

本地化 的未来



AI Business
电子书系列



如何利用语言和文化来
推进数字化转型

合作伙伴：

LIONBRIDGE

业务扩张的核心

本地化需要根据特定市场调整产品、通信、服务等内容,使其符合当地客户的期望。

在全球化的世界中,本地化让企业能够进军国际市场,从而增加收益,同时在此过程中发挥规模经济的优势。因此,对于任何商业战略来说,如果要将产品推广至本地以外的市场,本地化则是推行该战略的核心所在。即使一家公司在品牌宣传中着意与特定地区挂钩(例如 Apple 承诺其产品是由“Apple 在加利福尼亚州设计”),也必须开展本地化工作,以便吸引海外客户。

当然,翻译仍然是本地化的核心部分。正如我们将在本电子书中介绍的,其中部分工作语言将由专家人工完成,部分工作可交由先进的 MT 工具完成,还有部分工作则将通过两者相结合的方式完成,这种混合方法叫做增强翻译(参见第 5 页)。

但真正的本地化绝不仅限于语言。要想将一项服务或一种产品从一个市场成功移植到另一个市场,本地化人员需要了解各市场间的细微差别。“总的来说,我们使用另一种语言再创造一种体验。” NI (National Instruments) 语言服务高级经理 Inés Rubio 告诉 AI Business,“本地化的重点不在于语言,而在于文化。”

在这本由翻译与本地化专家 Lionbridge 提供支持的电子书中,我们将了解国际企业面临的一些挑战以及做好本地化所带来的好处。我们采访了多位专家,并引述了 Zynga、Royal Caribbean、Volvo 和 Wikimedia Foundation 等公司的观点。我们衷心希望这份出版物能为您提供所需知识,助您打造出真正的全球化产品和服务。

Luke Dormehl | 副主编 | AI Business

电子书合作伙伴:

LIONBRIDGE

本地化的未来 | www.lionbridge.com



3

现代本地化服务简介
语言和文化

4

市场概述
本地化成为主流

5

AI 的应用
什么是增强翻译?

7

案例研究
Lionbridge 如何帮助 Royal Caribbean 彻底改革本地化工作

8

服务全球受众
Zynga 如何借助本地化俘获数百万玩家的心

10

着眼于当地市场
使用语言服务提升 Wikipedia

12

MT 的进步
拐点

14

案例研究
Lionbridge 如何帮助 Volvo Cars 打造统一的线上市场

15

接下来将发生什么?
语言技术的未来

17

学术观点
对等问题

语言和文化

本地化的重点不在于语言，而在于文化。例如，提到西班牙语时，我们的目标受众是哪些人？是全世界说西班牙语的人？还是欧洲的西班牙人？如何让目标受众产生共鸣？由此可见，本地化绝不仅限于翻译词语，而是蕴含着 [更深厚的文化底蕴]。

翻译固然重要，但更关键的是了解目标受众

对网站、使用说明书和软件应用程序等资产进行语言翻译，以便受众能够阅读或理解，这是本地化的常见定义。

Lionbridge 产品主管兼战略师 Kajetan Malinowski 告诉 AI Business，“本地化的目的是让产品或服务真正契合当地市场，使用最终消费者的语言很重要。”但语言只是一部分。公司要想成功打入某些利润丰厚的市场，还需要借助文化方面的专业知识。

不止步于翻译

您是否听过这句话：“怎么说要比说什么更重要”？不能将本地化简单等同于翻译是有原因的。本地化要考虑到当地沟通的方方面面，其中不只包括语言，还包括地区差异。例如，测量单位就是因地而异的一项标准。法规或技术方面的要求甚至“语气”或语体等其他要素也因地点而异。

“总的来说，我们使用另一种语言再创造一种体验。”NI 语言服务高级经理 Inés Rubio 告诉 AI Business，“本地化的重点不在于语言，而在于文

化。例如，提到西班牙语时，我们的目标受众是哪些人？是全世界说西班牙语的人？还是欧洲的西班牙人？如何让目标受众产生共鸣？由此可见，本地化绝不仅限于翻译词语，而是蕴含着 [更深厚的文化底蕴]。”

本地化还可能涵盖图片的翻译（例如界面图标），使其符合目标受众的偏好。这些图标不一定具有普遍或明确的意义。例如，在美国和英国等市场，用灯泡表示一个想法是常见的做法，但在其他地方，意义却有所不同。“竖起大拇指”表情符号在美国和英国有积极的含义，但在拉丁美洲、中东和西非则有着冒犯的意味。颜色在不同地区也有不同含义。在印度，橙色象征着爱和勇气，但在中东文化中，这种颜色常与仙逝和哀悼联系在一起。

做好本地化可以带来巨大利益，做得不好或没有认识到当地习俗的重要性，则可能会造成严重损失。在 21 世纪 20 年代，本地化是所有现代企业都应采用的重要利器。



本地化成为主流

“刚开始时，公司可能完全想不到要进行本地化。”她说，“可能直到最后一刻才会想起它，而且也不认为本地化是一道上游环节。但随着公司从错误中吸取教训，意识到后期再做这些改变的成本有多高昂，公司和整个团队就会 [开始重视本地化]。

着眼于本地化市场的规模及其增长背后的驱动力

现代本地化的诞生可追溯到 20 世纪 80 年代，本地化 (localization) 通常简称为“l10n”，其中数字 10 指“l”与“n”之间的字母数。而计算机软件也是在这个年代开始成为全世界办公室和家庭的必备之物，这并非巧合而已。为了抓住这一不断增长的市场机遇，很多公司开始拥抱本地化，包括支持新的字符集，为特定市场添加或改动功能。这些工作最初在公司内部完成，但此类工作的规模和重要性快速促使本地化行业走向成熟。

