(10),177-179.

[2]王海蘭,譚韻華,劉栩妍 & 詹嘉琪.(2020).粵港澳大灣區廣播語言使用調查.(eds.)語言生活皮書——中國語言生活狀況報告（2020）(pp.71-79).doi:10.26914/c.cnkihy.2020.068110.

[3]曾岑.(2018).影視劇粵語配音人才培養的困境與對策.新聞研究導刊,9(07),238+244.

[4]徐靜.(2023).數字時代媒介對地方語言的影響：廣州年輕一代的粵語使用與認同.當代青年研究,(04),14-25.

題目：粵語句末助詞“啊 aa3”和“喇 laa3” 在話語情境裏所起的作用

作者：謝國強（Tse Kwok Keung）
Affiliation: Adjunct lecturer, Nanyang Polytechnic, Singapore
Email: tse_kwok_keung@hotmail.com

粵語裏有大量的句末助詞，一般都認為是用來表達語氣，主要出現在對話的語境，很少在比較“正式”的場合出現（例如電視新聞報導），常見的有“啊 aa3“和”喇 laa3“。

按照已故法國語言學家 Antoine Culioli 的理論框架，在話語情境裏，話語的座標 Sit 由說話者 S 和他當前的時間 T 構成：Sit (S, T)，一個謂語關係 (predicative relation）<p>跟 Sit 掛鉤便成為話語的句子。<p>成就句子的基本意義，而<p>跟 Sit 掛鉤則讓句子的解讀牽涉某些複雜的範疇，如情態和體貌（Culioli 1999）。

說話者 S 也包含共話者：S (locutor, collocutor）。很多時候，共話者的角色是潛藏的、無為的、消極的，在這個情況下，說話者和共話者無分彼此。但如果共話者的角色被激活，那麼說話者和共話者之間可以出現複雜的相互調整（adjustment）。我們認為粵語的句末助詞就是圍繞著這種相互調整來起作用的。

在現實語境裏，<p>跟 Sit 掛鉤而成為句子。掛鉤的句子可以不帶句末助詞（句末助詞“零”），也可以帶句末助詞如“啊 aa3”或“喇 laa3”，我們用一些簡單的謂語關係來做例子，如“老李下個月退休“和“佢去咗廣州”：

句末助詞“零”	句末助詞“啊 aa3”	句末助詞“喇 laa3”
1. 老李下個月退休。	1a. 老李下個月退休啊。	1b. 老李下個月退休喇。
2. 佢去咗廣州。	2a. 佢去咗廣州啊。	2b. 佢去咗廣州喇。

在（1）和（2），<p>跟 Sit 掛鉤而得出的句子，基本上就是對<p>的內容作陳述；在（1a）和（2a）帶“啊”的句子，<p>跟 Sit 掛鉤，而同時共話者也被喚起，說話者似乎在跟共話者說：“<p>跟 Sit 掛鉤，而掛鉤這件事，你也是持份者/我們是同夥”；在（1b）和（2b）帶“喇”的句子，除了喚起共話者在掛鉤這個關係上的角色外，還表示 Sit 和<p>之間有一種“有待“（awaiting）的關係，說話者似乎在說：“作為 Sit 的 S，我們對<p>跟 Sit 掛鉤有期待（這種期待可以很微弱），而<p>果然跟 Sit 掛鉤”。

除了對“啊 aa3”和“喇 laa3”作橫向的比較，我們也可以對這兩個句末助詞作縱向的比較。根據以粵語為母語人士的語言直覺，“啊 aa3“跟另外兩個句末助詞“吖 aa1”和“呀 aa4”可以相提並論，而“喇 laa3”則跟“啦 laa1”和“喇 laa4”可以相提並論（有學者認為“喇 laa3”和“啦 laa1”在意義上確實有關連，參看鄧思穎 2015，阮氏紅貴 2021）。這些同音不同調的句末助詞在語氣的表達上雖然各異，但在意義上有相通之處。從這個角度來看，我們認為可以把這兩組一共六個的句末助詞放在同一問題域來研究。

主要參考文獻：

Culioli, Antoine. 1999, Pour une linguistique de l’énonciation, Tome 3. Paris: Éditions Ophrys.
鄧思穎. 2015, 粵語語法講義. 香港:商務印書館.
阮氏紅貴. 2021, 中國語文通訊, 第 100 卷第 2 期. 99-119.

