

代表属性:发音人的选择和语料有效性¹

梁源

香港教育学院

yliang@ied.edu.hk

黄良喜

香港浸会大学

lianhee@hkbu.edu.hk

提要: 属于实证研究的语言学研究, 一般需要选择发音人作为研究对象的有效代表, 因此发音人必须符合研究所需的代表属性。通过设定外显研究对象主要特征参数, 本文提出: (1) 代表属性可最直接地通过参数来表述; (2) 越满足代表属性参数的研究对象就越能成为研究的有效代表。本文通过几个语言学研究示例展现了代表属性在语言学研究中的根本性。尽管研究的问题、目的、甚至方法都不同, 我们从中可以看到, 厘定代表属性直接影响研究所使用语料的有效性。

关键词: 代表属性; 参数; 发音人; 语料有效性

一、问题

方言研究需要选择具有代表性的发音人。通行做法是参考年龄、性别、教育和生活环境等等因素, 选择老派、识字、一直在当地生活的男性或女性。由于一般的方言调查选择发音人的数量不会太多(大约 3-4 名, 见《中国语言资源有声数据调查手册》(2010)), 如果发音人选择错误必然会造成很大影响。加大发音人的数量是一个解决办法, 但是, 选择多少名发音人才能降低发音人错误的影响呢? 事实上, 通行做法的目的在于辨认发音人的语言是否具备代表性。这种做法对于人口流动较小的语言社团比较管用, 但对于人口交际复杂的现代都市则很难保证发音人的代表性。

以深圳为例。深圳在经济特区成立(1980 年)以前, 本地常住人口主要通行两种语言: 本地粤语(南头话)和客家话(汤志祥、林建平 2006)。随着特区成立, 大量移民和建设者带来了不同的方言, 与深圳本地语言以及普通话互相接触、互相影响。如果以方言数量来讲, 深圳人群可以简单地分为单语者和双语者两大类。单语者人群主要包括为数不多的交际圈狭窄的本地客家或粤语者、零星的来深移民的年长家眷(操自家方言)。双语者则为数众多, 使用方言/语言包括普通话、深圳新粤语、闽语、客家话、省外各地方言、甚至外语(梁源、吴育珊 2011)。面对深圳的语言情况, 按照通行做法来选择代表性发音人就会困难重重: 单纯地选择南头话发音人作为深圳粤语研究对象已经不能代表深圳的实际语言环境; 同样地, 选择普通话或者广州/香港粤语者也不妥(我们总不能在深圳找一个广州人并以他的粤语作为深圳粤语的代表!)由此可见, 语言研究必须考虑到实际的语言生态(language ecology)对于不同发音人的影响, 发音人的代表性必须以实际语言环境为主要考量。

深圳的语言环境将聚光投射到了“代表性”这个老问题上。如果可以直接地鉴定发音人在语言因素²上是否具备代表性, 便可以有效地选择出具备代表性的发音人。

(1) 代表性的问题

在语言学研究中, 我们如何选择具有代表性的发音人?

¹ 本论文得到香港教育学院研究项目“多重语体格局下的母语习得与变异”(04090)和暨南大学汉语方言研究中心自立项目“深圳地区方言的地理语言学研究”(14FZ17)的资助。论文初稿曾在“第三届中国地理语言学”国际学术研讨会上宣读, 感谢与会专家的意见。

² 这里的“语言因素”不是指语言特征(即语音、语法、词汇等特点), 而是指发音人的背景如何使得他成为这个语言群体的代表。

像深圳这样的情况其实很普遍。所有大城市都有不少人口流入的情形，这势必带来语言的接触和变迁。北京、上海如是，新加坡、香港、纽约、伦敦亦如是。代表性问题在现代社会的语言学研究中将随着人口流动的增加而愈发突显。

二、代表属性

许多研究领域都是通过对有效代表的探索来推导对群体的认识。作为研究对象的群体往往是先验成立的。

(2) 研究对象：由先验形成的群体。

要弄清楚研究对象群体是否为一个有意义的集子，最好的方法是将其主要特征描述出来。特征描述如果借助参数，就可以较为清晰地判断给定个案是否能够作为研究群体的有效代表。参数设定越精细就越能外显研究对象的主要特征、并找到有力代表；同理，越有力的代表应当更符合这些参数。括而言之，研究对象有效代表的选择必须满足研究对象主要特征的参数。我们称之为“代表属性”，定义如（3）。

(3) 代表属性：一套准确描述研究对象主要特征的参数（parameter）。

代表属性的确立直接影响了实证研究的有效性，我们选择满足代表属性的个案不必过多、也不应过少。

参数的确立必须依据不同的研究目的。假定我们的研究题目是“中国古代文人居所”，那么，所有中国古代文人生活和休憩过的地方都属于我们的研究对象。可能是草堂、可能是四合院和寺庙、还可能是山洞和天桥底下等等（陈平 1992，丁俊清 1997）。这些个案既包括建筑物也包括非建筑物，它们之所以形成一个研究群体是在于“文人居住”这条先验的标准。而在少林寺、潭柘寺、杜甫草堂、白马寺、寒山寺、孔庙、妈祖庙、城隍庙、黄大仙祠、布达拉宫等等当中，只有杜甫草堂是有效代表，其他寺庙也许文人曾经游玩和借宿过（如白马寺和寒山寺），但不足以满足“居所”这个条件。支持我们作出如此选择的是一套参数，这套参数是依据研究题目而来的。就中国古代文人居所而言，参数可以包括：社会意义、功能、时间、材料和结构（高度、宽度、面积和形状等）。其中，[社会意义]定值为【中国文化人的生存空间】；[功能]定值为【居所】；[时间]定值为【民国以前】；[材料]和[结构]则无需定值。

