

# 大数据背景下的互联网翻译

## ——开源理念与模式创新

王家义 湖南工程学院/澳门大学 李德凤 李丽青 澳门大学

**摘要：**互联网自由与分享的开源理念和大数据的技术支持，使普通网民参与到翻译的社会化大生产之中，引发了翻译模式的变革与创新。众包翻译、云翻译与网络机器翻译等互联网翻译模式引领了这一潮流。众包翻译充分展现了用户的主动参与和群体智慧；云翻译整合了翻译的资源和技术，提高了管理效率，满足了客户对翻译产品“低价、高效和优质”的需求；神经网络机器翻译质量获得大幅提升，用户对翻译技术的依赖程度不断增加，译者需要适应机器翻译的结果，并将其修改为合适的译文。人机协助与人机对话也将成为今后翻译研究的新维度。

**关键词：**开源理念；众包翻译；云翻译；网络机器翻译

**中图分类号：**H059

**文献标识码：**A

**文章编号：**1000-873X (2018) 02-0078-05

数字人文由“人文计算”(humanities computing)发展而来，通过引入计算机、统计学等技术和方法对人文学科的问题进行辅助性研究。数据库、计算语言学、量化分析和超文本卡片等技术被大量的应用。随着互联网技术的疾速发展与普遍运用，开放性、协作性和多元创新性成为了数字人文的重要特征，两种趋势也正日趋显现，即：1)从小型手动输入的数据库向真正的“大数据”项目转化；2)降低使用电脑分析文本的技术门槛。以体量大(Volume)、类型多(Variety)、速度快(Velocity)和价值高(Value)为特征的大数据引发了更深层次的信息化，并渗透到翻译行业。随之兴起的众包翻译、云翻译和网络机器翻译等互联网翻译模式既满足了市场对翻译服务的多样性需求，增加了用户体验，又拓展了翻译研究的空间和维度，让人们重新去认识翻译的功能与意义。另一方面，正如Cronin(2013)所呼吁的，我们要重视数字化时代译者伦理研究和探讨普适计算(ubiquitous computing)的应用对翻译实践与理论研究的影响，以及云计算(cloud computing)技术等其他新型信息技术的应用对翻译的可能影响。在这一背景下，通过研究厘清互联网翻译产生的思想基础，剖析互联网翻译的模式与运行机制有助于揭示互联网翻译的本质，为新时期的语言服务和翻译实践提供借鉴和参考。

### 1. 互联网思维：自由与分享的开源理念

《大教堂和集市》(*The Cathedral and the Bazaar*)是开放源代码运动的主要领导者Raymond的集大成之作。该书被看成是互联网文明的伦理基础。也有学者把它比作“互联网文明的《圣经》”(赵绍辉, 2014)。该书对软件的开发模式进行了形象的比喻和深刻的描述。在Raymond(1999)看来，软件开发有两种截然不同的模式，即：大教堂模式(The Cathedral model)和集市模式(The Bazaar model)。大教堂模式下，源代码在软件的开发过程中由专门的团队掌控，在软件发行后才对外公开。该模式下软件的开发是集中、封闭和自上而下的，它反映的是一种“由权利关系所预先控制的级权制度”(赵绍辉, 2014)。与之相反，集市模式下，源代码在开发过程中便在互联网上公开，供他人免费使用或继续开发。这是一种并行的、点对点和动态的开发模式。前者如微软这种封闭的商业软件，后者如Linux这种开放的自由软件。前者彰显的是富丽堂皇、庄严肃穆和等级森严，后者体现的是朴实无华、平易近人和不分等级。这样，在集市里平等、分享和大众参与成为可能。

《大教堂和集市》的另一贡献在于它对互联网社会价值的探讨。该书首次从社会构造上区分了互联网文化与农业文化和商业文化的不同。农业文化看重天地君亲师、君臣父子尊卑

等级；商业文化谋求价值最大化，财富汇集在金字塔的最顶端；互联网文化使社会构造变得扁平化。在互联网的世界里，宇宙是平的，历史是平的，互联网更是平的。免费、平等、分享、参与和去中心化构成了互联网的基本理念。赵绍辉（2014）更是大胆地提出“网络的时代，应该就是集市长而教堂消的时代，就是草根化的时代，扁平化的时代。”

