

大数据时代的外语教育科研

刘润清

(北京外国语大学,北京,100081)

作者说明:本文系《外语教育中的科研方法》(修订版)一书的第十二章第四节。成稿之后,得知《当代外语研究》的杨枫主编正在研究大数据问题,就发给他向他请教。没想到被“看中”了:他们乐意提前发表本节。本章第一节讲大数据的特点;第二节谈机器学习;第三节说大规模网络教育前夕;第四节论大数据时代的外语教育科研。其中的逻辑是:先介绍大数据时代的特点;后关注大数据时代如何出现:由于计算机朝着智能方向突飞猛进——从机械地执行命令到能推理学习,再加上我们处理信息的能力空前提高,智能计算机给网络教育带来了巨大变化——各国著名大学陆续推出网络公开课;最后才回到大数据给外语科研带来的机遇和挑战。整章的讨论尽量回避复杂的技术问题,只为外语教师提供能看懂的背景知识。希望这样的讨论能让我们走在未来的前面。

[中图分类号] H319.1

[文献标识码] A

[文章编号] 1674-8921-(2014)07-0001-06

[doi 编码] 10.3969/j.issn.1674-8921.2014.07.001

现在讨论一下在大数据时代,外语教育科研会发生怎样的变化。基于前面三节的讨论,我们有理由相信,在今后二三十年里,外语教育(可能主要是英语教育)会发生根本性变化。由于教学模式、学习模式的变化,教师认知和学生认知都会发生变化,我们科研的方向和重点(题目、领域、所跨学科)也会有根本性的改变。由于计算机技术的高度发达和大数据时代的到来,科研方法和统计方法也会有根本性的改变。本节将讨论网上外语教育的变化趋势和实现网络教育后,外语科研中可能发生的重大改革。

1. 教师的角色转变

教师不再是知识的唯一占有者和提供者,而变为课堂的组织者和学习的引导者。英语描写这一变化更有趣:教师从 God 变成 Guide,从 Sage on the stage 变成 Guide on the side,从 teacher 变成 helper、councilor、facilitator 等。组织者和引导者并不好当:教师要对学习内容十分熟悉,并知道网上有何种公开课视频可以提供;同时最好能了解学习者的特征(性格、认知方式、学习方法、学习目的、特殊经历等),能比较准确地告诉他在特定阶段选择什么样的视频,只有这样,才能实现教学的个性化。当然,引导者必须熟悉学习内容,否则引导就是空话。换句话说,教师备课量不会减少,只是课堂呈现量减少了;而准备须回答的问题时,思维范围会加宽,内容也

会多种多样、千差万别。此外,教师对教育技术的掌握也就成为理所当然的了。

少部分教师会成为公开课视频的演讲者或导演。提供公开课的教师,其教学理念要彻底转变:他要知道网络公开课不是课堂搬家;他要面对几万、十几万世界各地的水平不同、兴趣不同、学习方法不同的选修学生;内容之精准、繁简之恰当、表达之清晰,以及吸引学生之牢固而持久等,都是对他极大的挑战。例如,视频要模块化——每个主题只占3~5分钟,方便学生利用零碎时间观看;重要知识点处须嵌入交互式问题,方便学生及时检查学习效果;涉及背景资料时要嵌入超链接,方便学生先链接到网上查看资源,然后立刻返回原处。将来以课程为载体的视频会不断地被生产、积累、淘汰和沉淀。可以肯定地说,通过互联网上的编辑与合作,在不久的将来,世界上每个领域、每门课程、各个层次的视频,都会出现一些经典之作,保留下来。它们将取代教师的课堂授课,成为最受学生欢迎的学习材料。随着在线平台的推广和普及,会有越来越多的教师把自己的授课视频放到网上。美国有个 Udemy 网站,教师可以根据自己发放视频的点击率获得报酬。除了课程视频,教案也可以放到网络进行交易。乔治亚州一位学前班老师的教案成交量已达79万美元,月收入6万美元。教案的交易还出现了由学校牵头组织团购,统一发给全校老师参考使用的形式。