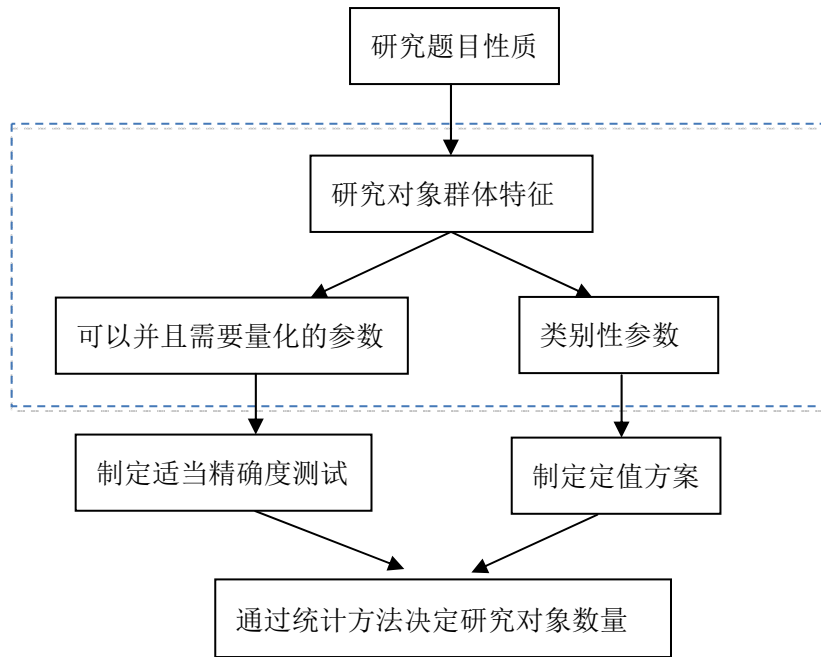
假定我们的研究题目是“女性的爱情观”，选择女性群体为研究对象似乎理所当然。然而，女性群体包括 tomboy³以及双性取向者。产生这一问题在于没有理解清楚研究题目的性质。“女性的爱情观”这一题目的性质是社会意义，因此，研究对象所要求的“女性”不可盲目地以生理属性作为选择标准，而要考虑是否以“社会角色”来定义女性。据此，选择研究个案的正确参数在“性取向”（反映其社会角色）与“生理性别”两个参数间，当以前者为更贴切；是否也选用后者，则取决于研究者对这个课题所设的涵盖面有多广。当然，如果我们的研究题目是“女性保健”，那么，选择研究对象的参数在“性取向”与“生理性别”两个参数间，就当以后者更为贴切。

一直以来，语言学研究都面临如何选择有效发音人的问题。解决这个问题的关键在于首先要理解清楚研究题目的性质，接着，就要根据研究目的罗列研究对象的主要特征并设定参

³ Tomboy 即女同性恋者中自诩为男性的人。

数。有了参数我们才有可能进行定值，有了定值才有可能决定要不要量化、以及如何量化。因此，代表属性问题是解决语言学研究方法论（甚至其他实证研究）的根本。我们可以归纳为图（1）。

图（1）：解决代表属性问题的程序



从图（1）可见，由研究题目到最终决定需要多少个研究个案，中间往往需要根据研究对象的性质选择适用的个案或个体，然而，这些性质并不一定都可以或者需要量化的。比如研究“中国古代文人居所”中选用的参数“高度”，是可以量化或细分为若干等级的。而研究“女性的爱情观”中的参数“生理性别”，一般来说不会量化或者细分为“很女性、明显女性、中等女性、稍微女性、微微女性、很不女性”，因为生理性别是类别性参数。当选定了能够帮助研究者辨认研究对象的参数以后，才可以制定测试和筛选研究个案的方案，最终，还应当按研究题目决定到底需要多少个个案。

图（1）的虚线框是本文重点，旨在说明如何寻找具备代表性的研究对象。这样，我们才能够对研究结果的有效性可以进行较客观的评价，避免单纯依靠多、庞、杂来鉴定研究价值。

三、语言学研究示例

本节将以几个语言学研究为示例展现代表属性在语言学研究中的根本性。尽管研究问题、目的、甚至方法不同，这些示例都让我们看到，厘定代表属性直接影响了研究所使用语料的有效性。

3.1 方言学研究

方言学研究一般要求到若干调查点采集语料，从而建立对该方言的描述。在交通相对局限的古代，这个做法基本没有什么问题。因为人口流动较小，各地方言的差异比较稳定（农村经济生活分散，见袁家骅(1989)）。因此，如果以地区作为选择发音人的参数，的确可以符合研究所需的代表属性，毕竟当地人就是操当地语言的有效代表。然而在当今社会，人口