粵語玉林話的兩個否定詞

郁孫豪（香港科技大學）　晨曦（廣西大學）

玉林位於廣西東南，根據《中國語言地圖集》，通行于玉林市區的土白話屬粵語勾漏片。從語法角度看，其否定詞系統與其他粵方言存在較大差異：廣州話「唔」的功能在玉林話裡大致由兩個否定詞承擔，「冇」[mau^{24}]和「覓$^{=}$」[mɛk^{22}]（張敏 2001、2019、周烈婷 2002）：

（1）玉林：a. 佢**冇**$_{不}$是學生，是老師。　b. 佢**覓**$^{=}$[mɛk^{22}]$_{不}$是學生，是老師。

其中，以「冇」對應廣州話「唔$_{不}$」的現象在珠三角外的粵方言中屢見不鮮，如下表[1]：

方言點	廣州	廣東連山$_{布田}$	廣東鬱南$_{平台}$	廣東信宜	廣東吳川	廣西北海	廣西平南
「不」的說法	唔[m̩^{21}]	冇[mo^{33}]	冇[mau^{13}]	冇[mau^{23}]	冇[mɔɒu^{33}]	冇[mou^{13}]	冇[mɐu^{23}]

同樣，「覓$^{=}$」雖不見於廣州，但在粵西也能找到相似的形式。如封開（南豐）就是用□[niɐk^{55}]來表達「唔」的意思：「唔好」即[niɐk$^{55\text{-}33}$ hɔ^{52}]，「唔使」即[niɐk$^{55\text{-}33}$ siə^{52}]（詹伯慧、張日昇主編 1998：665）。所以，玉林話的情況對西部粵語而言有一定代表性。其問題在於：玉林話為何需要兩個功能如此接近的否定詞？西部粵語這兩類否定詞的來源又有何異同？

首先，我們的觀察是：玉林話的「覓$^{=}$」區別於「冇」，在語義上，前者的否定意義基於由析取關係（disjunction）構成的梯級（scalar）「p ⊨ p∨q」，故而不完全等同於廣州話的「唔」，相對而言，「冇」則是玉林話中無標記的基本否定詞。具體地，「覓$^{=}$」往往伴隨著辯駁語氣：如（1a）可以是平鋪直敘的，而（1b）一般有「對方認為他是學生（而這不對）」的意味。根據該語境，可構造以下衍推關係：「他是學生」⊨「他是學生」∨「他是老師」∨「他是…」，「覓$^{=}$」並非單純強調命題為假，而是預設了上述梯級，強調「他是學生」為假，由此得出涵義「他是其他身份」，形成糾正意味。如（2）的「覓$^{=}$」字句就往往伴有申辯、反駁語氣：

（2）玉林：我**覓**$^{=}$$_{不}$識佢！（我才不認識他呢！）　（張敏 2019）

總之，玉林的「覓$^{=}$」與「冇$_{不}$」並非雷同，前者是一類梯級否定詞，故而帶有特殊涵義。這類複雜否定詞也見於老湘語，「冇$_{沒}$」亦可當「唔$_{不}$」用（3），同樣專用於糾正的語境（4）：

（3）新化（湘語）：我**冇**/ **唔**$_{不}$是長沙人。　（羅昕如 1998：269）

（4）邵東（湘語）：**冇**/ #**唔**是[我]$_{CF}$昨天做個$_{的}$，是渠！　（林素娥 2016：253）

比較可知：玉林的「覓$^{=}$」與老湘語的存在否定詞「冇」很像，都能作為「梯級否定詞」進而在一些環境下解作「不」。而粵方言內的證據也顯示，前者確實可能來自存在否定詞：

（5）賀州（粵語）：箇堵$_{這裡}$□[mak^{5}]$_{沒有}$人。　（陳小燕 2007：326）

賀州話對應廣州話「冇$_{沒}$」的形式是□[mak^{5}]（一說其中包含「得」[lak^{5}]），可能是玉林話「覓$^{=}$」的前身。此外，用於否定事物的存在時，「覓$^{=}$」的辯駁語氣就消失了，相當於「沒」：

