

「廣東話達人挑戰」的公共數碼人文學反思

Public Digital Humanities Reflections on the “Cantonese Master Challenge”

黎奕葆¹、李梓明²、黃錦鵬³、祁美瑩⁴、劉擇明¹、黃學轅⁵、葉家輝⁶

¹香港教育大學; ²香港城市大學; ³香港中文大學; ⁴紐約市立大學; ⁵香港都會大學; ⁶耶魯大學

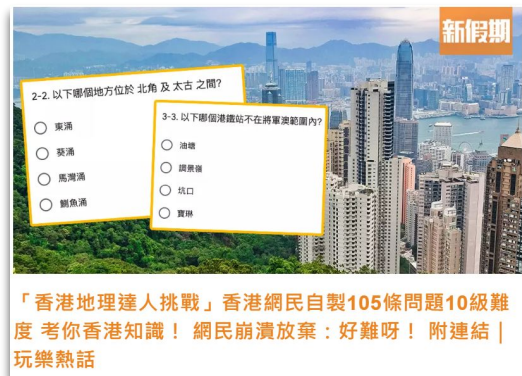
1. 「廣東話達人挑戰」
2. 公共數碼人文學
3. 「廣東話達人挑戰」作為一個公共數碼人文項目
4. 公眾需要與研究需要之間的矛盾與協調
5. 總結

1.「廣東話達人挑戰」

1.「廣東話達人挑戰」

緣起

- 於2022年3月，香港網絡上一個名為「香港地理達人挑戰」的 Google Forms 線上問答遊戲掀起熱潮



遊戲以高難度見稱，「超難」、「崩潰」等用詞皆見於報道標題之中

1.「廣東話達人挑戰」

緣起

- 其他不同的「達人挑戰」也在網絡上相繼湧現



730 **IG大熱！你係XX達人?!** **網上熱話**

7類達人測試整合
超過30種測試

MIRROR、港大挑戰都有

Music hub .hk

香港流行音樂達人挑戰 2022

一號
室：<https://forms.gle/L58rMVFWKm4E9tFH6>
←你在此

二號
室：<https://forms.gle/sUbT3qLmFqNnep2P8>

三號
室：<https://forms.gle/rvQXn.JoCZJu4TufM9>

本挑戰分為10級

The graphic includes two images: on the left, a hallway with a large calligraphic '香港大學' (HKU) sign on the wall; on the right, a group of people in colorful costumes performing a challenge.

主要是一般民眾興趣使然的創作，不過也有本地機構以此作為公關及教育大眾的手段

- 「香港九巴達人挑戰」
- 「香港歧視條例達人挑戰」

1.「廣東話達人挑戰」

- 「廣東話達人挑戰」就是在此熱潮中誕生

Ka-Wing Chan Kenith Chan Kevin Chan Casey Ho Mei-ying Ki Yik-Po Lai
Chaak-ming Lau Cindy Lau Tommy Tsz-Ming Lee Raymond Tse
Kam Pang Wong Oscar Wong Ka-Fai Yip Edward Yu Roy Yu May Zeng
(姓氏字母序)

- 2022年3月28日推出, 約一個半月內收到超過**40,000**份回應

1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言

廣東話達人挑戰

喂，你識唔識講廣東話？

根據政府統計處2019年嘅報告，喺全港6-65歲嘅人士裡面，有88.8%母語都係廣東話。個個都識講廣東話，唔通個個都了解廣東話咩？

我哋今次就想考你對廣東話嘅認識有幾深。

本挑戰由16個香港年輕語言學者合力寫成，分為4級，共1000分，需時大概半個鐘。題目類型涵蓋多個範疇，包括語音、音系、詞彙、語法、語義、俗語、歷史、社會等等。

話明係達人挑戰，即係唔係母語講廣東話就識做嘍，甚至喺 Google 都未必搵到答案。成日話廣東話博大精深，唔係講多句粗口就叫博，唔係講句「廣東話好難學」就叫深。希望大家玩完呢個達人挑戰之後，會明白一個日日都講嘅語言背後，可以有幾多鮮為人知嘅秘密，可以有幾多深厚嘅學問。希望大家真切認識到廣東話嘅博大精深！