Raymond 笔下的集市模式体现的正是开源运动所追求的自由和分享理念。Linux 和 Fetchmail 等开源软件的成功案例使“开源”作为一种重要的软件开发途径被广大开发者和使用者所接受。如今，“开源已不再被作为一种比较的参照或是区别于主流路径之外的另外一种存在，而是其自身具有了解释价值”（钱进，2016）。开源运动的影响力已经超越了最初的计算机软件开发行业，渗透至其它行业和领域。就翻译而言，Cronin（2013）敏锐地觉察到了数字化时代新翻译经济所追求的“量大、时短、成本低”的特点。并呼吁重视信息技术与工具革新对翻译前途的影响。相比传统的大教堂模式下译员的相对封闭和孤军作战，互联网翻译集市使“开源翻译项目”能够在成百上千的读者群以及合作译者们自己的热切审读情况下，获得更多的反馈和修正（殷燕，2016：113）。随着人类进入 Web 2.0 时代，以“去中心化，开放和共享”为特征的 Web 2.0 技术使得用户获取信息不再受时间和地域的限制。网民主动参与信息创造的热情得以激发。他们既可以是信息的接收者也可以是信息的发布者。网民之间的信息交换和协同合作变得十分便利。创造、分享和消费成为一种时尚和趋势。“互联网的技术和理念极大地推动了翻译技术和工具的革新，引发了翻译的观念、方式、过程、规模以及研究内涵等方面的划时代变革”（刘满芸，2016：17）。众包翻译、云翻译和网络机器翻译等新型的翻译模式应运而生。

## 2. 模式创新：众包、云与网络机器翻译

### 2.1 众包翻译

Web 2.0 模式下，开放式、自助性和大众化的网络平台使用户得以主动参与到信息的创

造、分享和消费之中。随着人们对跨文化交流需求的快速增长，普通网民对国外影视、旅游、文化、新闻等讯息有着大量的翻译需求，而且这种需求往往具有很高的时效性。于是一些年轻人就充当起“文化搬运工”的角色（曹艺馨，2015：80），以自愿参与为原则，采用“众包分译”的形式，在短时间完成海量的翻译任务，形成了网络技术平台设计架构下的翻译组织模式。一批外语程度较好的网民热衷于在互联网上搜索和浏览感兴趣的国外影视资源。同时，他们也乐意将这些影视资源翻译后推荐给朋友或其他的爱好者。早期由志愿译者组织发起的字幕翻译就是在这样的背景下产生的。“整个活动不以盈利为目的，更多是出于个人兴趣、对以外语为载体的信息的需求，呈现出一种寄托了理想的‘乌托邦’特征”（同上）。网络集市上的平民志愿者承担了原本属于大教堂模式下由职业译者来执行的翻译任务。这种在互联网技术支持下，由众多草根译者进行翻译大生产的新型翻译模式被称为众包翻译模式。Hossain 和 Kauranen（2015）考察了 2006-2014 年间不同学科研究者在 346 个期刊和学术会议上发表的关于“众包”主题的论文。该研究按关键词频率排序归纳了“众包”的特点，其主要概念依次为“社会”、“网站”、“创新”、“开放”、“信息”、“人文”、“在线”、“社区”、“机制”、“集体”和“网络”。具体到翻译领域，众包模式又呈现出规模化协作、开放式的工作流程、译者非职业化、译者即读者（陆艳，2012），用户生成（O'Hagan，2016）等特征。在自由和分享的理念下，众包翻译表现的是用户的主动参与和群体智慧，而自愿、协助、生产和消费是其本质特征。

以 Facebook 为例。这家全球知名的社交网站，每月的访问量达上亿次，用户遍布全球多个国家。为了解决网站本地化问题，Facebook 创造性的推出了由双语用户自愿为网站进行翻译的众包模式。Facebook 把翻译任务碎片化，英文网站内容被拆分成多个文本片段。翻译志愿者发挥其双语能力，花费少量的时间进行翻译就能使广大不懂英语的用户浏览

该网站内容。翻译志愿者可以在平台的“翻译反馈”和“翻译社区”论坛上进行互动交流,获得翻译反馈和评价。无需协议、没有报酬,只是出于对 Facebook 的喜爱和对创造语言的热情,这些分散的个体就能协作完成海量翻译任务。由于用户熟悉网站的界面和网络环境,相比职业译者他们更善于翻译 Facebook 的特定概念 (specific concepts) (Losse, 2008)。也有研究通过比较发现,熟悉网站内容的双语用户比“局外人”的职业翻译者更适合网站本地化的翻译 (Garcia, 2015)。在 Facebook 的 75 种语言中,绝大部分语言的翻译是由志愿者完成的。截至 2010 年, Facebook 的翻译平台有 40 万的翻译志愿者 (Van der Meer, 2010),他们当中每周都会进行翻译的志愿者就已经超过了 10 万人。“Facebook 依靠社交网络积累的人力资源可以翻译超过全球所有专职译员的翻译总量”(陆艳, 2012: 74)。Facebook 通过众包成功解决了社交网站翻译量大和涉及语种多的难题,使该社交网站取得了巨大的成功。

