

基于在线翻译系统提升中文科技期刊国际显示度的构想

吕雪梅 程利冬 张宏 程建霞

收稿日期:2018-08-02

修回日期:2018-11-19

哈尔滨工业大学学报编辑部,黑龙江省哈尔滨市南岗区西大直街92号 150001

摘要 【目的】为突破中文科技期刊语言的限制,基于在线翻译技术探讨提升中文科技期刊国际显示度的新途径。【方法】结合非中文读者对中英在线翻译的需求和机器翻译的技术特点,提出基于在线翻译系统提升中文科技期刊国际显示度的新途径,即中文科技期刊编辑部在自己的英文门户网站上构建在线翻译系统,非中文读者可将感兴趣的文章通过论文英文信息检索、DOI解析链接、中文论文拆解、中文文本翻译以及英文文本重新编排形成较为完整的英文论文,为非中文读者提供深入阅读中文科技论文的新手段。【结果】在线翻译系统可以较为准确地将中文科技论文在线翻译成英文,使非中文读者可以较好地领会论文的学术思想。【结论】中文科技期刊出版机构通过借助人工智能、开发平行语料库和中文再加工等方法,进一步提高在线翻译系统的准确性,对于提升中文科技期刊的国际显示度,促进非英语母语国家创办自己的优质科技期刊具有重要意义。

关键词 中文科技期刊;国际显示度;人工智能;在线翻译;平行语料库

DOI: 10.11946/cjstp.201808020686

科技期刊是科研成果的重要展示平台和交流传播的基本载体,是开展学术交流、推动科技创新和促进社会发展的主要平台,在创新型国家建设中发挥着十分重要的作用。近年来,在国家相关政策和经费的大力支持下,经过出版单位和广大科技人员的共同努力,我国科技期刊已经由改革开放初期的约400种,发展到今天的5052种^[1]。

中文科技期刊在促进我国科技交流和服务中文读者方面发挥了重要作用,但在高度国际化的今天,中文科技期刊迫切需要走出国门。然而由于受到语种的限制,被SCI收录的中文科技期刊数仅占我国全部被收录期刊数的10%;2016年,Web of Science (WoS)数据库TOP期刊中中文科技期刊的平均影响因子仅为0.244。中文期刊的学术影响力偏低,国际显示度不高,与我国科技事业的发展水平不相符,提升中文科技期刊的国际显示度是我国科技出版领域的重要任务之一^[2-3]。目前,我国中文科技期刊主要通过以下3种途径提升国际显示度。

(1) 采用中英双语出版。语言是影响中文科技期刊国际显示度的最主要因素,中文科技期刊采用中英双语出版,是提升期刊国际显示度的重要途径之一。出版语言的多元化不仅可以保持期刊服务我

国科技发展的特色,而且还可以在在一定程度上满足国际化的需求^[4]。目前,国内有46种期刊采用中英双语出版,大多数期刊被WoS、EI和Scopus等国际知名数据库收录,其影响因子和他引频次等指标也有不俗表现,可以看出适度发展双语期刊是提升国际显示度的途径之一,但是中英双语出版仍受到诸多方面的限制。

(2) 与国际知名出版机构合作办刊。与国内出版发行渠道不同,国际科技期刊的组稿和审稿通常由国际知名教授或大学承担,出版发行则由专业出版集团完成,国际知名出版集团如Elsevier、Springer Nature和Wiley-Blackwell等,拥有上千种科技期刊。与国际知名出版机构合作是我国科技期刊快速融入国际学术领域的重要途径。但是,中文科技期刊在与国际知名出版发行机构合作时受到语言的限制,无法使用其庞大的英文数据库。

(3) 创办期刊英文门户网站。许多中文科技期刊已建立英文门户网站,定期更新论文英文摘要、介绍和期刊信息等;也有一些期刊编辑部利用数字化平台进行优先发表和论文推送,拓展中文期刊在国际上的影响力^[5]。杨美琴等^[6]提出以中文科技论文为核心内容,通过英文长摘要、优秀论文英文简介

作者简介:吕雪梅(ORCID:0000-0001-8856-0763),硕士,副编审,E-mail:lxmq@hit.edu.cn;程利冬,博士,副编审;张宏,硕士,副编审;程建霞,硕士,编审,编辑部副主任。