还有的教师瞄准“引导者”角色,推出了提供专门辅导的家庭网站。学习者可以按自己的专业从网站公布的专家中选择自己喜欢的课程辅导者,通过网上视频辅导,10分钟的视频交流后辅导服务开始按分钟计费。学习者和辅导者可以用很低的成本和

作者简介:刘润清,北京外国语大学中国外语教育研究中心顾问、教授,兼任国家高等教育自学考试英语专业委员会顾问,曾任北外外国语言研究所所长、中国外语教育研究中心主任、《外语教学与研究》副主编、中国英语教学研究会秘书长。

很少的时间,实现有效对接,构建一对一的学习环境。涂子沛(2014)^①称之为“微学校”。由于计算机和网络无处不在,“微学校”可用随时随地地搭建。

多年来,我们总听说高排放、高污染、高能耗的重工业正经历着“结构性调整”。现在,轮到教育经历“结构性调整”了。这一调整不仅更换结构,更重要的是革新理念和生产方式。不过,看来教师们不用担心会丢掉饭碗,只是要调整教学理念,适应新的教学模式,提前做好准备,说不定我们能在新的环境下大显身手。

2. 学习者角色的转变

学习者成了真正的中心:他自己决定学什么、何时学、如何学、在哪学、学多少、学多快。学习者的主动性要充分发挥出来:首先他不能是计算机盲,而是要娴熟掌握网络技术。第二,他要主动了解有关课程的一切信息,包括他自己的责任、教师的责任、管理员的责任等,以便决定自己选修的专业和课程。第三,他必须在网络通知的限制之内,完成课程作业,按要求修够学分等。第四,他有大把的时间和绝对的主动权,个人的决心和努力决定了自己的未来。这要求学生有高度自觉性、自律能力、毅力、自我管理能力等;放任自流、自由散漫、不能自律的学生会遇到困难。也会有极个别意志薄弱或娇生惯养的学生,被网络上的八卦信息所吸引,迷恋于不三不四的游戏、大片、赌博等,成为大数据时代的牺牲品。

师生角色的变化比表面看上去要深刻得多。从前的学徒制和私塾制里,讲究“师徒如父子”,是很不平等的关系,知识的传授是一对少,老师是绝对权威;后来的学校制,开始了师生关系,解除了不平等的“父子”关系,打破了师生的小比例关系;但一对多的教授方式和工厂生产线式的批量生产,无法给学生提供个性化、持续性的指导和评价,而且,教师仍然把持知识的权威地位。用一个模子刻制零件的批量生产式也扼杀了学生的个性和创造性。大数据时代不同了:教师从“父亲”、“权威者”变为指导者、帮助者、协商者——师生间的关系平等多了。学生成了中心,是被服务的对象,无限扩大了自我发展的空间,不再被放入同一个模子里压榨成统一型号的产品。

3. 学习材料的转变

可能纸质教材和参考书会减少(但不会彻底消失),音频、视频资料会大大增加;电视使用会减少,MP3、MP4、笔记本和iphone等会更加普遍。2013年的全国国民阅读调查结果显示,传统纸质媒介中,2013年成年国民对图书、报纸和期刊的接触时长均有不同程度的减少;新兴媒介中,上网时长和手机阅读的接触时长呈增长趋势。中国新闻研究院院长魏