流动、教育普及、媒体传播模糊了发音人是否满足代表属性的界线 (Siegel 1993, Trudgill 1986)。譬如调查上海话, 选择两个在上海出生和成长的发音人, 他们在一起使用上海话交谈、中间却夹杂着不少普通话的词语。我们无法肯定也无法否定他们说的是地道⁴上海话。这种情形在当今上海非常普遍, 而且, 不仅上海如此、他处亦然 (甘于恩 2003, 邵敬敏、吴立红 2005, 包联群 2007)。

我们首先要明白方言调查只是一个研究方法, 它可以适用于不同的研究项目。如果研究目的是“了解现代上海话的语言状况”, 那么, 发音人所说的上海话夹杂着普通话、甚至外语都不构成问题, 因为这才确实实地反映了现代上海话的实际状况。从《上海方言词典》(许宝华、陶寰 1997) 可以看到, 上海话内部分为市区老派、市区中派和市区新派, 在语音上可以看到普通话的不同层次的影响。根据 (2), 为“了解现代上海话的语言状况”的研究个案的代表属性应当包括: 是否在上海长期居住、语言身份认同等。其中, [是否在上海长期居住]宜定值为【是】、[语言身份认同]⁵宜定值为【上海话使用者】。顾钦 (2007) 在探讨普通话对于现代上海话语音演变的影响时, 采用的就是这两个参数⁶来选择发音人。毕竟, 研究语言状况的先决条件是找出适当的使用者, 以便理清这个群体的语言使用状况。而什么是地道的上海话不是先决条件。但是, 如果研究目的是“通过上海话构拟中古语音”, 那么, 辨认地道的上海话就成为了先决条件。根据 (2), 研究个案的代表属性包括: 是否承袭上海话口传文化、经常使用上海话、是否被集体认定等。其中, [是否承袭上海话口传文化]宜定值为【是】、[经常使用上海话]宜定值为【经常使用】、[是否被集体认定]也宜定值为【是】。这三个代表属性分别承担不同的特征要求: 承袭口传文化保证了发音人具备稳定保留传统文化和语言元素的特性, 高频率使用保证了发音人上海话的鲜活性, 而集体认定则避免了判断的主观个性。当然, 根据更加具体研究的目的和情况, 应当适当增加年龄、生活环境、性别等参数, 以展示新老派或者更小地域之间的口音差异。有效使用年龄等等参数成功调查上海话, 在前人研究中比比皆是 (游汝杰 2006, 许宝华、汤珍珠、陈忠敏 1993)。

如果研究目的是“保存上海话”, 那么, 上述参数中也有适用的, 如[语言身份认同]、[是否承袭上海话口传文化]、[经常使用上海话]、[是否被集体认定]等。当然, 还有其他研究目的, 比如“探讨上海话的语音演变”、“上海地区不同经济阶层的口音特点”等等也同理 (参见沈同 1981, 许宝华、汤珍珠、钱乃荣 1981、1982、1983, 焦成名 2009)。

3.2 类型学研究

类型学研究注重厘清语言和语言之间的类型关系, 寻找语言不同表征现象的系统性不同, 进而给全世界语言分类和解释语言的多样性 (Comrie 1989)。语言社团的不同归类对于类型学研究至关重要, 因此, 类型学可以研究语言内部的类型差异 (比如世界各地英语的类型差异), 也可以研究不同语言之间的类型差异 (比如汉语和日语的类型差异)。

英语作为一种语言, 内部包括了不同的语言社团, 如英式英语、新加坡英语、港式英语、印度英语、南非英语、澳洲英语、菲律宾英语、美式英语等等。比较英式英语和新加坡式英语, 可以发现很多类型上的差异, 比如有无声调、有无格标记和词序。见表 1 (Wee 2009)。

表 1: 英式英语和新加坡式英语类型比较

⁴ “地道”是一个很模糊的概念, 涉及到如何定义不同的方言和语言: 从至小的认为个人语言都是一个独立的语言 (I 语法), 到至大的历史同源的种种语言为一个语言。中间还可能包括由于语言接触而产生的语言。

⁵ 这里的“身份认同”是指自我认同, 而下面的“集体认定”指被他人认同。

⁶ 除此之外, 顾钦 (2007) 还使用年龄 (10-25 岁) 作为参数, 他认为 25 岁以上发音人的上海话已经有前人调查, 10 岁以下发音人的语音情况还不稳定, 因而不做研究。

类型差异	英式英语	新加坡英语
轻重音/声调	轻重音 例: electric[ɪˈlektɹɪk] university[junɪˈvɜːsɪti]	词末音节高平调, 其他音节中平调, 首音节可能为低平调 例: electric[ɪ ¹¹ lek ³³ trɪk ⁵⁵] university[ju ³³ ni ³³ və ³³ si ³³ ti ⁵⁵]
有无格标记	只体现为代词的主格和宾格 例: he/him, she/her	同英式英语
WH 疑问词的前置	前置 例: Where did he go?	自由 例: Where he go?/ He go where?/ He where go?