规模超乎以往

在过去十年里，语言服务市场的规模几乎翻了一番，涵盖说明制作、翻译和更广泛的本地化工作。服务于语言服务行业的主要出版公司 Sator 最近发布了一份报告，声称 2020 年潜在的笔译、本地化和口译市场规模为 238 亿美元。科技、游戏、媒体和生命科学是增长最快的客户垂直市场，2021 年整个行业预计最多增长 10%。

根据 Statista 披露的信息，尽管欧洲人口低于全球人口的 10%，但却占据了全球语言服务市场的半壁江山。这是因为该地区使用多种语言。同时，从市场规模来看，美国仍然首屈一指。

这种差异并不仅体现在国家/地区之间，也体现在行业之间。本地化在某些行业已经是一个成熟领域，但在其他（尤其是新兴）行业和公司，其重要性仍有待巩固。

NI 语言服务高级经理 Inés Rubio 表示，可以通过对本地化的投入程度判断一家公司的成熟度。“刚开始时，公司可能完全想不到要进行本地化。”她说，“可能直到最后一刻才会想起它，而且也不认为本地化是一道上游环节。但随着公司从错误中吸取教训，意识到后期再做这些改变的成本有多高昂，公司和整个团队就会 [开始重视本地化]。随后，他们开始将本地化纳入整个过程。”

本地化技术的发展使得该行业所依赖的工具不断得到改进。神经网络机器翻译 (NMT) 等全新创新技术仅出现了几年时间，却已产生重大影响。这些突破将继续改变客户和最终用户的本地化世界。





什么是增强翻译？

与 *Lionbridge* 的 *Will Rowlands-Rees* 和 *Jaime Punishill* 一同探讨人与机器的关系

“MT 是 AI (人工智能) 的一种基础性应用。” *Lionbridge* 首席产品官 *Will Rowlands-Rees* 告诉 *AI Business*, “MT 首次出现在 20 世纪 50 年代早期。甚至在那时, 有人就预测不再需要人工翻译了, 但是其预测可能有点偏差。”

近 75 年后, 人工翻译仍然在翻译工作中发挥着重要作用。即使业内已引入 MT, 人工的作用仍不可小觑。免费公开的翻译引擎在某些情况下可能很奏效, 但基本不适用于商业用途。

“令人发笑的 MT 译文比比皆是。” *Lionbridge* 首席营销官 *Jaime Punishill* 告诉 *AI Business*, “这些公共引擎未经过训练。在商业环境中效果还好一点, 但仍然会造成奇怪的语言问题。从英文翻译成韩语可不简单, 更不要说从德语翻译成韩语了。MT 译文往往令人捧腹大笑。”

出色翻译的重要性

如果您只想了解一篇新闻文章的大意, 也许能接受 MT 版本。但坦白来说, 在某些其他情况下, 达到“合格线”的翻译还远远不够。有些内容的翻译需要近乎完美, 例如投资产品的招股说明书、医学研究标签或法律文本。

“如果我在为一家大型跑车制造商制作手册, 却给出了错误的说明, 那就会涉及健康和安全问题。” *Rowlands-Rees* 说, “又或者是临床研究或任何其他被视作可能 [关乎生死] 的领域。从公司的角度来看, 每件事都将影响品牌。我们要确保翻译正确, 因为如果翻译不正确, 就会对品牌造成不良影响。”

这一过程一般称为后期编辑, 即先使用机器翻译 (MT) 工具进行初步翻译, 然后将这些未经编辑的 MT 语段交给人类语言专家, 他们将纠正计算机造成的任何错误。从理论上来说, 虽然这种方法可以节省时间, 但也存在局限性, 例如仅在流程末期引入人工翻译。

走进增强翻译

增强翻译这一术语由 *CSA Research* 首创, 它利用一系列技术让人类语言专家更快地产出优质翻译, 同时提升一致性和准确性。增强翻译可以更好地结合人类能力与技术, 从而改善翻译体验。其实, 其中用到的诸多方法早已面世, 只要将这些方法结合起来, 就可以创造出更完备的体系。增强翻译可自动执行低价值且耗时的任务, 还能使人类语言专家充分调用自己的专业知识。▶

► 增强翻译包含七个主要方面：
翻译记忆库
自适应神经网络机器翻译
质量评估
自动内容充实 (ACE)
术语管理
无人化项目管理
翻译管理系统 (TMS)

所有这些都是并行工作的。例如，翻译记忆库可帮助完成重复性翻译任务。“我们每年可能都会翻译同一本产品手册。” Punishill 说，“该产品手册的大部分内容与前一年完全相同，但要放在以前，公司要花钱翻译两次或更多次。翻译记忆库可用于查看源文本并判断是否有匹配语句。”

如果有完全匹配的语句，文本可直接应用而无需作出任何其他修改。另一方面，如果有些部分不匹配，可在标记后由翻译人员更新，使其能够检查前后的词语和句子，以确保增加的内容不影响文本上下文。

增强翻译还有另外一项好处，即为特定翻译工作确定理想的人员和机器资源：也就是将工作交给翻译人员，或是了解哪种 MT 引擎将产出理想成果。

发挥优势

“在这一方面，您会发现每个机器擅长不同的工作。”Punishill 继续说道，“一个擅长将英语翻译成德语，另一个擅长将西班牙语翻译成韩语；一个擅长翻译药学内容，另一个擅长翻译营销内容。因此您需要混搭 MT 引擎，将它们的功能组合起来支持您的翻译系统。”

“Lionbridge 可以帮助公司打造并为其运行 [这些系统]。我们可以覆盖您需要开展的各种工作，从训练引擎到维护品牌和语言指南、术语词典及保持系统顺利运转的所有语言资产。我们不仅要翻译好，还要能够以客户的风格来翻译，例如 Apple 的风格或 Microsoft 的风格。”

Rowlands-Rees 提到：“我们能够告诉用户，‘我们认为这份文档最好由人工翻译，原因是这样的，’或者‘我们认为选择 MT 翻译效果不错，想使用 MT 引擎，原因是这样的。’我们已经了解了一些原因，能解释为何在不同情况下选择人工翻译或 MT 更合适。我们