=====

【Level 1 - 小學雞級】如果廣東話係你嘅母語，相信你會覺得 level 1 嘅題目好簡單，憑你嘅直覺作答啦！但係如果你嘅母語唔係廣東話，嗰就冇咁容易喇。

【Level 2 - 初哥級】就算你嘅母語係廣東話，睇怕都要諗吓先答到。

【Level 3 - 老手級】燒腦問題正式開始！

【Level 4 - 學者級】你即將面對嘅係語言學學者研究嘅問題！

0~140: 廣東話學習者

141~200: 流利嘅廣東話使用者

201~430: 對廣東話稍有研究

431~600: 對廣東話瞭如指掌

601~800: 獨當一面嘅廣東話達人

801~1000: 辦辦掂嘅廣東話神級達人！

同所有嘅達人測試一樣，做到後段會越做越難，希望大家有心機同時間先好進入挑戰！就算遇到唔識答嘅題目，都請你估一個答案，咁樣完成挑戰之後先會顯示番個正確答案。

1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言

廣東話達人挑戰

 laiikpo@gmail.com (未分享) [切換帳戶](#) 

Level 2 - 初哥級

Level 2 開始深入少少嚟，就算你嘅母語係廣東話，睇怕都要諗吓先答到。

【詞義】以下邊個講法係眨義嘅？ 6分

- 白雪雪
- 白茫茫
- 白晒晒
- 白矓矓

【詞義】以下邊個詞嘅常用語義係由隱喻引伸而嚟？ 6分

- 抽水
- 泵氣
- 開門
- 抽筒

1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言

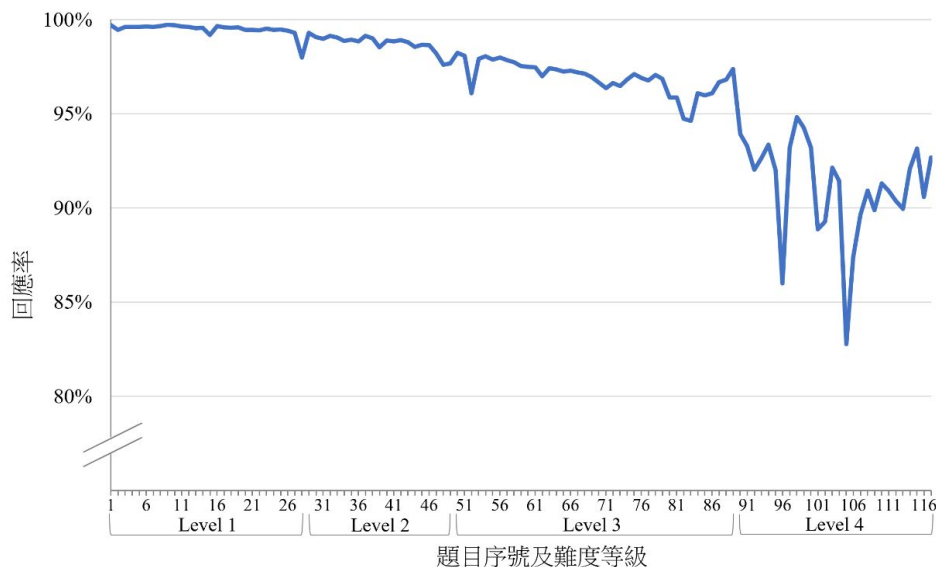
向參加者展示的介紹	製作團隊的內部指引
【Level 1 - 小學雞級】如果廣東話係你嘅母語，相信你會覺得 level 1 嘅題目好簡單，憑你嘅直覺作答啦！但係如果你嘅母語唔係廣東話，噉就 冇咁容易喇。	參加者只要是 粵語母語者就會知道答案
【Level 2 - 初哥級】就算你嘅母語係廣東話，睇怕都要諗吓先答到。	參加者按母語者語感或常理推敲後能確定答案
【Level 3 - 老手級】燒腦問題正式開始！	參加者可以按母語者語感或常理猜想答案
【Level 4 - 學者級】你即將面對嘅係語言學學者研究嘅問題！	一般參加者無法猜出答案

難度等級	Level 1	Level 2	Level 3	Level 4
題目的平均答對比率	89.3%	69.9%	48.3%	32.7%

1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言



1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言

難度等級	主題
Level 1	詞彙 詞彙配搭 語法 發音
Level 2	詞義 俗語 發音 變調文白讀
Level 3	詞嘅構造 句意邏輯 語法分析 虛詞用法俗語 粵拼 發音與語音系統 歷史
Level 4	社會 語法學 傳統音韻學 歷史語言學 廣東話嘅鄰居