玉山表示,超九成的数字化阅读方式接触者表示阅读电子书后就不再购买其纸质版,这一数据连续四年持续上升,反映出数字出版对传统出版物的冲击作用不断加强。学习材料的可及性极度增长:数字化材料应有尽有,有文字的、声音的和视频的,包括百科全书、词典、虚拟场景、虚拟地图、虚拟世界。因为媒介的彻底转变,将来教材编写的原则也会发生变化。比如,原来繁琐的文字阐述和描写可能被模拟视频所代替。只依靠视觉输入的文字材料会减少,要求手、眼、耳、口并用的材料会大大增加。技能层次类的课程大多具备可拷贝性,可以交给大投入的电影制作模式完成。文化性和学术性的教材制作起来挑战性大些,可能要经过几年的试验阶段。最重要的学习材料是公开课视频。每节时长为2个小时的课程视频,一般会被分解为若干个时长为8~15分钟的知识模块(modules)。这是为了让学生自己驾驭课程学习,学生自己在与学习材料的互动当中选择属于自己的路线和速度。只有按教师要求完成一个模块,才能进入下一个模块的学习。每个知识模块里包括的课程活动有:8~15分钟的讲座视频、嵌入式小测验、课后的测验和考试、同伴评价题目、编写程序作业、调查题目、课程讨论区、谷歌视频群聊,等等。当然,还有上传给学生的额外的背景材料。认知心理学告诉我们,回溯检索性学习(the retrieval task)可以提升学生的学习质量。MOOCs(massive open online courses)(可译为“慕课”=屏幕上的课程)课堂教学由于在讲座中成功地嵌入了测试问题(embed quiz questions),使得学习者投入的热情大大增加。但必须提到的一点是,在MOOCs中,这些测验的结果并不计入学生修课成绩,否则学生会变得焦虑,对课程及学习内容产生畏难情绪,甚至本来很想学习这门课程的都因此而弃选。

4. 学习环境的变化

大学校园的界限会变得模糊甚至被打破;学生在教室里学习的时间减少,而随时随地地学习会成为普遍现象;师生之间面对面的交流不多,而在网上视频互动的机会大大增加。来自不同国家的学生会在网上相遇、互动,成为学习伙伴——求学过程的国际化程度会发展到不可想象的地步。那时会有多个大网站,提供各个学科的各种资料,相当于多个大图书馆。输入关键词,想查阅的资料就会呈现于屏幕。只要学生能有上网的机会,就能克服地域限制,就能实现教育的相对公平,受教育机会相对均等。网络公开课有着极富生气的讨论区,选择同一门课程为数以万计的同学,以他们特有的方式互帮互助。有些授课教师会通过网络社交平台积极参与到同学们的讨论中去,有的教师则会邀请助教对同学们的

讨论和问题进行反馈和必要的提醒。前面提到,会有“微学校”或称“家庭辅导网站”出现,学生可用选择自己看中的辅导老师就某一门课程进行辅导,那绝对是一对一、个性化的交流,按时间付费即可。以社交媒体为特征的 Web2.0 技术已经把人际交流互动推向登峰造极的应用。现有的信息技术可以通过网络视频和智能软件,为每个学习者搭建一个学习情境,像活生生的人一样对学习者的学习行为提供持续性的诱导、评价和支持——例如,发出提示和建议,指出错误,教授学习方法,帮助养成良好的学习习惯等等。部分 Coursera 的学员已经在使用 Google Hangouts on Air Weekly video, Twitter feed Facebook page 进行在线聊天,制作在线教室,分享学习体验,交换意见和看法等。部分近距离的学员甚至会通过线上约定方便的时间、地点见面,讨论学习情况。还有,MOOCs 作为一种基于科学学习理论和经过精心设计的教学模式,首次把学习的本质、学习过程、学习动机、学习策略等学习科学的基本规律运用到教学当中,大大提升了学习体验和学习质量。放眼看去,几十年后,大学校园的围墙将不复存在;课程、教案、辅导、讨论统统被搬到网上;教师都在网上“上班”。这对去教育行政化和官僚化大概会起到不小的作用。以后的网络管理和网络政务不知道可以精简多少百万公务员。