上述研究有赖于大量语料。这些语料的采集会跨越社会各个层面以及语言所使用的各种文体, 还可能涉及多名发音人。选择发音人适用的参数包括[语言身份认同]和[是否被集体认定]。这两个参数保证了发音人无论从自我认同还是他人认同上都属于该语言社团⁷。

更普遍的类型学研究关注不同语言之间的类型差异。如果研究目的是比较汉语和日语的语音类型, 常见做法是分别选择标准汉语和标准日语的发音人。选择发音人的参数包括[官方认可]⁸、[是否被集体认定]。那么, 根据这些发音人得出的汉语和日语的语音类型差异如表 2。

表 2: 标准汉语和标准日语的语音类型比较

语音类型差异	汉语	日语
计算的韵律单位	音节	摩拉
轻重音/声调	声调	乐调重音
复辅音	不允许	不允许
韵尾辅音	只允许[n], [ŋ] ⁹	只允许[n]

可是, 如果扩大汉语和日语的范围, 不仅仅指标准汉语(即普通话)和标准日语(即东京日语), 那么, 汉语作为一个语群, 内部至少包含七大方言群。如果都考虑在内, 那么可以允许的韵尾辅音会包括[m, n, ŋ, p, t, k, ʔ] (李小凡、项梦冰 2009)。日语同理, 东京日语是乐调重音语言, 而鹿儿岛则更接近声调 (McCawley 1978)。从这个角度出发进行类型比较, 就无法以[官方认可]鉴定发音人的代表属性。解决方法可以是先进行语言内部比较, 抽出汉语和日语的核心特点, 通过核心特点的比较达到研究汉语和日语类型差异的目的。选择发音人的参数可参考语言内部类型差异的研究。

3.3 语音语料库研究

语料库涵盖多种主题。就汉语而言, 有汉字语料库、现代汉语口语语料库、古代汉语文献语料库、方言语料库等等。语料库也涵盖多种形式, 有基于文本的语料库、基于图像的语料库、以及基于录音或录像的语料库等等。

如果语料库的研究涉及语音, 那么, 我们会面临选择发音人的问题。假设要建立一个“深圳粤语语音语料库”, 那么, 发音人应当说什么样的粤语呢? 梁源、吴育珊(2011)发

⁷ 根据具体的研究情况, 还必须考虑发音人群体在年龄、性别和社会阶层等方面的数量。

⁸ 不同地方采用的“官方认可”的准则不同。比如汉语目前比较常用的是“汉语普通话水平考试(PSC)” (中国国家语言文字工作委员会)。

⁹ 本文暂不考虑儿化的情况。

现：伴随着特区的成立（1980年）和快速发展，深圳在种种动力（人口涌入、经济腾飞、即时城市化）相互交错下，出现了一种深受邻近大都会香港的影响的粤语变体。这个粤语变体是目前深圳的通行粤语（其音系特点详见梁源、吴育珊（2011））。可是，在特区成立以前，深圳本地人通行的粤语是南头话（属于粤语莞宝片，语音特点详见汤志祥、林建平（2006））。这和梁、吴所述的粤语变体差异颇大。建立“深圳粤语语音语料库”不可不考虑这些问题。因此，选择发音人的参数应当包括：[语言身份认同]和[是否在深圳长期居住]。其中，[是否在深圳长期居住]定值为【是】；[语言身份认同]定值也为【是】，同时需要考虑发音人认为自己讲的是南头话、深圳粤语新变体还是其他粤语方言，因为不同判定会影响到所收集语料的具体归类。而语料的不同归类将构成使用“深圳粤语语音语料库”进行研究的基础（详见下文4.4）。

3.4 代表属性的满足问题

有了代表属性的参数，当然不能保证一定找到完全符合一切参数的发音人（受到种种现实条件限制）。尽管如此，通过厘清这些参数，我们可以更具体地知道所找到的发音人在什么层面上符合研究需要。

前文（3.1节）提到，辨认地道方言发音人的参数包括：[是否承袭该方言口传文化]、[经常使用该方言]和[是否被集体认定]。其中第一个参数恐怕不容易满足。以广州话为例。广州话口传文化中的讲古、粤曲、南音、童谣、木鱼、龙舟等，有不少已经失传或者消亡，年轻一辈中满足这个参数的发音人可能绝无仅有。那么，这是否意味着所有涉及辨认地道方言发音人的研究都不可行呢？其实，任何研究都有其局限性，我们不必因此裹足不前。我们必须看到，语言地道性可以包括不同层面：满足[经常使用该方言]和[是否被集体认定]的发音人可以确保其广州话的延续性；而满足[是否承袭该方言口传文化]的发音人可以确保其广州话反映了语言历史文化的各种元素。

针对一个具体研究题目，我们根据研究对象的主要特征设定了一套参数，还可以进一步制定出各种测试，用以发现各发音人对于个别参数的满足程度，这是一种将参数量化的办法。诚然，参数（类别行参数）也可以通过二元制定值。具体采用哪种方法取决于测试的设计精密密度。以“经常使用该方言”为例¹⁰，我们可以（i）直接询问发音人是否经常使用该方言，找出二元制的定值【是】或【否】；（ii）通过问卷对发音人使用方言的频率进行量化（每周多少小时？使用场合的多少？等等）；甚至（iii）通过配置录音器随身录音若干时日，直接测量方言使用量（录音可能包括其他方言/语言使用量，因此可以用来测试不同方言/语言的使用比率）。这三级测试分别代表了量化的不同精密密度¹¹，可以依照研究目的进行适当选择和调整。

四、深圳语言研究

回顾文章第一节，深圳的语言现状给语言调查提出了难题。在这个移民城市，常用语言随着社会生态不断发生变化。如果以“深圳粤语”来概括在深圳通行的粤语，在1980年前“深圳粤语”还可以指南头话，在特区成立后“深圳粤语”就很难说清楚是南头话、还是在

