

谁是 Mr. Tan：揭秘华人姓氏的英文拼写

陈方豪

2021/7/15

一、问题的产生——从象形文字到拼音文字

在英语已经成为国际通用语言的当下，大家和外国友人打交道的场合与频率不断上升。越来越多的人在自己的微信号上同时标注自己的中英文姓名，以方便跨文化的交流。如果你的微信好友里有较多来自港澳台、东南亚的华人朋友，用心观察不难发现，虽然大家可能具有相同的中文姓氏，但是英文的拼写却大相径庭，很难直接从一个人姓名的英文拼写中判断出他/她的中文姓氏。

具体而言，有以下两种情形阻碍了从英文到中文的顺利转换：**同音不同字，同字不同音**。例如，同样拼作 Chow，既可以是周（周星驰，Stephen Chow），也可以是邹（邹至庄，Gregory C. Chow）；同一个“张”，也可以衍生出各种各样的拼写：Zhang、Cheung（张五常，Steven Cheung）、Chang（张震，Chang Chen）、甚至是 Chong。



图 1 两位都是 Mr. Chow (左为电影明星周星驰，右为经济学家邹至庄)

环顾全球，我们并不是唯一受此困扰的国家。以我们的邻邦日本为例，女生的名字“Rika”，其对应的汉字可以是“梨花”、“里香”、“梨香”；同一个汉字“惠美”则既可以是“Emi”，也可以是“Megumi”。仔细想来，其中要害在于我们和日本都广泛使用汉字这一象形文字，又为了与世界同步将象形文字转换为了拼音文字，但这个转换并没有得到协调一致的安排。欧美国家大多使用拼音文字，所写即所读，所以并不存在这个转换的问题。归根究底，这一问题产生的原因在于**拼音文字和象形文字的映射并不是一一对应的**。

那么问题来了，华人姓氏的这种拼音文字和象形文字的错乱对应关系是如何形成的呢？

二、家乡的影响——方言

我国对外移民的历史由来已久，最早可以追溯到唐宋时期，时至今日已经形成了遍布五大洲 160 多个国家和地区的庞大的海外华人华侨群体。根据统计，其规模可达三千万之众^[1]。特别地，大多数海外华人的祖籍集中于沿海的福建、广东、浙江三省。这三个地区共同的特点

在于方言的丰富而多变，且其发音与中国大部的官话差别较大（见图 1）。这为后来姓氏拼写的混乱局面埋下了伏笔。



图 2 中国方言地图

鸦片战争之后，清政府被迫打开国门，前往海外谋生的移民数量骤增，也增加了与国际社会交流的需求，其中文字的统一标准化便是首当其冲一点。为了便利中外人群拼读中国的人名和地名，历任英国驻华大使、剑桥大学汉学教授的威妥玛 (Thomas Francis Wade) 于 1867 年以罗马字母为汉字注音，发明了威妥玛式拼音 (Wade-Giles System) ^[2]。在很长一段时间内，它被用作中国和国际上通行的中文拼音方案。例如，广为流传“常凯申”——即蒋介石的错译名，其英文拼写 Chiang Kai-Shek，就是采用了威妥玛式拼音。

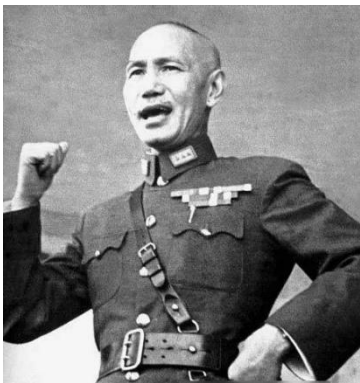


图 3 “常凯申” (Chiang Kai-Shek)

但偏偏威妥玛式拼音作为一项标准并没有被全面执行。另外一条有重叠但又平行的拼写标准——邮政式拼音 (Postal System)，也在实践中被广泛采用。邮政式拼音的设立主要目的在

