

香港粤语的文字化困境与出路

2019年9月，我在香港、广东、广西和重庆探亲访友并沿途进行简略的人类学/社会语言学考察，注意并确认了中国南方（加上云南、贵州和部分四川地域）其实属于受上古印度影响的所谓东南亚高地Zomia/ᄒᄒᄒᄒᄒ前沿¹，正在进行着已经基本完成了的汉化进程，而尤以南方各种语言/方言的文字化困境显著地提示着这个汉化进程（的程度），也进一步证实改进汉语—导入汉音元素—是唯一可以解决这个困境的方案。在“以汉音元素帮助创制少数民族文字的设想”²中加入“侗族语言的文字化”一节，本文考虑香港粤语的文字化困境。

我9月7日周六晚抵达香港机场。机场内访客稀松，没有遇到8月份来到机场的抗议者。我当晚乘7人座位的小巴（这个词来自香港粤语）直接从深圳离开香港进入中国，沿途也没有遇到任何抗议的迹象。9月23日周一下午，我再次从深圳乘列车进入香港西九龙站，在地下通道看到贴满两旁的各种抗议（以及少数反抗议）标语、传单。这里不是诉诸理性的地盘，不少内容充满政治性的对立和仇恨，其中还包含隐蔽的文化（讲粤语与非粤语者）的歧视（插图）。25日周三我离开香港，乘大巴（这个词也来自香港粤语）从尖沙咀开往机场，沿途通畅顺利，进入机场时被要求出示证明，这就是机场内见不到抗议者的原因。



我在九龙下榻旅馆附近的老旧建筑之间的夜间市场观察下层市民的生活（插图），与新兴深圳的欣欣向荣形成强烈的反差。香港的人均收入5万多美元，早已超过英国，但贫富差距(2016年gini指数就高达53.9)远超所有发达国家(英国2015年为33.2)。以房价为例，我看到实用200平方英尺的租金可以高达每月15,000-17,000港币，房价高达500万港币（因为抗议运动已经有所降低）。抗议运动的持久深入广泛，把最大的利益集团香港的土地开发商们也推到香港公众、中国内陆和国际社会的视野面前。各种矛盾的公开化，有助于防止权势各方黑箱造作的交易，多少防止了最糟糕事态的悲剧（如89年的天安门事件）。

24日周二我在维多利亚港周围的尖沙咀、中环、八一大楼、政府总部和立法会一带步行良久。除了立法会大楼周围的媒体车辆和记者，一切都平静如常（这一带禁止张贴标语）。我只身进入立法会的入口，询问参观的可能，被告知今天不对外开放。当我研究诸如西班牙内战³、俄罗斯革命的历史⁴时，很向往能够亲身到现场考核，现在置身于当代史事件



¹ 参见赵京：“当代人类学力作揭示的山地东南亚的安那祺史观”，2011年8月8日。

² 赵京，2019年10月1日第二稿。

³ 赵京：《西班牙内战的安那祺主义教训》，2015年1月25日第五版。

⁴ 赵京：《鲜为人知的俄国革命：马赫诺运动、喀琅施塔得起义及托洛茨基主义、列宁主义问题》，2014年2月11日第三版。

的平静现场，加深了对香港局势的认知。



当日晚，我参加香港文化中心为公众举办的免费大型露天音乐对话和伦敦交响乐团表演（约5百听众）。除了慢慢地习惯港式粤语的发音，我也注意到三个音乐家对话中自然讲出的英语词汇为听众自然地接受，简易英语已经成为香港粤语的一个不可缺少的部分。作为舞台背景的海岸对面的霓虹灯火映照下的升平气氛，似乎驱散了政治抗议和经济不安的阴影。

很多人从不同角度、立场或个人生活经验、信息量多少关注目前的香港问题，多同意没有任何“解决”方案，这可能反而是理性能发挥作用的唯一期待。对于香港目前面临的政治、社会、经济等综合症，我虽然从自己研究过的巴黎Commune/自治体“公社”⁵、Anarchism/安那祺主义运动⁶的历史自然地看出类似特征，并没有自己独特的见解，也没有精力像1993、2010年访问香港时会会见同志、朋友⁷。我更关注到粤语的文字化这个长期复杂的文化问题对讲港式粤语者和非粤语者的交流的阻碍。

粤语是一种声调语言，受五代十国时期人口南迁影响，保留了大量古汉语用词。如用“系”而不用“是”来正面答复，“走”则保留了古汉语中奔跑的意思（日语中也如此）。再看，“曲”字本身是Uighur/𐰇𐰺𐰽𐰾𐰿/𐰇𐰺𐰽𐰾𐰿儿语kuy/köy（音乐）的音译，唐代（和现在客家话）发音kiok，粤语发音kuk，闽南语发音giok，日语发音きょく/kyoku，都比现在的转自满族语的普通话（Mandarin/𐰇𐰺𐰽𐰾𐰿/“满大人”）发音更接近原初的𐰇𐰺𐰽𐰾𐰿儿语发音⁸；这个例子还说明粤语、客家话、日语、古汉语等的单字不必是普通话单字中的单音节拼音，也就是说，拼音不是表示多音节字和词的合适手段。不过，粤语虽然也被称为“广东话”，但在广东、特别是广州，汉语普通话已经普及为第一语言，很难遇到粤语文化的气息，只能到岭南“圣域”祖庙（佛山）才能体会到武术、粤剧等粤语/岭南文化。我听了没戴脸谱的粤剧男女对唱，感觉与京剧类似。只有在香港，粤语是第一语言（90%），保持和发展出独特的港式粤语⁹。