¹⁰ 本文只列举了一个参数量化的例子，其他参数的量化方针可以有不同方向，也可以有不同精密密度的要求，但需要注意，这些不同都必须依照研究目的而定。

¹¹ 必须注意，不是越精密的方针就一定越好。过于精密的测量会将发音人过分分层，这些分层很可能将本属同一群体的发音人错误分类（乃至结果找到的是个人语言 *idiolect*, Chomsky 1965），导致整个语言系统似乎完全没有一致性的效果。所以，量化测试的精密密度必须根据研究目的和研究对象作出适度调节。

深圳常见的其他白话。深圳作为一个语言调查点，除了深圳粤语，我们要必须考虑普通话和其他方言。对于像深圳这样语言生态复杂的语言点，如何选择具备代表性的发音人将直接影响其调查方案。此外我们还应当看到，不仅深圳（大城市）如此，其他地方如镇、乡和村等，也会随着社会经济、普通话教育和人口流动的冲击而出现语言生态的复杂现象。因此，深圳语言体现了一个问题的典型，理解它、厘清它，对于现代语言调查应该有着十分积极的作用。

对深圳语言点进行调查，表面上看是一个项目，实际上包含了若干个不同层面的小项目，如“调查深圳本地语言”、“调查深圳新移民的语言”、“考察深圳的通用语”等等。这些小项目都涉及到对深圳语言点进行调查，代表着深圳语言的不同方面。下面，我们姑且以建立“深圳语言点语音语料库”作为一个大项目进行探讨。

4.1 小项目：深圳南头话/本地客家话

根据汤志祥和林建平（2006），深圳在成立经济特区以前，本地常住人口主要通行两种语言：本地粤语（南头话）和本地客家话。如果想了解这两种语言的现状，研究对象应当是这两种语言的使用者，选择有效代表的参数包括：[语言身份认同]、[是否被集体认定]、[深圳居住者]。第一和第二个参数保证了发音人对所使用的南头话/本地客家话的自我认定，以及被他人集体认定，第三个参数则限定了这两种语言必须在深圳使用，反映的是深圳的语言现状。

4.2 小项目：深圳新移民¹²语言

深圳成立特区以后，大量移民涌入并定居在深圳，他们形成了不同的方言群体。要调查这些移民的语言现状，首先要对新移民人口组成进行调查和分析，从中抽样出具有代表性的主要语言群体，再从这些语言群体当中选择适量有效发音人，使用的参数依然是：[语言身份认同]、[是否被集体认定]、[深圳居住者]。但第一和第二个参数必须相对于新移民中不同群体的语言，比如表3。

表（3）：深圳市户籍人口的籍贯抽样

籍贯	2014年人数	1980年前人数	方言/语言
广东省紫金县	63503	779	客家话（粤台片：兴华小片）
广东省五华县	48610	482	客家话（粤台片：兴华小片）
广东省揭西县	47142	689	客家话（北部） 闽语（闽南区：潮汕片）（南部）
广东省潮阳市	46576	390	闽语（闽南区：潮汕片）
广东省普宁市	41972	250	闽语（闽南区：潮汕片）
广东省兴宁市	41491	276	客家话（粤台片：兴华小片）
广东省龙川县	40190	587	客家话（粤中片）
广东省陆丰市	39291	97	客家话（粤台片：新惠小片）（北部） 闽语（闽南区：潮汕片）（南部）
广东省惠来县	36604	146	闽语（闽南区：潮汕片）
广东省丰顺县	36082	170	客家话（粤台片：兴华小片）
广东省梅县	32156	403	客家话（粤台片：嘉应小片）
广东省大埔县	31613	198	客家话（粤台片：兴华小片）
广东省博罗县	29276	602	客家话（粤中片）

¹² 深圳新移民指特区成立后来移居深圳、并定居深圳的人。

广东省海丰县	26065	511	客家话（粤台片：新惠小片）（北部） 闽语（闽南区：潮汕片）（南部）
广东省河源市	23987	489	客家话（粤中片）
广东省惠东县	23619	1105	客家话（粤台片：新惠小片）
广东省陆河县	18447	39	客家话

表 3 中的户籍人口，是根据深圳市 2014 年人口数量最多的前 17 个籍贯（精确到县级）进行的抽样。户籍人口身份保证了他们均为深圳居住者。对比 1980 年前该籍贯的人数，可以推断他们大部分属于新移民。而使用的方言/语言是根据《中国语言地图集》（1987）来估算。由于不少新移民的第二代已经放弃母语方言而使用深圳新粤语¹³和普通话，因此，具体判定必须通过[语言身份认同]和[是否被集体认定]这两条参数。

4.3 小项目：深圳通用语

每一个地区都有自己的语言习惯，我们姑且把体现地区语言习惯的主要语言叫地区通用语。地区通用语对于当地而言是最常用的语言，也是当地陌生人见面时最倾向使用的语言¹⁴。对于中国（除港台地区）来说，通用语经常是带有当地语音特色的普通话。深圳目前的通用语就是体现当地语言习惯与特色（包括混杂粤语语码、卷舌平舌不对立等）的当地普通话。倘若深圳通用语具备独特的语音特点，容易被其他使用者辨认出来，那么，就可以认为深圳通用语具有了自己的、可以客观描述的语音系统，因而值得进行调查。但如果不具备这个先验条件，就意味着深圳通用语尚未定形，调查时必须按照 4.2 的方法将人口细分为各个语言群体进行语料采集。