（6）玉林：a. 老兄**冇**$_{不}$**有**錢。　≈　b. 老兄**覓**$^{=}$[mɛk^{22}]**有**$_{沒有}$錢。　（張敏 2019）

上例 a、b 可自由替換，表明「覓$^{=}$」兼有存在否定功能，暗示梯級否定或許是後續演變。根據徐毅發（2017），「存在否定詞變為基本否定算子」在漢語方言中存在多個路徑。因此，假如賀州的□$_{沒有}$[mak^{5}]和封開的□$_{不}$[niɐk^{55}]之間存在此類演化關係的話，玉林的「覓$^{=}$」以及老湘語的「冇」（例 3、4）作為（預設析取關係的）梯級否定詞或許就是中間階段。

另一方面，西部粵語的「冇$_{不}$」（1a）也一般認為來自存在否定詞（張敏 2002）。類似地，可能也經歷了「梯級否定詞」。不過，結合肇慶的情況看，應是預設了合取關係（conjunction）：

（6）肇慶（粵語）：我（平時）**冇**/ **唔**$_{不}$打籃球個$_{的}$。　（徐毅發 2017：2）

肇慶話的「冇」在否定慣常事件時可解作「不」。於是可構造一種基於「條件」的衍推：「平時」→「打籃球」⊨「平時」∧「體育課」∧…→「打籃球」。肇慶話的「冇$_{不}$」與玉林話的「覓$^{=}$」都預設了梯級，雖具體類型不同，但都能視作西部粵語存在-否定演化的中間階段。

[1] 語料來自詹伯慧、張日昇主編（1994：672、1998：675）、羅康寧（1987：168）、李健（2014：177）、陳曉錦、陳滔（2005：407）、劉春梅（2023：374）。

Towards Glyphon Frequencies in Written Cantonese

Jon CHUI
jon@canto.hk

Chinese glyph often maps to multiple readings. In Cantonese, 行, depending on the context, is read as hang$_{\searrow 4}$, haang$_{\searrow 4}$, hang$_{-6}$, hong$^{\nearrow 2}$, hong$_{\searrow 4}$, or hong$_{-6}$. Usage of the multiple readings may be mixed within the same sentence, *e.g.*, jau$_{\nearrow 5}$ **hang**$_{\searrow 4}$ jan$_{\searrow 4}$ **haang**$_{\searrow 4}$ jap$_{-6}$ naam$_{\searrow 4}$ bak^{-1} **hong**$^{\nearrow 2}$ gaak$_{-3}$ lei$_{\searrow 4}$ gaan^{-1} ngan$_{\searrow 4}$ **hong**$_{\searrow 4}$ gam$_{-6}$ 有 行 人 行 入 南 北 行 隔 籬 間 銀 行 撳 cin$^{\nearrow 2}$ zou$_{-6}$ sin$_{-6}$ **hang**$_{-6}$ 錢 做 善 行. Relative frequencies of the **glyphons** (a portmanteau of *glyph-* and *-phone*) in different text types is a useful metric for applied linguistics and pedagogy, such as designing a character learning progression. This, however, had not been available for written Cantonese.

The grapheme-to-phoneme (G2P) engine from the Cantonese Font 3.3 is a deterministic, rule-based method that attains > 99% accuracy in Jyutping assignment for both written Cantonese and standard written Chinese; it supports extraction after manual correction. I applied this G2P process to reveal the glyphons from the graphemes in several written Cantonese genre:

1. Thomas Tsoi translation of the *Animal Farm* (45,949 characters, Jyutping manually corrected),
2. HKCanCor (230,000 characters radio transcript), and
3. article texts in the 2023 Yue Wikipedia (23 million characters in total)

A list of ranked glyphons for each work was compiled and compared, and the G2P results can be further validated by comparing our extract with the manually assigned Jyutping from HKCanCor. Specific ranking of glyphons differs between the three genre of works (special themed novel, radio transcript, general text). Overall glyphon usage frequencies approximates a Power Law distribution.