5. 评估和测试的变化

首先,笔纸测试(pen and paper test)会逐渐减少,代之以机测、网上测试;机测的优点之一就是它能自动调整项目的难易度(参见项目反应论 item response theory);好处之二是它不仅记录下考生的答案,同时记录下答卷的过程——改动几次才“碰”对答案,并把过程错误加权后计入分值。第三,随着机器智能的提高,外语考试不再大量依靠多项选择题,机器可以对口语表达和作文进行评分。多年来难以对主观试题评分的棘手问题会逐渐解决,并不断完善。以后的面试,考生不用再跋山涉水,“帷幄私宅,即可决胜千里之外”。口试还可以录像,评分时参考视频效果;甚至对视频进行多模态分析(Multimodal Analysis)。一旦机器可以评阅主观试卷,“人情分”、“成见分”就会随之消失。考试作弊也会杜绝——机器上的身份认证、可视镜头、网上监控等会比安保人员更“六亲不认”。一旦发现作弊的机器,立刻屏蔽,三年之内不能登录(比方说);枪手躺着也得中枪。顺便说一句,那时不用几万人、几十万人同时考试;每个人都可随时考试,立刻知道成绩。网测的试卷也不同于目前的纸笔试卷。一定是听说读写译综合考试,语言地道自然,真实场景再现,图文声并茂,让学生有身临其境之感。再往后,机器会

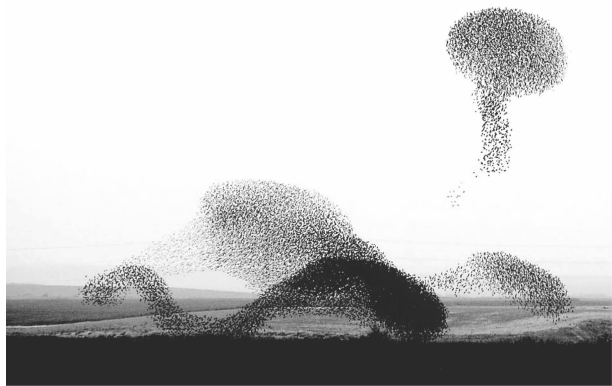
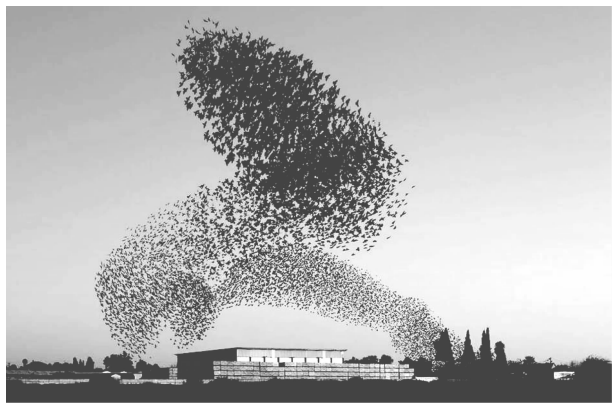
让考生“进入”虚拟场景,与里面的人物对话、解决问题、完成任务等,那才叫考查语言运用。

不难想象,在上述教学背景下的科研一定与现在大不相同。目前可以想到的大概包括以下几点:

(1) 大数据对哲学观点的冲击。近些年来,经验主义受到不少批评,因为人类对事物的感知极为有限,靠人工收集数据几十年充其量还只能算“样本”(sample)。自从数据库的出现,人们开始对经验主义产生新的认识:瞬间获取海量数据,从量变跳到质变(见本章第一节),让研究者突然观察到从来不敢想象的发现。维克托·迈尔-舍恩伯格和肯尼思·库克耶在《大数据时代》(Mayer-Schonberger & Cukier 2013)^②中谈到三大变化:第一,现在的科研能分析更多的数据,有时甚至是关于某个现象的全部数据,因此不再依赖随机采样。他们提出,我们的目标是“样本=总体”(这一“野心勃勃”的目标只有在大数据时代才有可能)。第二,我们拥有的数据如此之多,所以不再热衷于追求精确度;当然,也不是放弃精确,只是不再迷恋于它;适当忽略微观层面会让我们在宏观层面有更好的洞察力。第三,有了以上两个变化,我们不再热衷于寻找因果关系,而更注重相关关系。这三条会彻底改变我们在科研问题上的本体论、认识论和方法论。从前,因为我们能够收集的数据少,所以只能依赖随机抽样。因为是随机抽样,所以讲究精准。因为观察范围小,只能人工设计实验求因果关系。但是,抽样再精确,推理出的总体统计也有偏差;实验设计再好,也是人工控制了某个变量,总有几分不自然。迈尔-舍恩伯格和库克耶在书中说:“大数据对个人的影响是最惊人的。在一个可能性和相关性占主导地位的世界里,专业性变得不那么重要了。行业专家不会消失,但是他们必须与数据表达的信息进行博弈。如同在电影《点球成金》(Moneyball)^③里,棒球星探们在统计学家面前相形见绌——直觉的判断被迫让位于精准度数据分析。这将迫使人们调整在管理、决策、人力资源和教育方面的传统理念”。下面就根据迈尔-舍恩伯格他们的论断,引申出几个重要问题加以讨论。