1.「廣東話達人挑戰」

遊戲設計

- 大致仿效「香港地理達人挑戰」
- Google Forms
- 117道題目
- 四選一選擇題為主
- 由易到難以四個等級編排
- 越難的題目佔分越多
- 涵蓋粵語和粵語語言學的多個方面
- 完結時會顯示分數、答案和簡單解說
- 分數對應得出不同評價
- 參加者可以留言

分數範圍	評價
0-140	廣東話學習者
141-200	流利嘅廣東話使用者
201-430	對廣東話稍有研究
431-600	對廣東話瞭如指掌
601-800	獨當一面嘅廣東話達人
801-1000	瓣瓣掂嘅廣東話神級達人

廣東話達人挑戰

總分 128/1000



喂，你識唔識講廣東話？

根據政府統計處2019年嘅報告，喺全港6-65歲嘅人士裡面，有88.8%母語都係廣東話。個個都識講廣東話，唔通個個都了解廣東話咩？

我哋今次就想考你對廣東話嘅認識有幾深。

本挑戰由16個香港年輕語言學學者合力寫成，分為4級，共1000分，需時大概半個鐘。題目類型涵蓋多個範疇，包括語音、音系、詞彙、語法、語義、俗語、歷史、社會等等。

話明係達人挑戰，即係唔係母語講廣東話就識做嘍，甚至係 Google 都未必搵到答案。成日話廣東話博大精深，唔係講多句粗口就叫博，唔係講句「廣東話好難學」就叫深。希望大家玩完呢個達人挑戰之後，會明白一個日日都講嘅語言背後，可以有幾多鮮為人知嘅秘密，可以有幾多深厚嘅學問。希望大家真切嘅認識到廣東話嘅博大精深！

1.「廣東話達人挑戰」

遊戲達成了多方面的成果(黎奕葆等2023):

- (1) 為粵語語言學進行了一場受眾甚廣的科普
- (2) 檢視過往科普努力的成果
- (3) 協助評估一些語言學任務是否適合以眾包方法處理
- (4) 為特定研究議題獲取樣本相當大的量化數據
- (5) 窺探大眾對粵語及語言學的看法

1.「廣東話達人挑戰」

(2) 檢視過往科普努力的成果

問題：【粵拼】「見字飲水」用粵拼應該點標示？ 8分

選項： gin3 ji6 yam2 sui2 (42.3%) gin3 ji6 yam2 seoi2 (18.8%) gin3 zi6
jam2 sui2 (15.1%) gin3 zi6 jam2 seoi2 (23.8%) (N=39,377)

- 主流大眾對科學的粵拼缺乏認識，多依賴日常生活見於人名、地名的不嚴謹的「港拼」(Kataoka & Lee 2008; 林茵茵2009; 片岡新2014)

1.「廣東話達人挑戰」

(5) 窺探大眾對粵語及語言學的看法

「粵語是香港人身份的重要元素」

235: 編者們好有心機, 利申土生土長香港人, 但愈做到後愈慚愧

516: 後半部嘅題目好燒腦🤔, 要開始懷疑一下自己係咪香港人🤔

620: 真香港人測試!