假设深圳通用语确实具备了值得调查的语音特点，那么研究对象就是深圳居住的普通话使用者。在这个研究对象群体当中，要找出具有代表性的发音人可以借助于以下参数与定值。[长期居住]定值为【是】、[常与该地区社会各阶层的人交流]定值为【是】、[从事与语言相关职业]定值为【否】。前两个参数的作用相当明显，此不赘述。第三个参数的定值可排除语言教师、广播员等使用官方订立标准普通话的人。虽然这些被排除的人中也许有适合的发音人，但从具代表性的层面来看还是避开为好，因为按照参数找到的发音人应当是最理想和最典型的。

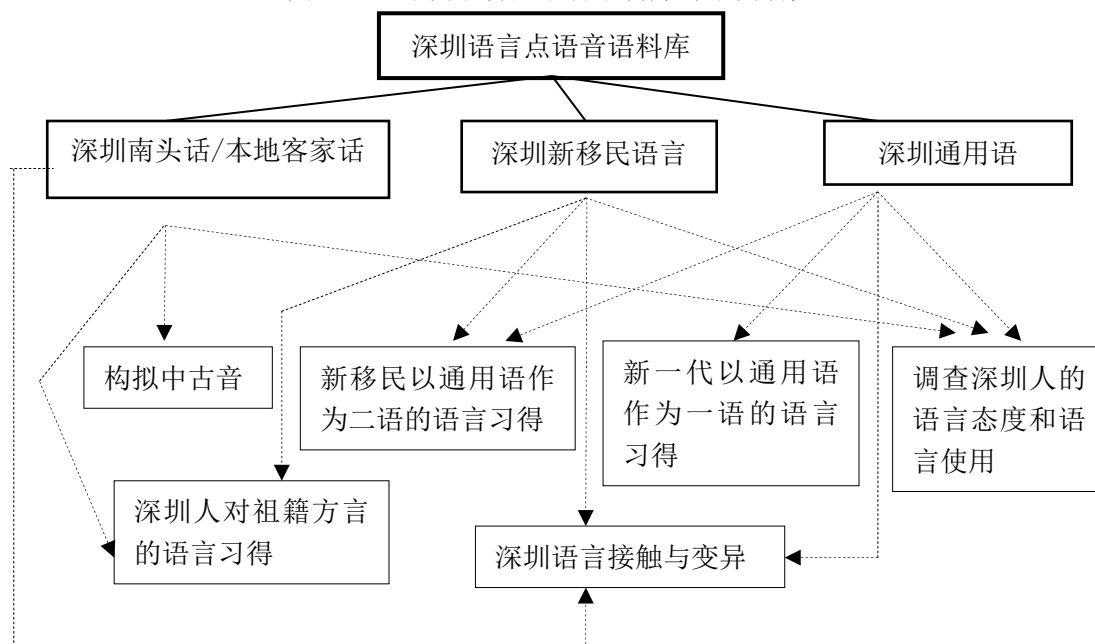
4.4 深圳语言点语音语料库的研究适用性

对深圳语言的语音进行调查而建立的“深圳语言点语音语料库”可以适用于不同研究目的。通过前面几个小项目可以看到，深圳语言至少包括三个层面：本地语言、新移民语言以及通用语。如果研究目的是构拟中古汉语，那么需要使用到的语料主要是南头话和本地客家话；如果研究目的是语言接触和语言变异，那么需要的语料将包括本地语言、新移民语言以及通用语；如果研究目的是语言习得，我们还需要区分是对祖籍方言的习得还是对通用语的习得（梁源、胡建慧 2015）、是一语习得还是二语习得。我们甚至可以使用语料库的录音作为刺激声进行语言测试或者实验，调查深圳人的语言态度和语言使用，等等。不同研究目的与语料之间的适用关系总结为图（2）。

¹³ 关于深圳新粤语的情况，看看梁源、吴育珊（2011）。

¹⁴ 虽然普通话在南京社区的接受度接近 100%，根据徐大明、付义荣（2005）的调查，70%的南京人与陌生人交际时仍然使用方言。

图 (2): 深圳语言点语音语料库与不同研究



五、结论

本文探讨在语言学研究中如何选择具有代表性的发音人。通过不同语言学研究的示例，本文提出：作为研究对象的有效代表必须符合研究所需的代表属性。最为直接表述代表属性的方法是参数。参数的设定能外显研究对象的主要特征，并且，参数设定越精细就越能找出研究对象的有力代表；同理，越有力的代表应当更符合这些参数。当然，在适当的情况下应该根据年龄、性别、教育水平等方面来平衡发音人数量，但不可盲目强求（比如研究儿童语言习惯就不能强行要求老中青三个年龄层都有代表）。厘清代表属性让我们更明确地看到常用参数的适用性。