(2) 随机抽样要不要坚持? 很长时间内,随机抽样是条捷径,它使得数字时代前的大量数据分析成为可能。但是,就像把一张数码照片或一首数码歌曲截成多个小文件似的,在抽样分析时,许多信息就丢失了——我们无法欣赏一首歌曲的抽样。中国有句流行的话是“解剖麻雀”,意思是研究一只麻雀就能知道所有的麻雀是什么样子。这话可能只说对了 90%,肯定有些信息没有涉及到。就跟麻雀差不多大的椋鸟为例,它们是生性喜欢群飞的鸟,以其群飞队形的变幻莫测而闻名,并成为摄影家的爱好。请看下面几张椋鸟群飞图。据说,它们飞行时靠得

很近,但从来不会相撞,这一技能的功用是迷惑捕食者(如隼鹰)。沙丁鱼的群游和蝙蝠的群飞也有这种特性。这一群飞能力单靠解剖一只棕鸟是发现不了的。这个例子说明了抽样研究的局限性。



迈尔·舍恩伯格还举过一个恰当的例子。具有革新性的 Lytro 照相机把大数据运用到了基本的摄影中,与传统相机只能记录一束光不同,它能记录整个光场里所有的光,达到 1100 万束光。具体生成什么照片则可在拍摄后根据需要而定。这就等于收集了全部数据,“样本=总体”。因此,这些照片就更具“可循环利用性”。普通照相机就像随机抽样,拍摄前必须决定聚焦点。所以,随机抽样虽然优点突出,但是它是没有办法的办法。

从前我们没有能力调查研究 10 亿人,所以只好随机抽样 200 人。现在的大数据常常包括几百亿、

甚至几千亿的被研究者——网民有多少,被试就有多少。在我们有能力全面利用大数据开展外语科研之前,只能坚持随机抽样,但在不久的将来,随着大数据的普及,我们会在某个题目上或某个小领域中,达到或接近“样本=总体”的境界。越到后来,此类现象会越多,直到某一天我们积累了各种各样的大数据,随机抽样不再唱主角,只在个别题目上偶尔/仍然使用。如今,我们掌握的数据库越来越全面,它不再只包括一点点可怜的数据,而是包括与研究问题相关的大量甚至是全部数据。我们不再需要担心某一小部分数据对整套分析产生不利影响。我们要接受这些纷繁、杂乱的数据并从中受益,而不是以高昂的代价去消除那点不确定性。迈尔·舍恩伯格(2013)说:“要想获得大规模数据带来的好处,混乱应该是一种标准途径,而不是应该竭力避免的”。他还说,世界上只存在一个唯一真理的想法已经过时了;再去追求它是对注意力的分散。传统数据库的设计要求在不同的时间提供一致的结果。但是,随着数据量的大幅增加以及系统用户的增加,此类一致性将难以保持。以后的大数据库不再固定在某个地方,而是分散在多个硬盘和多台电脑上。为了确保其运行的稳定性和速度,一个记录可能分开存储在两三个地方。当服务器上每秒钟都接受成千上万条搜索指令时,同步更新记录就不太现实了。因此,承认多样性是解决问题的办法。我们也许要学会容忍模糊性和不确定性——只要能对事物有个清楚的、完整的概念,其中掺杂点模糊也可以接受。