对发来的文档进行程序分析从而了解这一点，然后我们会安排理想的内容翻译方式，这样做能够提高我们的效率，更重要的是，也能够提高客户的效率。”

增强翻译也可以通过其他方式支持翻译工作，例如能够使用预测分析判断哪些文本能翻译好，从而提升上游和下游的内容创作水平。此外，它还可以对其他语言的搜索引擎优化 (SEO) 等方面进行综合评估，以显著提升网络曝光度。

增强翻译代表 MT 的最前沿技术，运用该技术可以大幅提高输出内容的质量。通过将翻译和质量预估(对 MT 可能达到的质量水平进行统计性估计)等技术与最新的自适应神经网络机器翻译工具结合起来，增强翻译将成为任何本地化团队装备库中的一件利器。未来几年，增强翻译这一术语预计会随处可见。



Lionbridge 如何帮助 Royal Caribbean 彻底改革本地化工作

大型游轮运营商希望赢得更多全球受众的青睐

远洋游轮环球航行，顾名思义，这是一项跨国业务。每年，全球约有 530 万游客准备乘坐 Royal Caribbean 的轮船远航到各国外目的地。为确保乘客享受到预期体验和维持 Royal Caribbean 品牌形象的完整性，该公司网站 royalcaribbean.com 需要呈现高质量的翻译内容。这样一来，用户可以方便地订购假期行程，还能了解到 Royal Caribbean 的一系列船舶、活动和目的地。

要解决的问题

对于 Royal Caribbean 这样久负盛名的品牌，缺乏规范性且繁杂冗余的内容翻译工作流程会对内容的准确性和传递效果产生不利影响。公司决定升级内容平台，并找到了经验颇丰的本地化专家施以援手。Royal Caribbean 选择使用 Lionbridge 的服务，这家公司提供语言服务已长达 25 年之久。本次升级的目标是利用集中式内容网站提高流量和销售额，改进翻译 workflow 并提供卓越的用户体验。

采用的解决方案

实施的解决方案包含两个工具：Lionbridge Connector™ 和 Adobe Experience Manager™。该解决方案提供高效的工作流和所需的多语言专业知识，不仅能够确保正确更新 Royal Caribbean 的内容，还进行了本地搜索优化。必要语言的材料在可接受的时间范围内发布，且成本合理，表达十分流畅。仅四个月，超过 80% 的现有 Royal Caribbean 内容（共计约 1,500 页）翻译成了七种语言。

“使用 Lionbridge Connector for Adobe Experience Manager 可实现网站翻译流程自动化，因此我们可以轻松翻译数千个网页。”Royal Caribbean Cruise Lines 产品内容优化经理 Ariadna Castro 说，“Lionbridge 提供的技术可帮助我们实现高效的翻译工作流程。Lionbridge 的客户团队非常出色，可随时为我们提供所需支持，助我们实现目标。”

本地化项目应首先为客户提供优良的用户体验。但除此之外，这些项目还必须能够实现切实成果，证明公司为何能从此投资中获益。以 Royal Caribbean 为例，效果几乎是立竿见影。在新技术采用后的头几年里，该公司报告有机点击量提高了 35%，有机搜索曝光量的涨幅更大，高达 37%，全球点击量增幅达到惊人的 3 倍。

1,500

页的内容在 4 个月内由
Lionbridge 翻译成了
7 种语言

随着全球客户获得更具个性化的体验，每一名参与者都从出色的本地化彻底改革中获得了回报。



Zynga 如何使用本地化俘获数百万玩家的心

“我们尽量不使用本地化一词，而是使用‘文化化’。”Zynga 发行总裁 Bernard Kim 说道，“所以我们不仅改变了产品的语言表达，还根据当地市场，对玩法和互动特性进行了灵活调整

发行总裁 Bernard Kim 和战略伙伴关系主管 Vineet Rajosi Sharma 都认为要开发能吸引全球玩家的游戏

游戏的普适性非常高且通常能够吸引全球受众。要在当前的游戏市场中蓬勃发展，大型游戏开发商必须能够适应各地。Zynga 成立于 2007 年，是全球社交游戏的领军开发商，全球日活跃用户高达数百万人。该公司的一款火爆游戏《FarmVille》于 2009 年 6 月在 Facebook 上发布，几周内便累积了 1000 万日活跃用户。十多年来，Zynga 大举转战手机游戏领域，凭借《Zynga Poker》、《Words With Friends 2》、《Game of Thrones Slot Casino》和《Harry Potter Puzzles and Spells》等游戏继续吸引了 150 多个国家/地区的数千万月活跃玩家。

该公司的“经典游戏”如此成功的一个秘诀是，着重强调产品适应全球各市场。毕竟，将英语作为第一或第二语言的人不到全球人口的十二分之一。如果游戏只有英文版，就会与大部分潜在用户群失之交臂。因此，以多种语言呈现

游戏具有重要的商业意义。但是，正如 Zynga 高管向 AI Business 所说的，要想成功进行本地化，仅翻译游戏内文本还远远不够。

我们称之为“文化化”

“我们尽量不使用本地化一词，而是使用‘文化化’。”Zynga 发行总裁 Bernard Kim 说道，“所以我们不仅改变了产品的语言表达，还根据当地市场，对玩法和互动特性进行了灵活调整。例如，韩国玩家可能在一天内的互动次数较多而且较痴迷于我们的一款游戏，那我们就为这些玩家打造符合当地市场文化的游戏选项和挑战。”

Zynga 战略伙伴关系主管 Vineet Rajosi Sharma 告诉 AI Business，让玩家感觉到游戏是专为他们而设计，这点至关重要。他说，“我作为用户，与一个应用交互时，需要感觉到它是为我而打造的。”>



▶要让一个游戏同时吸引多个国家/地区的数百万用户，同时还为用户提供个性化体验，这并非易事。游戏行业资深人士 Kim 说，他经常面临 IP 在全世界翻译不好的挑战。他得到的一个教训是，特定市场的游戏本地化不能是一项“附属”工作。