于便利国际邮电通信，于 1906 年春季在上海举行的帝国邮电联席会议中被通过。邮政式拼音规定，为了适合打电报的需要，拼写中不采用任何附加符号（例如送气符号等）。并且更重要的是，在广东、广西、福建等地区，一律按当地的方音拼写^[3]。这可就一下乱了套。因为南方诸省方言的复杂性（吴语、闽语、客家话、粤语等），大家根据其家乡方言的不同，同一姓氏可以有多种拼写、多种姓氏也可以共用一种拼写。

有趣的是，因为方言的差异，通过拼写可以大致推测一个海外华人的祖籍是哪里。例如同一个英文拼写 Ng，如果是“吴”的话，那么他祖上大概率是广东人，如果是“黄”的话，那么他祖上大概率是福建人。同一个中文姓氏陈，如果拼写为 Chan，那么大概率他是广东人，例如成龙的英文名为 Jackie Chan；而如果拼写为 Tan，那么几乎可以肯定他是福建人的后代了——因为 Tan 是闽南话“陈”的发言（读音为“dan”，第三声）。著名爱国华侨、厦门大学校长陈嘉庚的英文名字拼写就为 Tan Kah Kee。



图 4 爱国华侨陈嘉庚（Tan Kah Kee）

直到 1958 年 2 月 11 日，新中国全国人大批准颁布《中文拼音方案》后，威妥玛拼音和邮政式拼音才被停止使用。中国大陆至此终于有了统一的汉语拼音标准。自那之后，往海外移民的群体逐渐统一地采用汉语拼音系统。因此，如果你看到一位海外华人的姓名使用了汉语拼音系统，那么他/她大概率是新近的移民。

三、海外的影响——移入国的文字习惯与政策

在东南亚，华人文化的传承做的很好，从建立中文学校，出版中文报纸，组织华人社团，到组织各种利益祭祀，各种各样的组织和活动层出不穷。很多海外华人移民历经数代之后仍然保持着对于中华民族的认同感^[4]。

但毫无疑问，海外华人的姓氏拼写显然也会受到移入国的影响。1965 年前，印尼华人的姓氏拼写普遍采用汉语方言发音，只不过由于印尼曾经是荷兰殖民地的关系，在正字法上具备较多荷兰语特征，譬如粤籍的陈姓在新马是 Chan，印尼就是 Tjan，闽南籍的林在新马是 Lim，印尼则为 Liem，如著名印尼羽毛球选手林水镜就是 Liem Swie King^[5]。

在 1960 年代，海外华人的活跃引起了反作用。很多东南亚国家的本地民众认为，那些根据“血统主义”仍然持有中国国籍的华侨是潜在的“第五纵队”，他们效忠的对象是中华人民共和国。1965 年苏哈托推行反华法案，强迫华人放弃汉姓，改用印尼化的姓。祖籍福建福清的

著名侨商林文镜就有一个印尼化的名字 Sutanto Djuhar。

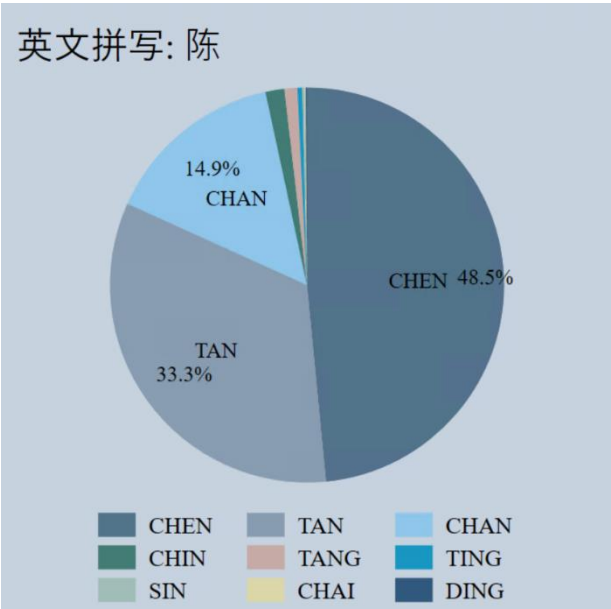
四、识别侨商的姓氏

那么问题来了，当我们只能得知一位海外华人的英文拼写时，是否能对他/她的姓氏有所判断？这正是我们的一篇研究^[6]中所面临的挑战：我们想要识别改革开放之后到大陆投资的侨商的姓氏，但并不是所有的侨商都以中文姓名进行工商登记。**我们的解决问题的思路是：让数据说话（let the data speak）。**