创建于 2006 年的译言网是众包翻译的另一成功典范。正如译言网的宗旨“发现、翻译、阅读中文之外的互联网精华”所描述的那样,该网站专注于把互联网上的外文精品如报刊、文章、书籍等翻译成中文,并在网上发布和共享。译言网的主要栏目包括图书翻译、视频翻译、协作翻译、及时译和译文学习区。译言网将翻译任务发布在不同栏目中,公开招募译者。网民通过简单的流程就可以注册成为译言网用户。用户在译言网上可以通过多种方式进行交流和互动,如向译言网推荐优质外文资源;翻译原文资源库的文章;浏览、点评、纠错和眉批他人的翻译作品;在及时聊天室与他人进行在线交流等。此外,用户还可以参加他人小组和自己创建专题小组,实现协作翻译。译者和读者在互动交流中取得了共同进步。根据译言网的统计,该网站的注册用户已接近 60 万,每周大约有 400-500 人发表各类译文,累计译文数超过了 40 万篇。

## 2.2 云翻译

云计算以大数据运算作为基础,“通过互

联网庞大的计算处理能力,将待处理程序自动分拆成无数个较小的子程序,再交给多部服务器所组成的庞大系统,经过搜寻、计算和分析将处理结果回传给用户”(Miller, 2008: 1)。云计算具有规模大、安全性强、可扩展和按需服务等特点,数据处理的能力获得快速提升,综合管理成本大幅降低。依据云计算的 SPI (SaaS、PaaS 和 IaaS) 三层架构服务理念,一些语言服务提供商通过构建云翻译平台把云计算技术引入了翻译行业。具体来说,就是语言服务提供商“将数据和程序存储在云端(服务器, IaaS 平台),借助一些开放的 API (应用程序界面)将网络变成一个庞大的多功能操作系统,网站之间可进行信息交流 (PaaS 和 SaaS 平台),用户在互联网上以注册用户形式拥有自己的数据,并能在不同网站上使用”(朱琳, 2016: 73)。云翻译平台利用功能强大的网络技术和语言处理技术将分散在各地的翻译资源整合到一起,通过对资源、技术和服务的优化配置和重新组合,极大提高了整个翻译行业的效率。这一过程实际上也是“众包服务商将网络用户创造的价值变为其有偿商业化资源的一部分”(同上),通过云技术构建的翻译平台和产品定制式服务提高了管理效率,满足了客户对翻译产品“低价、高效和优质”的需求。(Garcia, 2015)

2011 年传神公司推出了“云翻译服务平台”,业务模型包括任务管理、协同翻译和术语词典三大块,服务范围涵盖笔译、现场、会展、媒体、培训与认证、网站本地化和软件平台等多个语言服务相关领域。传神的“语联网”通过建立任务管理调度机制,实现任务进度监控和分工协同;通过建立翻译管理平台,实现翻译任务标准化、规范化管理;通过提供辅助翻译工具,实现高频词句译前分析及辅助翻译和质量检查;通过建立术语整理回收与管理平台,实现词典应用。在“语联网”云翻译服务模式,客户海量的翻译需求被拆分成若干个片段,分别由平台中智能化匹配到的不同译员分工完成。由于翻译任务的拆分和译员的筛选都是由语料库智能化匹配和生成的结果,因此相同原文在译文中可以保持统一的译法。

同时，“语联网”上译员正确翻译过的句子经专家审核后存储起来形成智能型“语料库”，当再次碰到相同的表达时，系统会自动匹配并显示为对应的翻译结果，省去了再次翻译的过程，提高了工作效率，又有效避免了同一表达出现不同译文的情况，保证了翻译的准确率。同时，语言专家也通过在线操作，解决语料、术语等问题。据介绍，在这一模式下，传神的翻译成本比一般翻译公司低 30%。

### 2.3 网络机器翻译

利用计算机对自然语言进行自动翻译的想法产生于 20 世纪 40 年代，其发展经历了 50 至 60 年代的初创期，60 年代中期的低潮期，70 年代的恢复期，80 至 90 年代的多样发展期和 2000 年以后的新时期五个阶段（Liu & Zhang, 2015）。机器翻译系统也由早期的基于规则的翻译系统过渡到今天的基于统计的和基于神经网络的翻译系统。2006 年 Google 在线翻译服务的推出标志着“数据驱动的统计机器翻译方法成为商业机器翻译系统的主流”（刘洋, et al., 2017: 1145），网络机器翻译的模式得以确立。该公司又于 2009 年推出 Google Translator 工具包，使免费的机器翻译用于更多的语言组合，并为网络用户提供免费的在线机器翻译和翻译记忆系统。此后，网络机器翻译呈现出蓬勃发展的态势。Hutchins 等（2010）在《翻译软件汇编》（Compendium of translation software）中就对 50 多个网络机器翻译软件的产品特征、语言组合和价格清单进行了详细的介绍，并指出 Yahoo! BabelFish, Click2 Translate, Prompt-Online, Microsoft Bing Translator, Transledium 等多个翻译系统都提供多语翻译，拥有广大的网络用户群体。