“[玩家知道] 他们是否会长期玩一款游戏。”Kim 说，“游戏的感觉、语言、表达方式，如果这些在投入特定市场前的最后一刻再下功夫，那就太晚了，公司会从互动数据中看到 [不良影响]。”

为了克服这一挑战，Zynga 在任何新项目伊始就考虑本地化(或文化化)战略，意在扩大潜在玩家数量。

“看一看我们的经典特许经营游戏，它们能够吸引全球玩家。”Sharma 说。在谈到哪些概念能获得通过时，他解释道，“我们将文化化看作重要的优先事项。我们需要让游戏在当地市场中的所有方面内容都易于理解。”

语言：重要，但不是唯一因素

Zynga 的工作室遍及全球，在美国、加拿大、印度、爱尔兰和土耳其都开设有办公室，所以语言是一项主要考虑因素，这点可以理解。Sharma 分享了一个可以简化本地化流程的小窍门：减少文字数量。“如果在游戏中加入太多文本，基本上就加大了工作的难度。”他说，“也没有人愿意阅读太多文字。”对于屏幕较小的移动设备而言尤其如此。

更有挑战性的是，任何文字在翻译后都必须能表达出原始游戏的“风格”。除了游戏内的文字和符号，这还包括货币、度量衡和日期格式等虚拟世界元素。此外还有应用商店描述、营销材料和其他必须翻译的相关文字。

“不管是内部翻译还是外包，都要有个译员。”Sharma 解释道，“我们非常重视 [翻译质量]。我们有本地化 QA，或者我们所说的‘国际 QA’，确保译文连贯合理。”

文化化工作范围之广，书面和口语表达只是冰山一角。例如，有些符号、性别表达方式和称谓必须纳入考虑。与当地流行的社交媒体平台集成以进行营销、认证或社交分享时也是如此。Kim 表示，因市场而异的一个显著细节是，游戏中的“应用内购买包的形状”通常需要改变，以增加对特定地点玩家的吸引力。过去，该公司曾收缩游戏市场(例如其在某些亚洲市场推出的 Harry Potter 游戏)以确保最佳体验。

Kim 说：“正因如此，我们选择了由翻译人员进行人工翻译，他们专门研究我们的游戏，长期以来为我们提供帮助并合作开发这些游戏。”

在免费游戏的世界里本地化的重要性

Zynga 的高管提到，在免费游戏领域中，本地化甚至更为重要。不同于收取固定购买费的高端游戏，“免费”游戏通过吸引玩家反复玩来获得收益。这意味着，浅尝辄止的本地化远远不够；例如，当地营销活动组织得很好，但游戏内的改动极少。Kim 说，免费游戏要确保质量，因为在游戏商品繁杂的市场中，玩家只会反复玩有吸引力的游戏。而这通常指使用本地化方法开发的游戏。根据 Statista 发布的信息，全球智能手机用户约为 38 亿，约占全世界人口的 48.33%，因此本地化是手机游戏工作室的一项明智的投资。



“总体而言，我们将文化化视为一项创造性事业。”Kim 说，“从用户通过创意广告或当地营销活动首次了解到我们的游戏，到他们成为玩家之后的 CRM(客户关系管理) 拓展服务，都属于文化化的范畴。是采用电子邮件还是通知？什么时间发出？毕竟公司不希望用户在睡眠时发送推送通知，或发出语言不当的通知。所有这些因素共同构成文化化。”

使用语言服务改进 Wikipedia

“机器学习引进后...我们可以使用多种模型改善读者和编辑的体验。”Albon 说,“例如,我们有多种模型用于研究并试图预测编辑的质量,以及是否某个编辑是否是破坏性的。”

Wikimedia Foundation 机器学习总监 Chris Albon 探讨项目的使命和当地数据的重要性。

Wikipedia 是全球由用户生成的最大的百科全书,应尽量提高其普及程度。本地化服务是否有助于为世界各地的用户提供更优质的宝贵资源?这是 Wikipedia 创建者面临的问题。Wikipedia 拥有 300 种不同语言的 4000 多万词条,每月服务的用户高达 5 亿。这一问题的答案是“当然可以”。

确保每个 Wikipedia 版本尽可能全面的有效方法是,将它们聚合起来,然后实施自动翻译,例如将英语词条(英语是 Wiki 中占比最大的语言)翻译成用户数量较少的其他语言。

Wikimedia Foundation 使用 Foundation 支持的内容翻译工具整合了一小部分内容,该工具首先对一个 Wikipedia 词条进行 MT,然后由编辑人员审核,根据需要改进,最后再发布。这一点至关重要,因为无人编辑的 MT 可能会造成问题。有这样一个明显的错误,英语中的词语“village pump”(互助客栈)翻译成了葡萄牙语的“炸掉村庄”。但让人参与到翻译过程,就可以尽量减少此类错误。迄今为止,自动化工具已经协助翻译了超过五十万词条。

然而,正如 Wikipedia 机器学习总监 Chris Albon 对 AI Business 所说的,该组织没有完全采用自动化翻译,而是采用了“充分考虑当地目标客户群体观点”的方法。

本地化的机器学习

Wikipedia 采用机器学习模型的范围之广远超过许多用户的想象。“机器学习引进后...使用多种模型改善读者和编辑的体验。”Albon 说,“例如,我们有多种模型用于研究并试图预测编辑的质量,以及是否将编辑视为破坏性举动。”

Wikipedia 使用数百个复杂度各异的“机器人”,从提示可进行操作的词条主题,到解决破坏行为以保持页面内容的可读性,各种工作都由它们完成。但是,Albon 及其同事(他们不会亲自为 Wikipedia 编写内容)只希望尽可能提高免费的百科全书的质量,而不想冒险将自己的偏见引入他们构建的 AI 模型中。