A practical application is demonstrated in constructing progressive learner's editions of the *Animal Farm,* where I selectively occluded Jyutping annotations for the top glyphons of this work, presumed learnt by the reader. Occlusion of the top 1% glyphons results in 85% reduction of Jyutping annotations; removal of visual noise helps focus the learner's attention towards glyphons yet to be learned.

大型語言模型（LLM）粵語知識盲點分析

劉擇明 (lchaakming@eduhk.hk)　**香港教育大學**

大型語言模型（Large Language Model，下稱 LLM）已經可以處理各種英文嘅自然語言處理任務，諸如拼音轉寫（G2P）、同音字判斷、押韻判斷、句子續寫、語法判斷、撮寫、翻譯等等；但係去到粵語呢類欠缺標準化、開源材料貧乏嘅低資源語言，經常會連基本任務都無法完成。Cheng et al. (2025) HKCanto-Eval 評測（Benchmark）指出截至 2025 年初，LLM 處理粵語嗰陣非常依賴繁體中文數據、普通話（官話）知識，而缺乏粵語語言學嘅專門知識。無論係開放權重（open-weight）模型抑或專有（proprietary）模型，就算可以接收粵語指令，都冇辦法完成涉及字音、字形嘅任務。

本報告集中探討 HKCanto-Eval 未公開嘅模型測試數據入面各模型嘅偏誤。

語音能力
共 100 題
平均 23.4%

除咗 GPT4o，各個模型都未有粵語音節結構、押韻、平仄嘅基本認識，亦受到不同程度受到強烈嘅普通話干擾。

同音字判斷 Phon#2 作 zok3
A 做 (19/24) B 桌 (0/24) C 昨 (3/24) D 左 (0/24) E 皆非 (2/24)

文字處理
共 100 題
平均 43.8%

大多數模型揀得啱粵文用字，但係未可以用繁體香港字形分析部件。

粵字填充 Char#75　　我而家__屋企
A 系 (1/24) B 喺 (17/24) C 係 (5/24) D 起 (0/24) E 皆非(1/24)

形似判斷 Char#1　　方
A 正 (12/24) B 形 (2/24) C 万 (7/24) D 放 (3/24) E 皆非 (0/24)

拼音轉寫
共 150 題
錯字率 (CER)
55.1%

所有模型都明白 G2P 嘅字母加數字聲調嘅格式要求，多數模型有折合音、受普通話、人名地名、其他方案干擾嘅字音，略有偏差嘅音，多音字錯讀。

G2P#2 鉛係一種重金屬，
yan6 sei6 jyun1 zaam1 zung1 jin6 shu1, **Qwen 7B**
yim4 hai6 jat1 chong2 cung4 gam1 jek6 **Llama 8B**
jin4 hai6 jat1 zung2 cung5 gam1 zuk6 **Gemini 1.5 Pro**
jyun4 hai6 jat1 zung2 cung5 gam1 suk6 **GPT-4o, CLLM 0.5 34B, …**

翻譯
共 20 題
平均 87.8%

可以做到基本要求，但係各別粵語特色詞彙會有空隙。

Tran_YueEng#18　　間酒店有啲隔涉喎，等咗三個字都冇的士肯入嚟。
The hotel has some **reservations**,
and taxis haven't been willing to come in for the past **three characters**. **Yi 9B**
There's something **off about** this hotel,
even after waiting for **three minutes** no taxi will stop. **Qwen 7B**
The hotel is quite **remote**,
and we've been waiting for a taxi for over **15 minutes** with no luck. **CLLM 0.5 34B**

整體而言， LLM 處理粵語時出現嘅偏誤反映咗靠估嘅限制。各模型明顯缺少咗粵語語言學嘅結構性資訊同粵文書寫材料。要有效處理粵語以至其他低資源漢語，呢個問題必須正面解決。

Cheng, T. C., Cheng, C. S., Lau, C. M., Lam, E. T. H., Wong, C. Y., Yu, H. O., & Chong, C. H. (2025). HKCanto-Eval: A Benchmark for Evaluating Cantonese Language Understanding and Cultural Comprehension in LLMs. *CoNLL-2025*.