近年来，随着深度学习的研究取得较大进展，端到端神经网络机器翻译（end-to-end neural machine translation）技术异军突起，翻译质量获得大幅提升。有研究对比了统计机器翻译与神经网络机器翻译（Neural Machine Translation, NMT）在 30 种语言对上的翻译质量，结果表明 NMT 在其中的 27 种语言对上超

过统计机器翻译（Junczys-Dowmunt, 2016）。百度技术委员会联席主席、自然语言处理部技术负责人吴华博士指出 NMT “克服了传统方法将句子分割为不同片段进行翻译的缺点，而是充分利用上下文信息，对句子进行整体的编码和解码，从而产生更为流畅的译文”（吴华, 2016）。继百度之后又有谷歌、微软等多家公司发布了 NMT 系统并很快取代统计机器翻译成为这些主流系统的核心技术。如今，多数网络机器翻译系统拥有网页版和手机 APP 等多种产品形态。除了传统的文本翻译外，谷歌、百度和有道的翻译工具还推出语音、图像、摄像头取词、照片扫描等不同模态的多语种免费在线翻译功能，极大地满足了不同用户对翻译的多样性需求。依靠人工智能技术，彩云小译在语音识别技术上取得突破进展，实现了中英文间的同声传译。可以说，早些年在科幻小说中描述的情节正在变成现实。“如今，不借助技术手段进行翻译是难以想象的”（Kenny, 2016: 10）。传统的依赖人工翻译的方式正在发生变化，“机器翻译+译后编辑”的模式“已经成为当前和未来职业译者的主流工作模式”（王树华, 2017: 86）。随着机器翻译质量的不断提升，可以预测今后译者的工作更多的是适应机器翻译的结果，并将其修改为合适的译文。人机协作与人机对话也将成为今后翻译研究的新的维度。

### 3. 结语

大数据背景下翻译行业出现了新的变化。自由和分享的开源理念、Web 2.0 技术加上云计算和人工智能等新兴元素的应用助推了翻译模式的变革与创新。众包翻译、云翻译和网络机器翻译等新型翻译模式的出现吸引了大众的主动参与，使协同翻译和集体智慧得到充分发挥，也把翻译引入了规模化和非职业化。译员结构、翻译进程及其管理机制均随之而发生改变，翻译的社会化属性日趋显著。我们有理由相信未来的翻译实践将走向平台化、线上化和智能化的发展道路。

另一方面，我们在张开臂膀迎接新型翻译模式的同时，也要用批判的眼光审视翻译行

业的这场惊天巨变 (Cronin, 2013; Kenny, 2016; 叶子南, 2017)。大数据背景下互联网翻译依然存在不足。众包翻译与官方权力及主流话语体系的对抗与对话 (曹艺馨, 2015); 众包翻译尚存在伦理悖论与乱象 (郝俊杰, 2016); 网络机器翻译的质量与评估问题 (Doherty, 2016; 叶子南, 2017), 大数据背景下译者如何自处 (Cronin, 2013; Kenny, 2016) 等问题值得人们深入思考和进一步探讨。

**基金项目:** 本研究获湖南省哲学社会科学基金项目 (15WHL11), 湖南省社会科学成果评审委员会基金项目 (XSP17YBZZ022), 湖南省教育科学“十二五”规划项目 (XJK014YYB003), 湖南工程学院青年科研重点项目 (XJ1506), 澳门大学跨年研究项目 (MYRG2015-00150-FAH, MYRG2016-00096-FAH, MRG2017-00139-FAH) 资助。