对于 Wikipedia,偏见是指什么?简而言之,对于 Wikipedia 的必备要素和优质 Wikipedia 词条的构成元素,旧金山用户与肯尼亚用户的看法不同,因为他们的文化预期不同。

“到目前为止,我读过的大多数 Wikipedia 词条当然为英语版本。”Albon 解释说,“所以,对于什么是优质词条, >





► 或者词条的特定主题、词条结构、所用语气应该如何，所有这些都基于我作为 Wikipedia 项目的一种目标用户的体验。我的看法并不完全适用于其他目标用户群体。参与到 Wikipedia 项目是共识，但对于什么是高质量的编辑，或什么是破坏性的编辑，或者如何处理 [特定主题]，这些都与本地化息息相关。这些应该由使用该当地语言的当地用户群体来管理。”

本地化是 Wikipedia 的一项使命

Wikipedia 当前运行着约 100 个完全使用当地社区进行数据训练的机器学习模型。这些模型部署在大约 30 个不同的 Wiki 上。Albon 在提到为全世界的所有人提供母语版的免费百科全书目标时，说道：“本地化之所以对于我们的 Foundation 如此关键，是因为它融入在整个项目中。”

要使用当地数据训练每个模型，因为不同国家/地区的用户对于优质 Wikipedia 词条的构成要素可能有着完全不同的预期。

“这一至关重要，鉴于各地文化背景不同，斯瓦希里用户群认为可能重要或不重要的词条元素，或者与之关联的重要事项，或者未研究的主题，这些与英语版本 Wikipedia 有所不同。”Albon 说，“例如，在英语版本 Wikipedia 中，目标用户并不觉得出租车（一种私有的小巴士）之类的东西 [举足轻重]。但在斯瓦希里版本的 Wikipedia 中，用户很感兴趣。因为在非洲东部、南部和西部，这是当地人日常生活的常见交通工具...我们不想强加 [我们的观点]，把我们使用英语 Wikipedia 创建的模型施加给斯瓦希里 Wikipedia。”



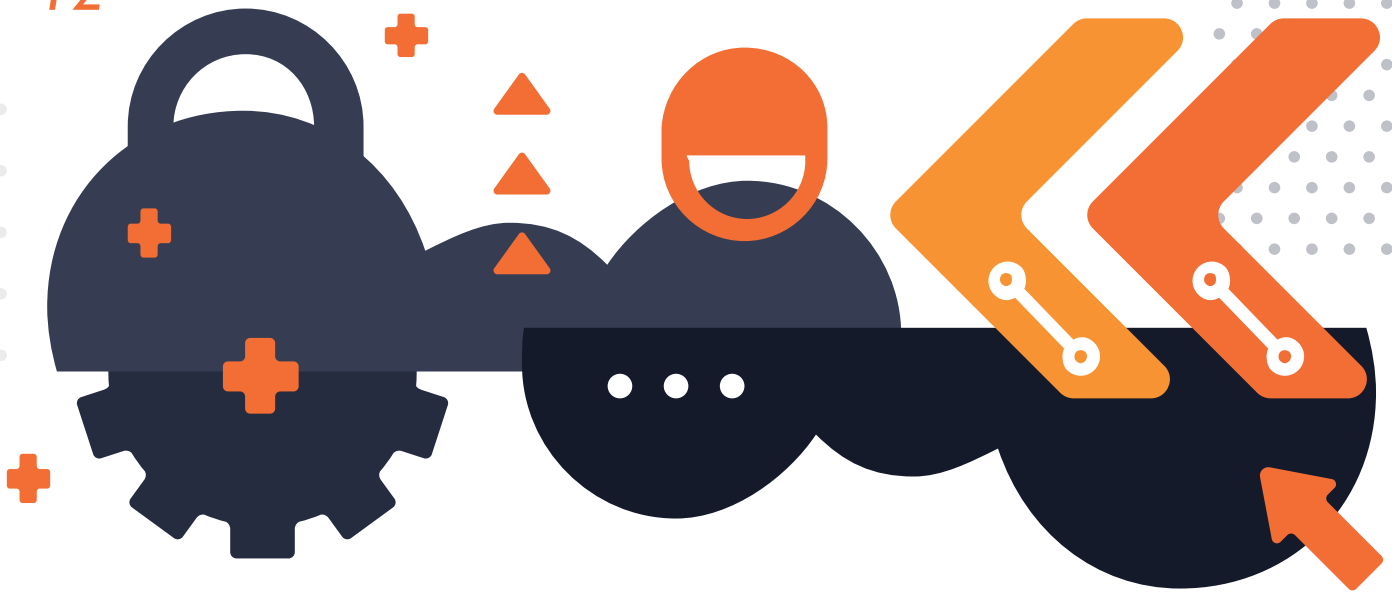
Wikipedia 使用的一个机器学习模型旨在建议何时将一个词条中的词语链接到另一个页面。因为 Wikipedia 上的大多数名词可能都有自己的页面，因此创建出来的词条可能包含诸多超链接，这会让页面显得杂乱，把潜在读者的注意力吸引到意义不大的链接上。因此，Wikipedia 的立场是，词条不应该链接到所谓的“常识”上，即访客几乎肯定都知道的非常基本的信息。但有一个问题，常识可能因地区而异。因此，为此任务部署的 AI 模型必须采用本地化方法来认定常识。

理想标准

“我们制定模型的理想标准是，使用模型所服务的用户群体的数据来训练机器学习模型。”Albon 说，“例如，斯瓦希里 Wikipedia 部门说，我想要一个模型，用来检测编辑质量，那么我们满足这一需求的理想标准是，从该用户群体内收集优质词条的训练数据。然后，我们训练该模型，再让模型服务于他们。”

提高不同版本 Wikipedia 的质量会提高其实用程度，从而吸引更多用户。然后，这些新用户可以提供反馈，让使用当地数据进行训练的机器学习模型变得更聪明，甚至更加本地化。这是一种独特的“良性循环”本地化方法，进一步提高了宝贵的在线资源的质量。