依据研究目的而订制代表属性的手段一直内在于前人研究的方法论当中。本文将这些概念外显出来，从而使得方法论上的优缺点更容易得到检验。

参考文献

- 包联群. 杜尔伯特蒙古族的语言使用和语言态度. *中国社会语言学*. 2007, (2): 117-129.
- 陈平. *中国居住文化*. 香港: 中华书局(香港)有限公司. 1992.
- 丁俊清. *中国居住文化*. 上海: 同济大学出版社. 1997.
- 甘于恩. 四邑话: 一种粤化的混合方言. *中国社会语言学*. 2003, (1): 95-100.
- 顾钦. *语言接触对上海市区方言语音演变的影响*. 上海师范大学博士学位论文. 2007.
- 焦成名. 上海土著学生语言行为报告. *语言文字应用*. 2009, (1): 27-37.
- 李小凡、项梦冰. *汉语方言学基础教程*. 北京: 北京大学出版社. 2009.
- 梁源、胡建慧. 多重语体下的母语方言不完全习得与语音变异——以深圳为例. 中国语言研究与教学学术研讨会, 香港: 香港教育学院, 2015年5月9日.
- 梁源、吴育珊. 从语言格局看深圳新粤语音系. *暨南学报*. 2011, (4): 132-137.
- 邵敬敏、吴立红. 香港社区英文词语夹用现象剖析. *语言文字应用*. 2005, (4): 37-44.

- 沈同. 上海话老派新派的差别. *方言*. 1981, (4): 275-283.
- 汤志祥、林建平. 深圳粤语的分布以及代表话南头话的音系. *粤语研究* (创刊号). 2006: 62-73.
- 徐大明、付义荣. 南京“问路”调查. *中国社会语言学*. 2005, (2): 143-150.
- 许宝华、陶寰. *上海方言词典*. 南京: 江苏教育出版社. 1997.
- 许宝华、汤珍珠、陈忠敏. 上海地区方言的分片. *方言*. 1993, (1): 14-30.
- 许宝华、汤珍珠、钱乃荣. 新派上海方言的连读变调. *方言*. 1981, (2): 145-155.
- 许宝华、汤珍珠、钱乃荣. 新派上海方言的连读变调 (二). *方言*. 1982, (2): 115-128.
- 许宝华、汤珍珠、钱乃荣. 新派上海方言的连读变调 (三). *方言*. 1983, (3): 197-201. 袁家骅. *汉语方言概要* (第二版). 北京: 文字改革出版社. 1989.
- 游汝杰. 上海话在吴语分区上的地位——兼论上海话的混合方言性质. *方言*. 2006, (1): 72-78.
- 中国社会科学院、澳大利亚人文科学院. *中国语言地图集*. 香港: 朗文出版社. 1987.
- 中国语言资源有声数据库建设领导小组办公室. *中国语言资源有声数据调查手册: 汉语方言*. 北京: 商务印书馆. 2010.
- Bao, Lianqun (2007). Duerbote mengguzude yuyuanshiyong he yuyan taidu [Language Use and Language Attitude of the Dorbed in Heilongjian]. *Zhongguo shehui yuyanxue* [The Journal of Chinese Sociolinguistics], vol.2007(2): 117-129.
- Chen, Ping (1992). *Zhongguo juzhu wenhua* [Culture of Chinese Living]. Hong Kong: Chung Hwa Co..
- Ding, Junqing (1997). *Zhongguo juzhu wenhua* [Culture of Chinese Living]. Shanghai: Tongji University Press.
- Gan, Yu'en (2003). Siyihua: yizhong yuehuade hunhe fangyan [Siyi dialect: A Yue-influenced Pidgin]. *Zhongguo shehui yuyanxue* [The Journal of Chinese Sociolinguistics], vol.2003(1):95-100.
- Gu, Qin (2007). *Yuyan jiechu dui shanghaishiqu fangyan yuyin yanbian de yingxiang* [The Influence of Language Contact on the Phonetic Evolution of the Dialects in Shanghai Urban Districts]. PhD dissertation, Shanghai Normal University.
- Jiao, Chengming (2009). Shanghai tuzhu xuesheng yuyan xingwei baogao [A Report on Language Behavior of Shanghai Native Students]. *Yuyanwenzi yingyong* [Applied Linguistics], vol.2009(1):27-37.
- Li, Xiaofan & Xiang, Mengbing (2009). *Hanyu Fangyanxue jichu jiaocheng* [Foundations in the study of Chinese Dialectology]. Beijing: Peking University Press.
- Liang, Yuan and Hu, Jianhui (2015). Duochongyutixiade muyu fangyan buwanquan xide yu yuyin bianyi: yi shenzhenweili [Underacquisition and Variation of native dialects in multiglossic environment: the Shenzhen Case]. *Zhongguo yuyan yanjiu yu jiaoxue xueshu yantaohui* [Symposium of Chinese Language Study and Pedagogy]. The Hong Kong Institute of Education, September 2015.
- Liang, Yuan and Yushan Wu (2011). Cong yuyan geju kan Shenzhen yueyu yinxi [Exploring the phonological system of Shenzhen Cantonese from perspective of language pattern]. *Jinan xuebao*, vol.2011(4):132-137.
- Shao, Jingmin and Wu, Lihong (2005). Xianggang shequ yingwen ciyu jiyong xianxiang pouxu [Analysis of the Phenomenon of Chinese Mixed up with English Words in the Community of Hong Kong]. *Yuyan wenzi ying yong* [Applied Linguistics], vol.2005(4):37-44.