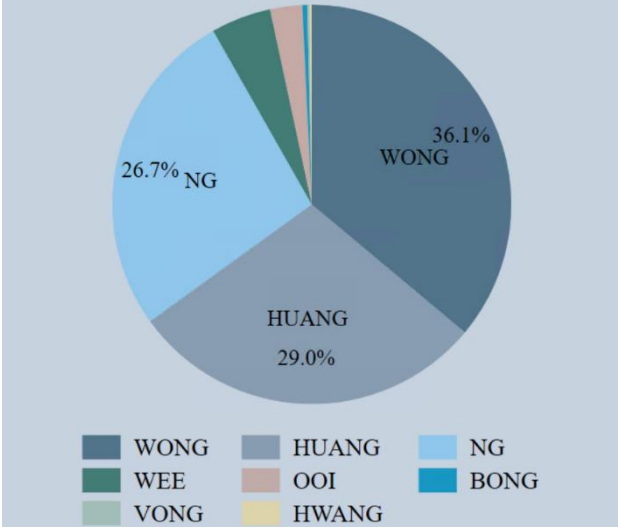
我们使用了截止至 2015 年初的全国工商注册数据，其中，我们筛选出所有在中国大陆曾经注册过的外资企业（包括港澳台企业），并提取出外资企业相应的人员信息。我们按照以下两个标准进一步将人员限制在海外华人群体：(1) 具有华人姓氏；(2) 以非大陆身份证登记。对于海外华人企业人员，我们提取其姓名信息以供分析。

可能由于历史的原因，工商注册数据中的姓名信息登记十分不规范，但反而为我们的研究提供了独特的优势。具体来说，数据中的姓名变量可以大致分为三种：纯中文（如周星驰）、纯英文（Steven Chow）、中英混合（周星驰 Steven Chow）。对于第一类和第三类的姓名数据，我们可以比较简单地提取出华人的姓氏，只有对于第二类姓名数据我们面临前文所述的拼写并非一一对应的问题。乍看之下我们似乎束手无策，但其实中英混合的姓名数据为我们构造侨商群体的姓氏的中英文对照提供了桥梁。用现在热门的机器学习的语言来说，**中英混合的姓名数据提供了珍贵的“语料库”。**

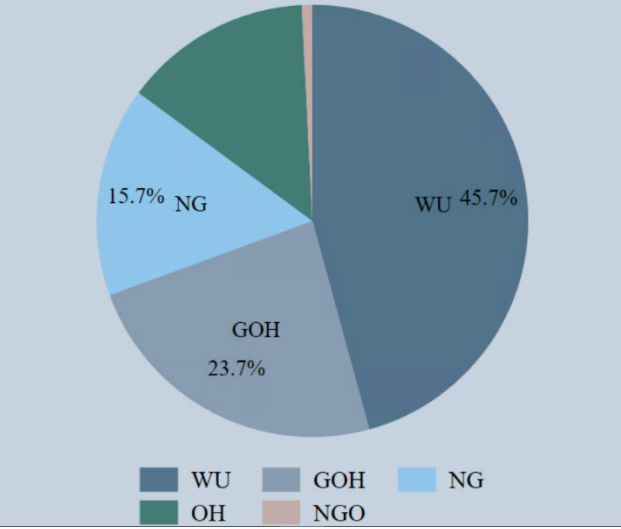
经过一系列算法的操作，我们从中英混合的姓名数据中提取这部分海外华人姓氏各自的中英文拼写，由此我们可以对每一种姓氏的英文拼写计算出其对应中文姓氏的概率。图 5 展示了具有代表性的几个姓氏中英文的相互对应关系。