“因此，让用户群体参与这样的工作非常重要。”Albon 说，“因为如果我们不这样做，就会有来自某个地区的一群人将他们对 Wikipedia 的特定观念强加给 [其他所有人]。人们会谈论通过机器学习模型洗白偏见。这正是我们不愿意看到的。我们希望用户群体对于自己的模型有发言权，参与这些模型的创建和管理过程。最终，我们要使用那些 [AI] 模型创造更好的用户体验。”



拐点

与 Lionbridge 的 Rafa Moral 和 Kajetan Malionowski 一起探讨机器学习与语言的关系

长期以来，翻译一直是 AI 发展的重点领域。1933 年，即 AI 成为独立学科的二十多年前，苏联科学家 Peter Troyanskii 向苏联科学院展示了“从一种语言翻译到另一种语言时选择和打印单词的机器”。在 1964 年的纽约世界博览会上，IBM 展示了一种 MT 工具，能够在英语和俄语之间进行基本的翻译。

人们很容易将 MT 视为 AI 的众多分支学科之一，而 Lionbridge 创新副总裁 Rafa Moral 认为，它是这项技术的核心主题之一。“AI 完全模仿人类智能，而最能体现人类智能的就是语言了。”Moral 告诉 AI Business，“因此，如果 AI 研究人员试图制造出行为与人类似的东西，从概念上就意味着要处理语言问题。我想这就是为什么语言从一开始就是 AI 的一个主题，乃至一个主要主题。”

MT 飞速发展

就像 AI 的其他学科一样，MT 在历史上也经历诸多变化，为近几年的突飞猛进奠定了基础，这项技术也终于走出研发阶段，向市场推出了许多成熟工具。

单词嵌入概念就是其中一项突破性进展。单词嵌入是机器学习和深度学习工具集的一部分，它是指一种单词表示方法，即意思相近的单词可以用相似的方式表示。单词嵌入过程自然地捕获单词的意思，可显著改善翻译流程，而不像以前的方法那样，不同的单词之间即使意思相似，表示也完全不同。

“单词在嵌入之前，是一个文本字符串。”Moral 说，“伴随嵌入概念而来的是矢量化概念，以及我们如何将单词转换成数字，这些数字可以含有语义。嵌入是指根据上下文来识别要与单词关联的数字或矢量。”

神经网络机器翻译：明显拐点

神经网络机器翻译是最近几年来的一项重大突破，Moral 将其称为该行业的一个“明显拐点”。神经网络机器翻译 (NMT) 不同于以前使用的基于规则的 MT (RbMT) 和统计 MT (SMT)。它是完全自动化的 MT 工具，依赖于神经网络。神经网络是受人脑启发构建的工具，为大部分的现代机器学习提供支持。NMT 会考虑所用单词的上下文，而非独立分析每个单词，因此产出的译文更好、更准确。它不仅着眼于单词，还着眼于单词之间的关系。

NMT 这样的技术为 Lionbridge 等公司的本地化服务提供了飞速发展的动力，这些公司使用此类技术高质量地翻译文档和其他文献。>

“如今的内容体量远超我们的翻译能力，是世界上所有译者可翻译数量的数百倍。”Lionbridge 产品主管兼战略师 Kajetan Malinowski 告诉 AI Business，“从这个意义上说，MT 正在帮助我们彻底改变这个行业，帮助我们管理更多的内容，帮助人类专注于可增加更多价值的领域。”

Malinowski 也像 Moral 那样表示，NMT 等方法的出现对于自动化翻译至关重要。但要产出高质量的翻译，就需要洞悉要使用何种方法，因为有些方法更适合特定的内容类型。

洞悉使用何种方法，取得理想结果

“不同类型的内容有着不同的目的。”Malinowski 说，“有些内容的曝光期非常短，而且可见性低，因此不需要超高质量的翻译即可使用。但也有些内容需要非常高的翻译质量，例如营销材料或商业广告，或者特定类型的法律、医疗保健或监管文件。MT 可让我们处理更多内容，并使人们转而开展‘附加价值’更高的任务。”

例如，Lionbridge 公司在进行翻译前就先使用 AI 工具帮助确定使用哪种方法可实现理想结果。

“我们通过一个混合过程汲取每种方法的精华，试图识别词汇密度和多样性、复杂性和变异性、质量等所有这些问题。”Malinowski 说，“这是我们的工作起点。一旦我们清楚地界定了客户的内容，就可以确定该内容类型的适合流程，即工作流程。这包括要使用的理想翻译记忆库、适合用于语言或领域的理想 MT 引擎以及适合该特定内容类型的工具集。”



一旦我们清楚地界定了客户的内容，就可以确定该内容类型的适合流程，即工作流程。这包括要使用的理想翻译记忆库、适合用于语言或领域的理想 MT 引擎以及适合该特定内容类型的工具集。

人类参与翻译流程

在许多情况下，正如本书第 5 页增强翻译部分所说，理想方法是结合 MT 与人类的专业知识。

“[MT 引擎] 已经发展得非常好，能够处理各种低可见性的内容类型。对于这些内容类型，翻译质量只需要足够好就行了。”Malinowski 说，“然后，如果需要更高的质量，可以让人类参与进来，改进内容。这样，人工翻译能专注于需要重点关注的高价值内容，确保这类内容易读、流畅、能激发出目标受众的特定情感。而这些，MT 目前还做不到。”

语言的主观性以及优秀翻译要素的模糊性对于最好的 MT 工具而言仍然是一个挑战，但现代翻译还是从这一领域的进步中受益匪浅。“与一两年相比，人类生成的内容呈指数级增长，而且在不断增长。”Malinowski 说，“MT 以及我们拥有的所有 AI 和机器学习工具正在帮助我们创造更多内容，向更多受众传达信息并更好地处理这些内容。”

Lionbridge 如何帮助 Volvo Cars 打造统一的线上市场

“我们想摆脱这种零散的感觉，打造一致的品牌形象和统一的线上平台。”