(3) 如何处理因果关系和相关关系? 我们喜欢求真,这是好的。但是,相关关系比较容易发现,因果关系很难判定。前面讲过,这个世界充满相关关系,而且是想象不到的相关关系:我们能预测许多事情,但无法解释其背后的原因。例如,医生依靠大数据取得了病理分析,制定出治疗方案,但却无法说明其病学干预的合理性。我预料,将来在外语学习上也会发现此类现象:比如说,凡是读过英语《哈利·波特》的学生都能通过“雅思”,但解释不出为什么。相关不是因果,但相关本身就很重。迈尔·舍恩伯格说:“大数据标志着人类在寻求量化和认识世界的道路上前进了一大步。过去不可计量、存储、分析和共享的很多东西都被数据化了。拥有大量的数据和更多不那么精确的数据为我们理解世界打开了一扇新的大门。社会因此放弃了寻找因果关系的传统偏好,开始挖掘相关关系的好处”。反过来讲,如果多次、大面积地发现 A 与 B 密切相关,也许就有理由相信它们之间存在一点因果关系或间接的、逻辑不通的、理性不接受的因果关系,暂且称之为“尿布和啤酒式的因果关系”(参见本章第一节)。况且,相关关系分析本身意义重大,它也为探讨因果关系奠定

基础。如果存在因果关系,我们可以进一步找出背后的原因。我们还可以从相互联系中找出某些重要变量,将它们用到验证因果关系的实验中去。

已有许多例证说明相关关系重于因果关系。世界上最火的网上书店亚马逊最初雇佣了一批书评专家写书评(等于推荐),以期增加销售量。结果发现书评的作用微乎其微,于是便开始使用一种协同过滤技术(靠书的内容联系)推荐图书。最后由于通过数据推荐产品所增加的销售量远远超过书评专家的贡献,亚马逊于是解散了书评小组。谷歌也能预测流感:在某个特定的地理位置,越多的人通过谷歌搜索跟流感相关的词条,就表明该地区有越多的人患上了流感。还有沃尔玛通过分析无数个年轻女客户的购物趋势,便可预知哪些女孩成准妈妈了,于是提前给她们寄去孕妇必需品说明书。这些例子都说明,“知道是什么就够了,没必要知道为什么。”这等于说,知道相关关系就够了,没必要知道因果关系。在大数据背景下,相关关系大放异彩,它让我们能够更容易、更快捷、更清楚地分析事物。

(4) 实验设计还要不要坚持? 近期,我们在外语教育中还收集不到很多大数据,所以,人工设计的实验还要持续一段时间。但是,由于有了网上教育,我们的实验在规模上和控制手段上都会发生变化。例如,随机抽样可在网上进行;实验组和控制组的分配可让机器完成;前测、后测、改卷和评分一律网上完成;实验不用总是面对面,可以在线进行;实验教学过程也可以用录像记录下来;问卷的回答和提交都在网上完成等。传媒手段的变化也会引起内容的变化。例如,问卷不再是纯纸质的,而是有录音、录像、虚拟情景的。我预料,当我们能够全面收集外语教育大数据时,也许再设计大型实验的意义就不那么重要了。那时我们能观察和分析的场景、因素、学生、教材、教育技术等等,已是数量巨大、种类繁多、层次分明、角度多维,比不够自然的实验揭示的东西不知要多几百、几千倍。我们现在还想象不出到底那时能开展什么研究以及如何研究。当然,话说回来,作为辅助验证手段,小型实验、勘查、观察、民族志研究等本书前面提到的方法永远不会全部消失;它们是新手研究者的入门教材,是刚接触、了解、体会科研时的蹒跚步履,也是大数据分析后的三角核查工具。

(5) 统计学可能发生的变化:统计学、会计学、经济学、管理学、矩阵计算、矢量计算、概率论、正态分布等知识和原理不会发生重大变化,因为它们揭示出世界的某些自然规律。但是,为应对大数据,计算技术和范式会向前跨出一大步。现在云计算、云存储、众包和外包模式等崭露头角,方兴未艾,以后会开发出更多、更完美的计算方式。目前在用的诸如 SPSS 之类的计算软件还会使用相当长一段时间,