粤语讲者在正式场合里普遍使用二十世纪初兴起的白话文书写系统，其语法、词汇与现代标准普通话相符，与粤语自身的语法、词汇差别很大，文言脱离，只有“粤语白话文”，而没有正式的“粤文”。粤语白话文书写会用到大量粤语独有的粤语汉字，这些字在BIG5系统或使用简化中文的电脑字符系统中没有收录或收录不全。为解决粤语的电脑文字处理问题，香港政府制定了一个香港增补字符集，收录了常用的粤字和其他BIG5扩增汉字约5000字左右，如“啲”、“嘅”、“攞”、“揸”、

⁵赵京：“巴黎公社悲剧的意义”，2010年11月8日第二稿。

⁶赵京：《安那祺/自由社会主义》，2015年12月18日第四版。

⁷赵京：“台港澳掠影”，2010年12月4日。

⁸赵京：“中亚文明简史新译初步”，2019年4月5日第一稿。

⁹<https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF%E7%B2%B5%E8%AA%9E>

“嘢”、“㗎”等。在最新一版的增补字符集中，收录了一些所谓的“粗口字”。这套增补字元集基本上解决了粤语白话文的电脑处理问题，大部分粤语口语文句都可以录入电脑。但广东以外的地区的粤语讲者对粤语固有字词的了解近乎空白，倾向于使用北京话借音字书写粤语，造成书写混乱。¹⁰很显然，这些主要表示发音的粤语汉字应该用规范的汉音元素来表示，统一发音。

语气助词在粤语语句情貌的表达上有非常重要的作用，不可缺或。但其用法复杂，有单式、二覆叠式、三覆叠式和多覆叠式等形式，如：喇-表示完成或引起注意，喺-提问，或引起注意，喎-表示意识到，“原来如此”意味，喇-暗示否认某事物之结果或叙述之意味，嚟-表示情况就是如此，嘍-期待某事物之时机即将达至之语气，咩-表示怀疑，或轻微不屑，咁-表示不敢肯定自己的判断，啫-表示不屑语气，或表示“而已”的意思，㗎-第二声，表示抱怨语气，嚟-表示完成语气，嘍啫-表示事情小，不必放在心想，㗎啫/嘍啫-同“而已”，可用单字“啫”，啫嘛-轻视、不屑意味，啦喂-提醒或积极回复意味，㗎喂-寻求最后确定，或表示已经不耐烦，㗎嚟-表示确认，或希望得到确认，㗎咩-表示强烈怀疑、不赞同，㗎喇-已经完成，或完全自信、确认，咁喺-对其之前所作的行为表示不满，㗎嚟啫-表示确认，或希望得到确认，㗎喇喂-表示事情已经发生，且带有不满的意味，㗎喇吓-表示对已经发生的事情进行确认，㗎喇可-通常用于询问他人某物是否已经完成使用，做晒㗎啦吓咩/㗎啦啊嘛-“已经做完了，是吧？”带半信半疑意味。¹¹这里引用较多，说明这些粤语汉字并不具备任何表意功能，应该用表音的汉音元素来代替，既简便又准确。

粤语外来词主要来自英语，广州从清朝起已经出现外来语。而到香港殖民地史时期，香港粤语(港式粤语)中吸收外来词特别多，影响着广东境内的粤语区。从1980年代开始，不少粤语外来词渐渐进入了汉语普通话，例如“巴士”(bus/ㄅㄨˊ)等。有时，这些词被汉语普通话吸收的时候变形，如粤语“搭的”(乘搭的士)被汉语普通话当作“打的”吸收。有些外来词很多是汉语普通话没有吸收的，如“士多”(store/ㄊㄨˋ)；有的是普通话吸收了但译法不同，如汉语普通话中的“沙拉”是在粤语中带有入声的“沙律”；不少外国人名在粤语中的译法，与汉语普通话存在很大差别，如美国总统Bush/ㄅㄨˊ在普通话汉语中翻译成“布什”，台湾的国语译作“布希”，香港粤语则译成圆唇的“布殊”，还有Reagan/里根/雷根/列根等。这里正是汉音元素的用武之处，只有用标准的汉音元素，才能与英语原音接近，避免各当地语的发音不同而滥用不同的表意汉字来替补。

结论是：这些与发音相关的粤语汉字，都可以、也应该用汉音元素表示。进一步的问题在于：是否，以及如何补充代表粤语发音独有的“粤音元素”到汉音元素方案中？这超出了本文作者的能力，期待别的学者的进一步工作。

[赵京，中日美比较政策研究所，2019年10月4日]

¹⁰ <https://zh.wikipedia.org/wiki/%E7%B2%A4%E8%AF%AD>

¹¹ <https://zh.wikipedia.org/wiki/%E7%B2%A4%E8%AF%AD>