该公司曾拥有约 100 个不同的网站

当 Volvo Cars 决定在 2030 年之前实现百分百在线销售并只销售电动汽车时，他们与 Lionbridge 和商业服务与云解决方案提供商 Avanade 联手做出必要的改变。这些措施包括彻底改变线上形象，将他们的网站重新打造为一个线上市场，而不仅是一个营销工具。这就需要打造一种用户体验，让全球的潜在客户都能获得信息并感到满意。

要解决的问题

在彻底改革之前，Volvo Cars 的网络覆盖面很强，但非常分散。该公司曾拥有约 100 个网站，涉及 45 种语言。这些网页由一个中心团队创建，但交由各地市场团队翻译。这影响了整个品牌的凝聚力，因为各市场采用了不同的网站布局和销售战略。

通过将网站集中化，Volvo Cars 希望能够更加自上而下地控制本地化工作，同时集中监督翻译，并与从事这项工作的人直接联系。“随着我们 [转变为] 线上销售，我们 [需要] 重新掌控发布的信息。”Volvo Cars 全球在线数字化生产主管 Cecilia Ernby 说，“我们想摆脱这种零散的感觉，打造一致的品牌形象和统一的线上平台。”

采用的解决方案

为优化翻译流程，Volvo Cars 利用了 Smartling 翻译管理系统，将专门用来管理翻译任务的时间减少了大约 1,000 小时。其中有一项措施是，集中使用以前每个市场独有的翻译记忆库。该公司还改进了翻译工作流程，增加了一个步骤，确保根据本地搜索引擎优化 (SEO) 要求正确优化内容。

集中化和本地化：不可忽视的力量

归功于这些变革措施，现在的 Volvo Cars 网站可以面向不同市场展现统一的形象。这意味着，在英国、德国或比利时查看产品页面的客户将体验相同的用户旅程，且内容翻译质量优异。这使得 Volvo Cars 能够跨多个地域启动营销活动和发布产品。

Ernby 说：“这么多营销活动同时启动，按照以前的布局，根本不可能实现。”

此案例研究非常有价值，呈现了本地化和集中化如何协同提供一致的用户体验，而不牺牲本地客户要求的任何优势。

1,000
使用新系统节省的工时

语言技术的未来

展望可穿戴设备、增强现实和超大语言模型的潜力

关于 MT 的局限性，有这样一则经典轶事，也可能是杜撰的。据说，有人使用早期的 MT 工具翻译英文句子“The spirit is willing, but the flesh is weak”（心有余而力不足）翻译成俄语。结果翻译成：“威士忌很烈，但肉烂了。”

自行业发展初期以来，MT 和语言技术已经取得了长足进步。但是，语言技术发展还有很长的路要走，正如它的倡导者一直相信的那样，它正在成为我们生活中必不可少的一部分。

关于新一代语言技术，令人印象深刻的一些演示都与硬件相关。这些创新技术通常涉及迅速发展的可穿戴技术产品，预示着翻译将无处不在地顺畅进行，终端用户甚至都不用去考虑这件事。这非常接近科幻小说中长期以来设想的“全能翻译器”，让人能即时理解任何语言。

全能翻译器之路

这项技术的一个现实应用示例是翻译耳塞，它将语音处理和 MT 结合起来，提供口语的实时翻译。值得注意的产品包括 Pilot 智能耳塞（还能将对话转录成书面文字）、Google Pixel Buds 和 WT2 Plus AI 实时翻译耳塞。

这项技术的一个可视化产品是 Google Translate 的增强现实 (AR) 功能，叫做 Google Lens。用户使用此功能，可以将智能手机摄像头对准另一种语言的标志、菜单或其他文本，并在手机屏幕上看到覆盖原始文本的翻译，且字体通常相同。这一切通过板载机器学习工具实现，即 Lens 无需联网就能工作。在 2021 年 1 月，该公司披露，独立版 Google Lens 应用程序在 Google Play 应用商店中的下载次数已超过 5 亿次。Google Lens 支持大多数 Google Translate 语言，当前语言数量超过 100 种。

Google Lens 可能看起来很先进，代表了第一代基于 AR 的翻译工具。Lens 的初始版本发布于 2017 年年末，直到后来才支持如此多的语言。但随着智能眼镜等流行的平视显示器的出现，AR 的应用可能会发生巨大变化。

这些可穿戴设备让 AR 技术不再局限于手机和平板电脑，而是与佩戴者的周围环境更加无缝融合，无处不在。从本地化角度来看，这意味着翻译将更容易地显示在佩戴者眼前。

底层技术的进步

面向公众的演示无疑让人深刻地意识到语言技术的进步。但对于那些需要高质量翻译的人来说，真正的创新仍然是深藏于软件之中的底层技术，而非智能耳塞或智能眼镜。

正如第 5 页“什么是增强翻译？”专题中所说，在不久的将来，几乎可以肯定，翻译技术▶



16

▶将使用包含 AI 和人类译员工作在内的混合增强方法。专攻知识产权、法律和生命科学领域的语言服务提供商 Absolute Scientific Translation (ASTV) 的首席执行官 Domenico Lombardini 认为，人类与机器对立的流行观点大错特错。

“虽然新技术可以在一定程度上抑制或减少人为干预，但同样也衍生出新的职业角色和新的工作方式。”Lombardini 告诉 AI Business，“目前，越来越多的译者成为 MT 输出的后期编辑，由于机器仍然容易出错，因此人类译者的贡献至关重要。正因如此，语言专家的训练是保证最终翻译质量的一个必要条件。”

即使翻译质量最终仍然由人类裁定，但不可否认，AI 工具在增强翻译任务中依然非常实用，例如创建推荐系统，以根据译员的特定技能组合推荐处理特定项目的理想译员。专业人员不用把整个翻译任务交给算法，而是可以使用 AI 分析文档的内容，挑选出关键词和风格要求。他们还可以在交叉引用这一信息前标记文本并进行分类。