- Shen Tong (1981) Shanghaihua laopai xinpaide chabie [Distinctions in Old and New Shanghai]. *Fangyan* [Dialects], vol.1981(4):275-283.
- Tang, Zhixiang and Lin, Jianping (2006). Shenzhen yueyude fenbu yiji daibiaohua nantouhuade yinxi [Distribution of Shenzhen Cantonese and the phonology of Nantou dialect]. *Yueyu yanjiu* [Linguistic Study], vol.2006(1):62-73.
- Xu, Daming and Yirong Fu (2005). Nanjing “wenlu” diaocha [“Asking-the-way” Survey in Nanjing]. *Zhongguo shehui yuyanxue* [The Journal of Chinese Sociolinguistics], vol. 2005(2):143-150.
- Xu, Baohua and Tao, Huan (1997). *Shanghai fangyan cidian* [Dictionary of the Shanghai Dialect]. Nanjing: Jiangsu Education Press.
- Xu, Baohua, Tang, Zhenzhu & Chen, Zhongmin (1993). Shanghai diqu fangyande fenpian [Dialect sub-regions in the Shanghai region]. *Fangyan* [Dialects], vol. 1993(1):14-30.
- Xu Baohua, Zhenzhu Tang and Qian Nairong (1981). Xinpai Shanghai fangyande liandu biandiao [Tone sandhi in New Shanghai]. *Fangyan* [Dialects], vol. 1981(2):145-155.
- Xu Baohua, Zhenzhu Tang and Qian Nairong (1982). Xinpai Shanghai fangyande liandu biandiao (2) [Tone sandhi in New Shanghai 2]. *Fangyan* [Dialects], vol. 1982.2:115-128.
- Xu Baohua, Zhenzhu Tang and Qian Nairong (1989). Xinpai Shanghai fangyande liandu biandiao (3) [Tone sandhi in New Shanghai 3]. *Fangyan* [Dialects], vol. 1983(3):197-201.
- Yuan, Jiahua (1989). *Hanyu fangyan gaiyao* [An Introduction to Chinese Dialects](2nd edition). Beijing: Wenzigai Press.
- You, Rujie (2006). Shanghaihua zai wuyu fenqushangde diwei: jianlun shanghaihuade hunhe fangyan xingzhi [On Classification and Mixed Nature of the Shanghai Dialect]. *Fangyan* [Dialects], vol.2006(1):72-78.
- Chinese Academy of Social Sciences & Australian Academy of the Humanities (1987). *Zhongguo yuyan dituji* [Language Atlas of China]. Hong Kong: Longman.
- Zhongguo Yuyan yousehgziyuan shujuku jianshe lingdaoxiaozu (2010). *Zhongguo yuyan ziyuan yousheng shuju* [Audio Data Resource of Chinese Languages]. Beijing: Commercial Press.
- Chomsky, Noam . 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press (2nd edition).
- McCawley, James D. 1978. What is a Tone Language? In Victoria Fromkin (ed.) *Tone: A Linguistic Survey*. New York: Academic Press, pp. 113-131.
- Siegel, Jeff. 1993. Dialect contact and koineization. *International Journal of the Sociology of Language*, vol.99: 105-121.
- Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Blackwell.
- Wee, Lian-Hee. 2009. Englishization as an aspect of building the Singapore identity. In K.K. Tam ed. *Englishization in Asia*. Open University Hong Kong Press, pp. 46-69.

Parameters of representativeness: subject selection and data integrity

LIANG Yuan

Dept. of Chinese Language Studies.
The Hong Kong Institute of Education
yliang@ied.edu.hk

WEE Lian-Hee

Dept. of English Language and Literature.
Hong Kong Baptist University
lianhee@hkbu.edu.hk

Abstract: As an empirical study, linguistic investigations rely heavily on the selection of suitable subject speakers as representatives of a language community, which in turn underline the necessity of determining representativeness. Through explicit description of the parameters that define a target population, this paper argues that (i) representativeness can be most straightforwardly captured using the same parametric description; and (ii) the degree of representativeness of the subject increases with the degree of satisfaction of the parameters. On this footing, this paper exemplifies the fundamental quality of the parameters of representativeness across different linguistic research topics, purposes and even methodologies. Evidently, the clarification on what counts as representative through the use of parametric description is a founding stone to ensuring reliability of language data in the establishment of any linguistic theorizing.

Keywords: representativeness, parameter, subject speakers, language data reliability

About the authors

Yuan LIANG is Assistant Professor at the Department of Chinese Language Studies. Her research areas are in Chinese dialects and their phonologies, and also in language development and change. Her publications touch mainly on Yue dialects, including recent new Cantonese communities in Shenzhen.

梁源，語言學博士，現為香港教育學院中國語言學系助理教授。在汉语方言音系、语言形成与历时变化等研究领域發表數十篇文章，主要涉猎粤语研究，包括深圳新生粤语。

Lian-Hee WEE is Associate Professor at the Department of English Language and Literature. His linguistics research focuses on the phonological aspects of Chinese and Asian varieties of Englishes, though his publications also extend to teaching. His most recent volume is *Capturing Phonological Shades within and across Languages* (Cambridge Scholars Publishing, co-edited), adding a list of works that already includes four other volumes, and numerous journal papers and book chapters.

黄良喜，英国语言文学系副教授，主要从事汉语及亚洲新生英语音系研究，偶尔涉及语言教学，著有论文数十篇、编著图书五本，最近期著作为合编的论文集 *Capturing Phonological Shades within and across Languages*。