869: 多謝! 呢個測驗令我更加鍾意香港

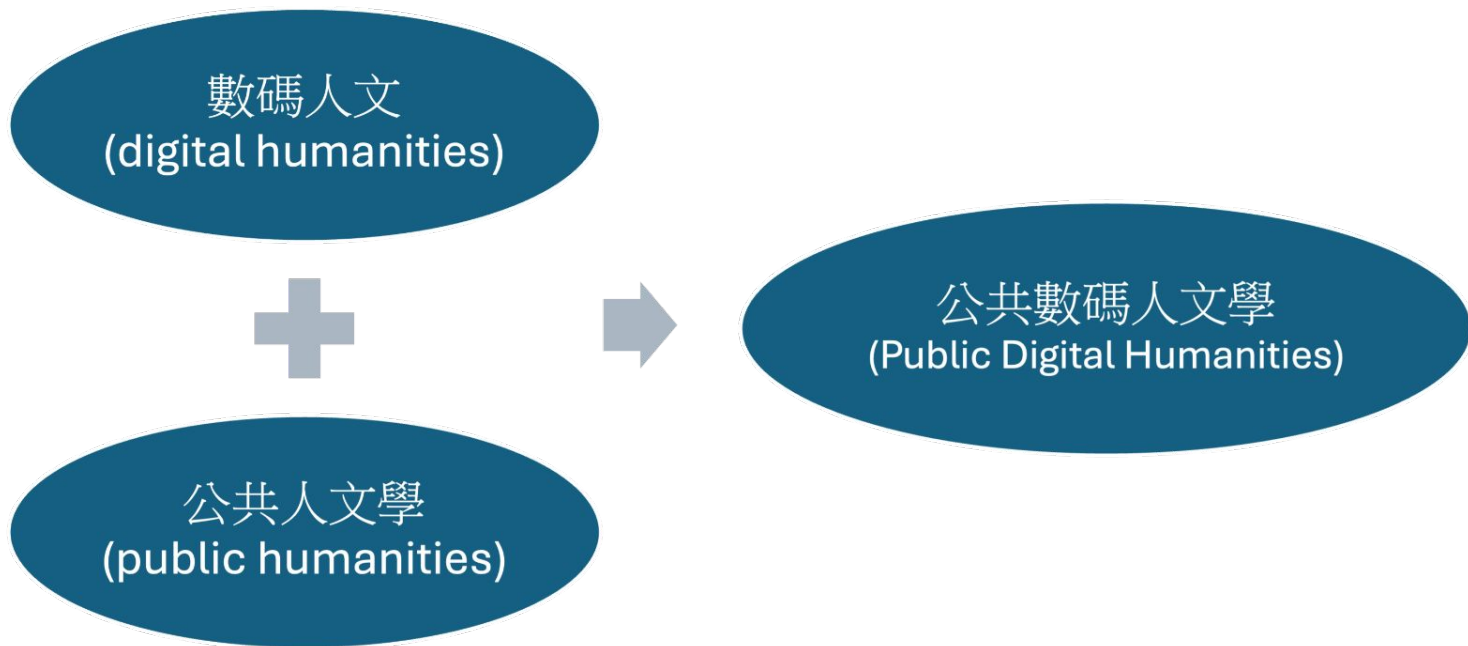
「語言運用比語言分析重要」

544: 母話廣東話嘅人講乜文法咩, 一開口就用, 理得你咩聲母韻母, 粵拼仲用英文

1302: 問題懶深入, 想食蕉根本唔洗知蕉點種。

2. 公共數碼人文學

2. 公共數碼人文學



2. 公共數碼人文學

傳統上不少數碼人文企劃都包含公眾參與的元素，眾包(crowdsourcing)是其中一種常見形式

- 早期典型的眾包實踐中，參與群眾被動地處理着被委派的碎片化的任務
- 研究人員不關注參與群眾有何得著，獲取研究數據往往是其唯一目的

近年公共人文學的思路引進到數碼人文學(參Christie et al. 2014; Brennan 2016; Stommel 2018; Cox & Tilton 2019; Schwan & Thomson 2022)

- 公眾在數碼人文企劃中的參與形式得到更多討論和反思

2. 公共數碼人文學

- 是實踐 (doing, practice), 而非單純的學術研究
- 研究課題回應大眾的需要, 而非單純出於學術上的研究興趣
- 重視公眾參與
- 公眾從參與中能有所得著 (超越金錢利益)
- 重視公眾的觀點
- 研究者與大眾之間關係平等

(參Schwan & Thomson 2022)

2. 公共數碼人文學

- 是實踐 (doing, practice), 而非單純的學術研究
- 研究課題回應大眾的需要, 而非單純出於學術上的研究興趣
- 重視公眾參與
- 公眾從參與中能有所得著 (超越金錢利益)
- 重視公眾的觀點
- 研究者與大眾之間關係平等

(參Schwan & Thomson 2022)

個案: MultiHTR (Tikhonov et al. 2022)

- 於德國研發多語手寫文本識別模型
- 參與群眾能得到破譯家傳或社區古老文件的幫助
- 提供接觸演算法訓練的機會以改變大眾對AI的負面印象

3.「廣東話達人挑戰」作為一個公共數碼人文項目

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」是公共數碼人文學的一次嘗試。
 - 透過社交平台的演算法接觸大眾
 - 這次項目的首要目標是回應大眾的需要。收集數據，又或是邀請大眾一同完成一些特定的語言項目等只是較為次要的目標。
- 當時而言，我們認為香港大眾的需要有兩個：
 - 期望社會上出現關於粵語的「達人挑戰」來挑戰自己對粵語的認識，
 - 期望自己能為粵語的保育出一分力。
- 大眾透過協助我們的研究，便滿足他們為粵語出一分力的願望。

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」在遊戲之中滲入了：
 - 學術研究數據收集的元素
 - 科普元素

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」在遊戲之中滲入了：
 - 學術研究數據收集的元素
 - 採用眾包式數據收集，通過遼闊的互聯網接觸到大量粵語群體的成員。