## 参 考 文 献

- [1] 曹艺馨. 互联网大众翻译模式微探: 历史, 现时, 未来 [J]. 中国翻译, 2015 (5): 78-82.
  - [2] 陈伟, 莫爱屏. 《数字化时代的翻译》述评 [J]. 外语教学与研究, 2014 (2): 309-313.
  - [3] 刘满芸. 翻译技术时代翻译模式的裂变与重构 [J]. 中国科技翻译, 2016 (4): 17-20.
  - [4] 刘洋. 神经机器翻译前沿进展 [J]. 计算机研究与发展, 2017 (6): 1144-1149.
  - [5] 钱进. 作为开源的数据新闻 [J]. 新闻大学, 2016 (2): 6-12.
  - [6] 陆艳. 众包翻译模式研究 [J]. 上海翻译, 2012 (3): 74-78.
  - [7] 陆艳. 云计算下的翻译模式研究 [J]. 上海翻译, 2013 (3): 57-61.
  - [8] 斯介生, 宋大我, 李扬. 大数据背景下的谷歌翻译——现状与挑战 [J]. 统计研究, 2016 (5): 109-112.
  - [9] 殷燕, 刘军平. 互联网翻译行业伦理基础解读——雷蒙德开源理念观照下的互联网翻译伦理研究 [J]. 西安外国语大学学报, 2016 (3): 111-114.
  - [10] 王华树. 语言服务的协同创新与规范发展——2016 中国语言服务业大会暨中国译协年会综述 [J]. 中国翻译, 2017 (1): 85-88.
  - [11] 吴华. 提升翻译质量: 百度机器翻译技术领跑 NMT 时代 [EB/OL]. [2017-7-16] [http://www.sohu.com/a/122342894\\_155403](http://www.sohu.com/a/122342894_155403).
  - [12] 朱琳. 翻译众包的崛起及其对翻译行业发展的影响 [J]. 上海翻译, 2016 (2): 71-77.
  - [13] 赵绍辉. 再谈互联网思维之二——大教堂和集市 (The Cathedral & the Bazaar) [EB/OL]. [2017-7-14] [http://blog.sina.com.cn/s/blog\\_ad8cab090102v0bz.html](http://blog.sina.com.cn/s/blog_ad8cab090102v0bz.html).
  - [14] Cronin, M. *Translation in the Digital Age* [M]. Routledge, 2013.
  - [15] Doherty S. Issues in human and automatic translation quality assessment [J]. *Human Issues in Translation Technology: The IATIS Yearbook*, 2016: 131-148.
  - [16] Garcia I. Cloud marketplaces: Procurement of translators in the age of social media [J]. *The Journal of Specialised Translation*, 2015, 23: 18-38.
  - [17] Hossain M. & Kauranen I. Crowdsourcing: A Systematic Literature Review [J]. *Strategic Outsourcing: An International Journal*, 2015, 8(1):2-22.
  - [18] Hutchins J., Hartmann W. & Ito E. Compendium of translation software [EB/OL]. [2017-7-16] <http://www.hutchinsweb.me.uk/Compendium-16.pdf>.
  - [19] Junczys-Dowmunt M., Dwojak T. & Hoang H. Is neural machine translation ready for deployment? A case study on 30 translation directions [J]. *arXiv preprint arXiv:1610.01108*, 2016.
  - [20] Kenny D. The Translator and the machine [J]. *The Linguist*, 2016, 55(6):10-13.
  - [21] Miller M. *Cloud computing: Web-based applications that change the way you work and collaborate online* [M]. Que publishing, 2008.
  - [22] O'Hagan M. 2 Deconstructing translation crowdsourcing with the case of a Facebook initiative [J]. *Human Issues in Translation Technology: The IATIS Yearbook*, 2016: 25-44.
  - [23] Raymond E. *The Cathedral and the Bazaar* [M]. Sebastopol: O'Reilly Media, 1999.
  - [24] Losse K. *Facebook-achieving quality in a crowd-sourced translation environment* [J]. XIII Localisation4All, Limerick, Ireland, 2008.
  - [25] Van der Meer J. Where are Facebook, Google, IBM and Microsoft taking us [EB/OL]. [2017-7-18] <https://www.taus.net/think-tank/articles/translate-articles/where-are-facebook-google-ibm-and-microsoft-taking-us>.
  - [26] Liu Qun and Zhang Xiaojun. Machine translation: general [A]. In Sin-Wai C. *The Routledge Encyclopedia of Translation Technology* [C]. London: Routledge, 2015:105-119.
- [作者简介] 王家义, 湖南工程学院外国语学院副教授, 澳门大学文学院博士生。研究方向: 语料库翻译学、基于认知视角的翻译过程研究。  
李德凤, 澳门大学文学院教授, 博士, 博士生导师。研究方向: 翻译教学、翻译研究方法论以及基于认知视角的翻译过程研究。  
李丽青, 澳门大学文学院副教授, 博士, 博士生导师。研究方向: 口译研究、基于认知视角的翻译过程研究。  
[作者电子箱] boriswjy@126.com; defengli@umac.mo; viclcl@umac.mo