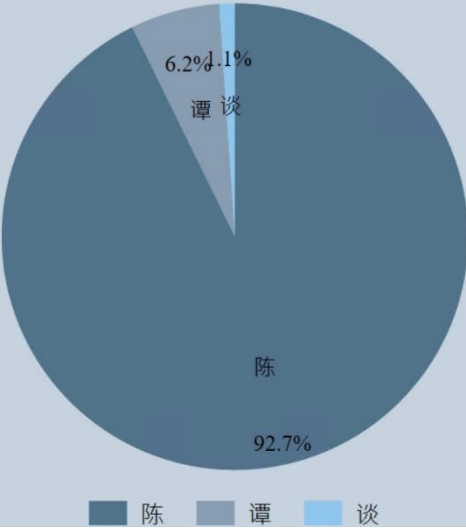
英文拼写: 黄



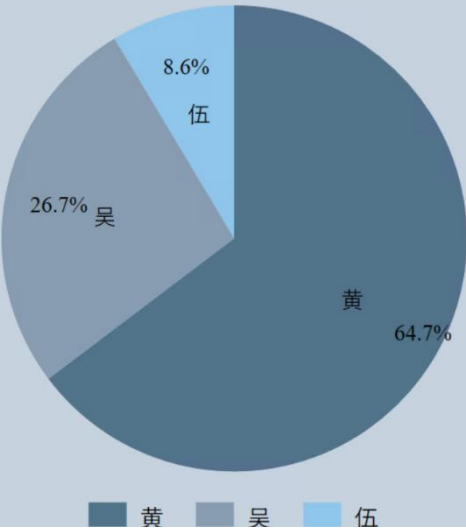
英文拼写: 吴



中文姓氏: TAN



中文姓氏: NG



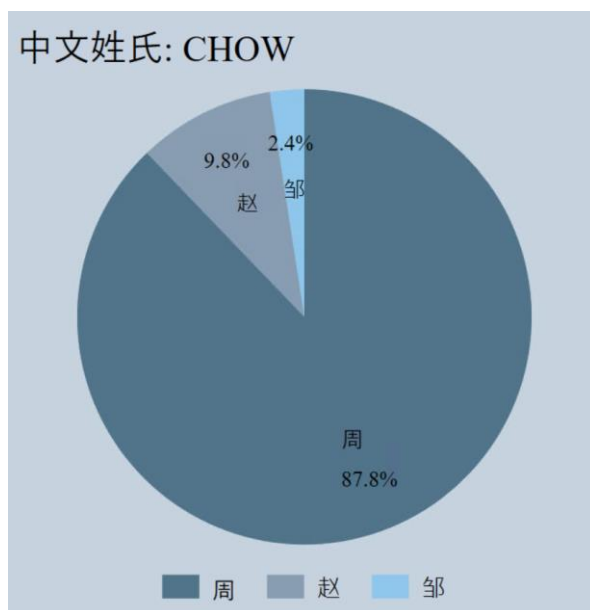


图 5 华人姓氏中英文对应关系

*数据来源：全国工商注册数据

下次当你遇见一个名片上印着 Mr. Tan 的海外华人，那么他大概率是一位陈先生，而较小可能是一位谭先生！

参考资料

- 【1】 《海外华侨华人与中国改革开放》，任贵祥，中共党史出版社
- 【2】 威妥玛式拼音法，百度百科
<https://baike.baidu.com/item/%E5%A8%81%E5%A6%A5%E7%8E%9B%E5%BC%8F%E6%8B%BC%E9%9F%B3%E6%B3%95/5404053>
- 【3】 邮政式拼音，百度百科
<https://baike.baidu.com/item/%E9%82%AE%E6%94%BF%E5%BC%8F%E6%8B%BC%E9%9F%B3/10240424>
- 【4】 《他者中的华人：中国近现代移民史》，孔飞力，江苏人民出版社
- 【5】 谁是 Mr Ng：海外华人如何拼写姓名？
<https://cul.qq.com/a/20141228/008958.htm>
- 【6】 Fanghao Chen, Ruichi Xiong, Xiaobo Zhang. Familiar Strangers: Overseas Chinese, Lineage Connection, and Foreign Firms in China. Working Paper.