




The 5th Pan-Cantonese Offline Meeting of
Lingnaam Jyutjam and
the 2nd Forum on Languages in South China
嶺南粵音第五屆泛粵線下交流會
暨第二屆嶺南語言討論會

July 24, 2022 (Sunday)

2022 年 7 月 24 日 (週日)

Program and Abstracts
程序表及論文摘要

Organized by
the  Jyutdict Editorial Board of
Lingnaam Jyutjam

**The 5th Pan-Cantonese Offline Meeting of
Lingnaam Jyutjam and
the 2nd Forum on Languages in South China**
July 24, 2022 (Sunday) 14:30 pm – 17:35 pm (UTC+8)

Zoom Information

<https://tum-conf.zoom.us/j/69201308733>

Meeting ID: 692 0130 8733

Passcode: 250714

Time	Presenter & Paper title
14:30 – 14:45	Opening
14:45 – 15:15	指南：Introduction to Train a Cantonese BERT Base Model
15:15 – 15:45	陳洞龔：論四邑方言知章有別嘅層次
15:45 – 16:00	Break
16:00 – 16:30	白朔：電白閩語的前《切韻》層次
16:30 – 17:00	以成，瞰明：部分粵語詞讀音的原始形式
17:00 – 17:30	瞰明：桂平木根鎮方言音系簡述及心母的演變
17:30 – 17:35	Closing

* 30 minutes for each report (5 minutes for brief description and 25 minutes for detailed explanation)

Introduction to Train a Cantonese BERT Base Model

指南

ABSTRACT

This article uses HKCancor, Mozilla Common Voice, Cantonese Wikipedia, and Cantonese corpora collected from the Internet to train a Cantonese base BERT model based on transfer learning from a Traditional Chinese language model trained by ckiplab/BERT-base-Chinese. BERT was a language model introduced by Google in 2018^[1], trained by a large corpus. Using BERT as a base model in downstream NLP tasks can significantly shorten the training time, and achieve good results. This article briefly introduced the process of training the BERT model, including data parsing, data processing, and finding what the breakthrough the Cantonese BERT model can bring to the Cantonese NLP downstream NLP task.

References

- [1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. doi:10.48550/ARXIV.1810.04805.

論四邑方言知章有別嘅層次

陳洞龔

摘要

本文通過方言調查，發現粵語四邑方言入邊存在中古效山咸流臻深攝三四等字嘅中古知章組聲母同影以雲聲母條件下韻母不同於其它聲母條件嘅現象，並進一步發現咗一個嘅中古效山咸流臻攝三四等字範圍區分中古知組同章組嘅層次。

本文發現知組嘅後接嘅韻母大部份條件下與章組嘅表現相同，但係有啲又同端組嘅表現一致，即係同時帶高化同低裂化傾向。此外中古知組二等同莊組未發現區分。

本文根據呢個發現重新構擬咗四邑方言嘅祖語，將章組擬作 j 聲母同部位嘅 $tɕ$ $tɕ^h$ $ɕ$ (dz z)；將知組三等擬作齶顎音 t t^h (d)。相關成果發表嘅知乎文章《台開江 深邑歸源音系》。

關鍵詞：四邑方言；知組；章組

電白閩語的前《切韻》層次

白朔

摘要

電白閩語作為沿海閩語的子方言，在語音表現上具備《切韻》所不具備的層次，可以通過將方言材料與上古漢語、原始閩語比較來討論這部分層次的特殊性。

本文通過比較分析電白閩語特殊字音（如「年」讀 hi2，「試」讀 ts^hi4，「書」讀 tsu1 等）在上古漢語和原始閩語中的對應得出其從原始閩語乃至早期的閩南語發展而來的音變路徑，並間中印證了祖語的清響音模型與部分傳世諧聲的構擬；兼而印證了黎話、海話在不同時期進入電白，但與潮汕、雷瓊、泉漳等地閩語具有共同的祖語。