就像新一代的语法或拼写检查器一样，机器将能够越来越多地协助人类语言专家从事翻译工作。这可能意味着简单改正一个拼写错误，也可能是告知语言专家风格存在不一致或异常之处、搜索引擎优化 (SEO) 分数，或是提醒他们（和雇佣他们的公司）可能要注意的文化敏感性问题的。

大规模语言模型

AI 语言模型还可用于其他一些任务，以帮助实现本地化要求。2020 年 AI 举足轻重的一项进步是 GPT-3 的创建，这是 OpenAI 开发的一个大规模语言模型。GPT-3 的神经网络拥有 1750 亿个参数，而且模型本身使用了 45 TB 数据进行训练，对于网络上可用的内容来说，这个数量非常可观。通过吸收如此多的数据，GPT-3 可以执行准确度惊人的工作，它能根据统计信息，预测哪些字词有可能跟在其他字词后面。例如，它知道“橙”后面通常跟着“汁”。

通过利用庞大的训练数据，GPT-3 可用于各种用途，包括在执行翻译任务时切换不同语言“风格”，例如将法律文本转变为通俗易懂的英文，反之亦然。

GPT-3 这样的工具开辟了语言技术的新机遇：不只是翻译现有的文本，还生成崭新的版本。“AI 提供的语言帮助不仅限于翻译。”Lionbridge 产品主管兼战略师 Kajetan Malinowski 说，“它还帮助我们编写内容。我们一直在测试最近的 GPT-3 模型，它可应用于多种使用场景，其中最精彩的当属内容生成。”

在当今世界，用户期望更精巧和个性化的宣传内容，大型语言模型这类的创新将不仅能针对各个国家/地区提供本地化文本，甚至能针对每个用户提供个性化文本。



语言技术的未来

机器距离完全掌握人类语言能力还有多远？

在 2020 年的一篇著名的研究论文“A Set of Recommendations for Assessing Human-Machine Parity in Language Translation” (评估机器翻译与人工翻译等效性的一套建议) ([网页链接](#)) 中, 作者声称, “过去几年里, MT 的质量得到显著提高, 多项实证研究发现, MT 与专业人工翻译几乎没有区别。”

两位合著者 (苏黎世大学的 Samuel Lübli 和都柏林城市大学的 Sheila Castilho) 向 AI Business 讲述 MT 的力量和潜力。

AIB: 您在论文中指出, 过去几年, MT 有了显著进步。取得这些进步的主要原因有哪些?

Samuel Lübli (SL): 我认为主要有三个因素: 资源的大众化、算力的提升和商业利益的增加。Paracrawl 这样的计划基本上为有兴趣创建 MT 系统的所有人提供了充分的训练数据。由于神经网络机器翻译方法在概念上比早期的方法 (如统计方法) 更简单, 而且多种实现工具均可免费使用, 因此构建一个用于学术甚至商业用途的 MT 系统比以往任何时候都更容易。

使用大量的训练数据训练神经网络模型需要极大的算力, 但相比几年前, 算力价格低了很多, 所以算力问题不再影响组织或者个人, 不再是只有 Google、Microsoft 或 Amazon 等大公司才能获得的东西。另外还有商业利益。虽然事实证明使用前代技术已经

可以在翻译过程中节省大量的时间和金钱, 但神经网络建模首次能够生成真正流利的句子, 这在公众以及决策者中赢得一片赞叹。

AIB: 评估翻译准确性的最佳指标是什么?

Sheila Castilho (SC): 这个问题几乎无法回答。关于翻译评估的著作成千上万, 许多作者在方法甚至翻译质量定义上都持有不同意见。在我看来, 评价翻译质量的最好方法是, 首先确定翻译内容的质量标准, 然后由此开始设计方法。

SL: 我同意 Sheila 的看法。如果某篇翻译的目的是帮助用户解决其笔记本电脑的一个问题, 那么也许没有必要使用专业译员去找语言上的错误并进行归类。不过, 我还要加一点: 无论做什么, 都要在某个时刻引入人力。翻译质量的提升已经证明, 评估翻译质量的自动化手段变得越来越不可靠, 例如 BLEU ([网页链接](#))。

AIB: 最引人注目、最有前景的新兴翻译技术是什么?

SC: 我认为混合不同方法的一切技术都很值得关注。例如, 交互式后期编辑, 以及将语音处理与 MT 结合。

SL: 除了混合式翻译外, 我认为眼下还将出现两个变化。第一, MT 系统将能够更多地考虑上下文, 即翻译完整文档, 而非将文档拆分成句子, 独立翻译每个句子, 然后再 [粘] 到一块。

在研究环境中, 这个问题基本得到解决 ([网页链接](#)), 一经商业化, 文档连贯性将有所提高。

第二, 对 MT 系统输出的控制力将得到提升。机器已能产出质量不错的翻译, 但用户无法决定 MT 的风格等方面。未来, 人们不仅能够控制文本翻译遵从正式还是非正式风格 ([网页链接](#)), 还能控制用词是简练还是繁复, 或反映作者或说话人的性别等许多其他特征。

为了简明清晰起见, 此次采访内容已经过编辑。





AI Business
电子书系列



LIONBRIDGE

Lionbridge 与全球多家企业建立了合作关系，“冲破藩篱，共筑同心”是我们的不懈追求。25 年来，我们一直致力于以 350 多种语言提供翻译与本地化解决方案，帮助各公司与全球客户及员工建立联系。得益于我们出众的平台，我们汇聚了全世界充满热忱的专家，与各大品牌一道，以语言为砖瓦，构筑文化的桥梁。出于对语言的热爱，我们聘用了优秀的语言专家，采用了先进的机器智能技术，全心建立沟通的纽带，也使得与我们合作的企业能够打动他们的客户。Lionbridge 的总部设在美国马萨诸塞州沃尔瑟姆市，并在 23 个国家/地区设有解决方案中心。

更多详情，请访问 www.lionbridge.com



了解更多信息