直到需要修改来应对新型数据。换句话说,在今后的 20 年里,我们应该进一步学习统计学原理和方法,尽可能多地掌握计算方法,比如,如何对音频、视频、各种图像等进行计算;同时开始了解大数据的有关情况,为科研做好长足准备。迈尔-舍恩伯格他们在书中(2013)略显狂妄地说:“行业专家和技术专家的光芒都会因为统计学专家和数据专家的出现而变暗,因为后者不受旧观念的影响,能够聆听数据发出的声音”。“大数据的先锋们通常并不是来自他们做出极大贡献的领域。他们是数据分析师、人工智能专家、数学家或统计学家,但他们把自己掌握的技能运用到了各个领域。”此话意味着,我们不但不会放弃统计学,还要加强对它的学习和研究,还要开始学习数据分析,而且是大数据分析。

(6) 如何取得大数据? 有关外语教育和学习的大数据从何而来? 这是我们现在就应该考虑的问题。我们现在就应该意识到收集外语教育大数据的必要性和迫切性。一旦开始网络教学、网上辅导、网上考试、网上报名、网上缴费、网上取得毕业证书等等,就要有意识地对有关信息开展物联网、云储存、云端分析、云端分享等,为开展各式各样的科研提供条件。英语教师必须承担开辟收集、挖掘、发现外语教育大数据的责任。近期我们也许还建立不了自己的引擎,但可以借助百度、谷歌等引擎或者其他出路。重要的是,我们首先应该阅读关于大数据的文献,深刻了解大数据时代的特点及其可能性。我们尤其要发挥最大想象力,记住“数据,从最不可能的地方提取出来”之名言,苦思冥想,挖掘再挖掘。举个例子。日本先进工业技术研究所的越水重臣教授研究了一个人的坐姿。很少人会认为坐姿能提供什么有用信息。但它真能提供不少宝贵信息。当一个人坐着的时候,他的身形、姿势和重量分布都可以量化和数据化。该教授和他的工程师团队通过在汽车座椅下部安装总共 360 个压力传感器以测量人对椅子施加压力的方式。把人体屁股特征转化成数据,并用从 0~256 这个数值范围对其进行量化,就会产生独属于每个乘坐者的精确数据资料。这个系统能根据人体对座位的压力差异识别出乘坐者的身份,准确率高达 98%。这项技术可以作为防盗系统安装在汽车上。有了它,汽车能够识别驾驶员是不是车主;如果不是,系统就会要求司机输入密码;如果输入的密码不正确,汽车就会自动熄火。把一个人的坐姿转化为数据后,就能孕育出切实可行的服务和一个前途光明的产业。我们也可以分析事故发生前的人体姿势变化,找出坐姿与安全的关系。该系统还可以在司机疲劳驾驶时发出警报或自动刹车。另外,如果汽车被盗,还可利用坐姿数据识别盗贼身份。细想起来,这位教授的想象力实在惊人。我们

恰恰需要这种钻研精神。

总有一天,我们会发现,被英语学习者搜索得最多的英语词是哪些,有何特征,用于何种目的,反应出搜索者有何种倾向,英语学习成绩如何等等。拥有大数据之后,我们可以分析什么视频被观看得最多,什么教材的借阅率最高,中国英语学习者写作中最常犯的错误是什么,每个学习阶段的词汇量是多大,等等。从大数据的细节里,我们可以分析出学习者的性格、学习风格、认知倾向、学习策略、动机类型等,也可以分析出教师的教学理念、认知方式、习惯性教学行为、与学生的互动倾向、反馈时的风格、知识面的宽窄、教师的魅力、学生的评价等。将来,我们肯定会去分析为什么某些视频最受欢迎,为什么某些“微课”最讨学生喜欢,为何某人的教案被大量售出,等等。总之,任何教授和学习行为都在网上留下“指纹”,我们的信息会被第二次、第三次加以利用,而且不同的利用必将创造出不同的价值。