本文採用了反映較新觀點的上古漢語構擬體系，有人亦稱之為「最簡化上古漢語」（minimal old Chinese），避免疊床架屋的聲母構擬。

關鍵詞：電白閩語；上古漢語；原始閩語

部分粵語詞讀音的原始形式

以成，噉明

*本文係附屬於正在進行的項目「原始粵語音系構擬」（噉明，以成）的研究進展報告

摘要

「泛粵典」是嶺南粵音·泛粵典編委會在粵語語言資料建設的重要成果之一。其中，富有粵語特色的單字詞會被收錄入「泛粵字表」（下稱「字表」）中，現今條目已逾 4000 條，而有效填寫超千條的方言點有 22 個。

本文主要利用字表中各方言點數據，結合鄰近民族語言對比，以及現有音系構擬成果，試構擬百餘條粵語單音節詞讀音的原始形式，並標註出可能的詞源及演化路徑。構擬完成後，依據詞源不同、詞語間關係親疏等來分組講解。字表中新材料的共時對比支撐，不僅可能有助於揭開今日某些粵語詞的原始面貌，還可能為原始粵語的音系構擬補充證據。

附構擬條目組註解「*sj.laap9（燂 1，燂 2，蠟）」：

- 以成：百越底層詞^[1]，或與「閃」同源。
 - 噉明：**/m̩lAp/。其中，**/m̩l/ > **/mj/ > **/n̩/。ml- mj- 讀法見侗台語。
/n̩/ > /ɕ/ 的路徑可由侗水語窺探^[2]：/n̩/ [佗佬] > /ɕ/ ≈ /j/ [莫話] > /ɕ/。
但須注意，清鼻音可能是更早於原始粵語的讀法。
 - 以成：據噉明，事實可能為：
 - **/m̩lap/9 > **/n̩ap/9 > **/ɕap/9 > *sjaap9 > saap9 燂 1 燂 2；
 - **/m̩lap/9 > **/n̩ap/9 > **/jap/9 > *jaap9 = jaap9 燂 1 燂 2；
 - **/m̩lap/9 > *laap7 > laap8 蠟。
- 綜上，可擬為 *sj.laap9。

關鍵詞：泛粵典；泛粵字表；粵語詞；原始粵語；讀音

[1] 余靄芹. 粵音構擬之二：聲母[M]. 2006.

[2] 龍海燕. 高洋侗話的鼻化擦音[J]. 民族語文, 2019(2): 63-68.

桂平木根鎮方言音系簡述及心母的演變

噉明

摘要

本演講主要分兩部分，第一是簡單介紹廣西壯族自治區桂平市木根鎮粵方言的音系並和鄰近的社步鎮、麻垌鎮、尋旺鎮方言做比，總結出其方言的音系特點，如：

1. 古全濁聲母清化後，今一概讀不送氣音；
2. 精組聲母全部讀 s；
3. 梗開三四和曾開三讀 $\text{eiŋ} / \text{eik}$ ，與尋旺鎮同；
4. 咸開一、宕開一等主元音為 $\text{ø}[\text{ö}]$ ，與社步、麻垌、尋旺同，但山開一主元音為 ɔ ；
5. 一般粵語中的「短 $\text{a}[\text{e}]$ 」音位被「短 $\text{ɔ}[\text{e}]$ 」替代；

第二是提出一個心母演變的假說。今兩廣粵方言中，不乏心母有讀 ɬ 者。但往前鮮有學者提出從中古漢語 *s 到 ɬ 音變過程。木根方言中 s 的調音部位為舌尖-齒齦（或可記為 ɕ ），與廣州話的 ɕ ，普通話的 ɕ 聽感迥殊。結合木根鎮方言其它粵方言的情況，筆者認為這個讀音是心母演變過程上的重要一環，具體演變鍊為「中古心母 > 清舌尖-齒齦有擦音 ɕ > 清齒齦無擦音 θ > 清齒近音 $\text{i}!$ > 邊擦音 ɬ 」。

去擦化這一步當發生在幫、端母內爆化之後；無擦音變清近音當發生在「原清響音派入陰調 > 濁響音陰調」之後。