	修(T1)	手(T2)	秀(T3)	愁(T4)	壽(T6)	(無)	總計
收(T1)	<u>98.0%</u>	0.5%	0.5%	0.2%	0.2%	0.6%	100% (N=40307)
守(T2)	0.3%	<u>98.2%</u>	0.4%	0.3%	0.4%	0.4%	100% (N=40257)
瘦(T3)	0.5%	0.4%	<u>88.3%</u>	0.4%	9.6%	0.9%	100% (N=40278)
仇(T4)	0.2%	0.2%	0.3%	<u>97.8%</u>	0.8%	0.8%	100% (N=40309)
受(T6)	0.5%	0.6%	13.9%	1.5%	<u>82.8%</u>	0.9%	100% (N=39973)

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」在遊戲之中滲入了：
 - 科普元素
 - 我們希望大眾能夠透過回答個別問題來學會不同的語言學知識。相關題目的答案分佈並不是我們首要的考慮因素。
 - 通過問題的內容來帶出不同的概念和訊息，而不是答案和解說。
 - 以免參加者跳過答案和解說。
 - 閱讀題目與分數攸關，但事後看解說則沒有分數的鼓勵。參加者有更大的動機閱讀題目。

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」在遊戲之中滲入了：
 - 科普元素
 - 通過問題的引導和選項之間的對比，不少參加者似乎都能馬上領會其意。

問題：	【變調文白讀】以下邊個字冇文白異讀？ 6分
選項：	生(6.2%) 命(11.6%) 青(7.4%) <u>春(74.9%)</u> (N=39,580)
解說：	「生」文讀eg「生產」白讀eg「生仔」；「命」文讀eg「命令」白讀eg「冇命」；「青」文讀eg「青春」白讀eg「青色」

3.「廣東話達人挑戰」作為一個公共數碼人文項目

- 「廣東話達人挑戰」在遊戲之中滲入了：
 - 科普元素
 - 部份題目除非參加者本身就知道答案，否則他們基本上沒有辦法把答案猜出來，我們實際上也不預期參加者能答對這些問題。
 - 我們想要傳達的訊息已經在這些問題的前設之中。
 - 「語言是會隨時間而演變的」

問題：	【歷史語言學】以下邊個字喺19世紀嘅廣東話中同其他三個字唔同音？12分
選項：	<u>將</u> (60.0%) 彰(23.5%) 章(8.3%) 張(8.2%) (N=37,025)
解說：	可以留意吓，「將軍澳」個「將」同姓氏個「張」，佢哋嘅英文拼寫唔一樣

4. 公眾需要與研究需要之間的矛盾與協調

4. 公眾需要與研究需要之間的矛盾與協調

- 矛盾1: 同途 殊歸
 - 公眾需要: 與粵語相關的達人挑戰, 同模式下的不同主題
 - 研究需要: 大眾對粵語及語言學知識的掌握、觀感
- 協調: 以公眾需要為大原則, 研究需要大幅度讓步
 - 貼緊達人挑戰的模式(選答題、由淺入深、量化難度、覆蓋範圍全面)
 - 避免重複性的問題(例如最小對比對), 減低作業感
- 為開拓新的研究方式及資源自願受限(例如課題的選擇及深度)

4. 公眾需要與研究需要之間的矛盾與協調

- 矛盾2: 應否收集個人資料?
 - 公眾需要: 娛樂性質, 不必透露個人資料
 - 研究需要: 個人資料有助資料解讀
- 協調: 以不改變娛樂性質為前提, 不犧牲普遍性
 - 沒有收集性別、年齡、語言背景
 - 沒有限制作答區域(凡擁有連結者均可參加)
 - 沒有限制作答次數
- 為提升了科普效能, 但削減了有助詮釋數據的資料
 - 透過接觸大量受眾, 從而達到科普成效, 並以量取勝, 以大量數據克服極端例子

4. 公眾需要與研究需要之間的矛盾與協調

- 矛盾3: 研究意圖應否隱藏？
 - 公眾與研究者之間資訊不對稱, 目標亦不一致
 - 公眾參與的目的是挑戰而非協作研究
- 協調：
 - 列明「呢個達人測試收集到嘅資料可能會作學術研究用途。」
 - 最低限度的提示
- 反思：
 - 如何拿捏研究的透明度
 - 如何說明大眾在當中的角色
 - 如何避免「大眾被視作研究的棋子」的負面觀感
 - 研究同時須要顧及社會觀感

5. 總結