附注

① 根据复旦大学计算机研究所所长施伯乐的观点,涂子沛的“《大数据》这本书揭示了数据在未来社会中的革命性作用。大数据将是下一个社会发展阶段的‘石油’和‘金矿’。无论是个人,企业还是国家,谁能更好地抓住数据、理解数据、分析数据,谁就能在下一波的社会竞争中脱颖而出。关于数据的知识,将成为个人知识结构中的必备要素和基础”。该书被列入《亚洲周刊》年度十大好书。其主要观点包括:(1)数据竞争,数据兴则企业兴,数据强则企业强;(2)数据治国:足够多的眼睛,将使所有的错误无所遁形;(3)数据开放:隐瞒政府实情的力量,就是摧毁这个政府的力量。该书

最后一章谈到,大数据时代教育可能发生革命性的变化。

② 该书已有汉译本,译者盛杨燕、周涛。宽带资本董事长田溯宁在该书汉译本序中写道:这“是我看到的最好的大数据著作,不管对于产业实践者,还是对于政府和公众结构,都是非常具有价值的”。他还说:“现代历史上的历次技术革命,中国均是学习者。而在这次云计算与大数据的新变革中,中国与世界的距离最小,在很多领域甚至还有着创新与领先的可能。重要我们以开放的心态、创新的勇气拥抱‘大数据时代’,就一定抓住历史赋予中国创新的机会”(迈尔-舍恩伯格、库克耶 2013:1)。该书翻译得也好,一是懂行的人翻译的;再就是译者时不时在脚注里写出自己的看法或解释。译文排版也费了心思,如把重要观点专门排成黑体,把例证变换字体排在方块内。本文所引中文均来自该汉译本,下引此作仅注页码。

③ 《点球成金》改编自 2003 年出版的《魔球——逆境制胜的智慧》,它被《财富》杂志推荐为 75 本商业必读书之一。这本书记录了美国体育界一个真实的商业奇迹:“烂队”奥克兰运动家队自从采纳了一种基于精算、概率论等的计算程序后,横空出世,拉扯着一帮“衰兵败将”型球员,创下 103 场比赛连胜的纪录。这是一部财经电影:统计学、会计学、经济学、管理学、矩阵计算、向量计算、精算、概率论……搭建起一个传说中的完美数学模型。这是一部励志电影,充满丰沛的美式“挑战”精神:个人挑战体制,衰人挑战大牌,被钉死的钉子挑战命运的板子。

参考文献

- Mayer-Schonberger, V. & K. Cukier. 2013. *Big Data: A Revolution that Will Transform How We Live, Work and Think* [M]. New York: Houghton Mifflin Harcourt.
涂子沛. 2014. 大数据[M]. 桂林:广西师范大学出版社.
维克托·迈尔-舍恩伯格、肯尼思·库克耶. 2013. 大数据时代:生活、工作与思维的大变革(盛杨燕、周涛译)[M]. 杭州:浙江人民出版社.

(责任编辑 林玉珍)

布朗大学招聘启事

Employer: Brown University
Cognitive, Linguistic & Psychological Sciences
<http://brown.edu/Departments/CLPS/>

The Department of Cognitive, Linguistic and Psychological Sciences announces a 2-year Visiting position at the Assistant Professor level in Linguistics/Language Processing at or above the word level, beginning September 1, 2014 or January 1, 2015. This includes syntactic, lexical, semantic, and/or pragmatic (including discourse) processing.

Expertise at the intersection of theoretical and experimental approaches is expected, and the successful candidate will be able to teach, conduct and supervise experimental research engaging theoretical issues in linguistics and cognitive science of language. Applicants should be able to teach both introductory courses and advanced courses in language processing and

psycholinguistics. While not required, ability and willingness to teach an introductory course in syntax is highly desirable.

Applications will be considered on a rolling basis, starting July 8, until the position is filled. Curriculum vitae, reprints and preprints of publications, a maximum two-page statement of research and teaching interests, and three letters of reference should be submitted on-line as PDFs at the application website below.

Brown considers applicants for employment without regard to, and does not discriminate on the basis of, gender, race, protected veteran status, disability, or any other legally protected status.

Application Deadline: Open until filled

Website: <http://apply.interfolio.com/25176>

Contact information: Prof. Pauline Jacobson,
Pauline_Jacobson@brown.edu