等手段再加工国内优质科研成果,从而有效提升中文科技期刊的国际显示度。

上述措施和方法虽然可以在一定程度上提高中文科技期刊的国际关注度,但是无法让国外学者深入了解优秀中文科技论文的学术思想和创新之处。近年来,随着人工智能、深度学习和大规模平行语料库的发展,机器翻译技术取得了令人瞩目的成就,Google、Facebook、微软和百度等国际知名互联网公司先后推出自己的在线翻译平台^[7-9]。中国知网 China National Knowledge Infrastructure (CNKI) 推出的“CNKI 翻译助手”,成为辅助学术文献阅读和翻译的海量中英文在线词典和双语平行语料库,在科技专业术语和句子翻译准确性方面具有独特优势^[10]。为此,本研究提出一种深入阅读中文科技论文的新途径。首先,中文科技期刊编辑部首先创建英文门户网站,及时发布最新公开发表论文的标题、摘要或长摘要、关键词、作者及其单位等英文信息,并注明每篇文章的数字对象唯一标识符(Digital Object Unique Identifier, DOI);其次,建立基于 DOI 的中文论文全文链接,便于非中文读者进一步了解感兴趣的论文;最后,在网站上推出期刊的在线翻译工具或利用成熟的翻译平台,使非中文读者能够在线翻译和阅读论文。

1 中文科技期刊在线翻译系统的构想和实践

1.1 中文科技期刊在线翻译系统功能设计

结合非中文读者对中英在线译文的需求和机器翻译的技术特点,本研究将期刊在线翻译系统的功能归纳为以下 6 个部分。

(1) 论文英文检索。读者可以使用期刊编辑部所发布的论文的题目、摘要或长摘要、关键词、作者及其单位等英文信息及 DOI 进行文献检索。

(2) DOI 正文解析链接。建立中文论文正文与其 DOI 的链接,通过点击所检索论文的 DOI,获取中文论文的正文。

(3) 论文拆解发送。接收读者论文翻译请求后,该模块将论文拆解为有、无英文信息两部分,将后者进一步拆分为章节或段落单元,并依次发送到文本翻译平台。

(4) 文本翻译平台。文本翻译是系统的核心模块,既可以委托专业公司单独开发,也可以利用现有的在线翻译软件。首先根据论文题目和关键词选定实时更新的英文专业词库,然后基于该词库进行在

线翻译。

(5) 英文文本编排。该模块将分单元翻译的英文和原有英文摘要、图表等重新编排,形成一篇完整的英文论文,供读者阅读。

(6) 读者评价反馈。读者可针对每一翻译单元的翻译质量给出“好”或“坏”的评价,该信息反馈给翻译平台以辅助深度学习,不断提高翻译的准确率。

将上述功能模块进行组合,得出所设计的在线翻译流程(图 1)。非中文读者首先通过互联网或期刊英文网站以关键词、题目、作者等信息搜索中文科技论文的英文信息,如果读者对某一篇文章的相关信息感兴趣,则可以通过 DOI 访问该论文网页,利用编辑部提供的在线翻译系统将中文论文翻译成英文进行全文阅读,突破语言限制,为非中文读者和中文科技期刊搭建沟通和交流的桥梁。

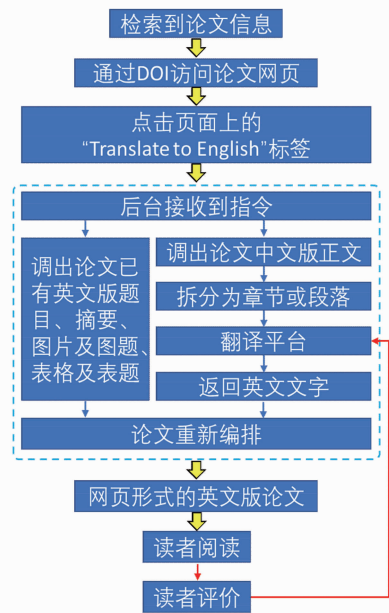


图 1 中文科技期刊网站在线翻译流程

1.2 基于百度翻译的科技论文在线翻译实践

为验证所设计的期刊在线翻译系统的可行性,本研究利用百度翻译进行中文科技论文在线翻译实践。

实例 1:以《材料科学与工艺》2018 年第 26 卷第 2 期刊载的《高强度钕合金的发展与应用》为例,百度翻译情况如下。

摘要原文:钕及其合金具有较好的尺寸稳定性、抗辐照和耐腐蚀等性能,因此在航空航天、航海、核反应堆和生物医用等领域具有重要的应用前景。而纯钕及应用较为广泛的钕合金的抗拉强度较低,提升钕及其合金的力学性能成为其能成功应用于结构

件的关键环节。本文简要概述了锆及其合金在核工业、化工、医疗及航空航海等领域的发展现状。此外,本文还着重介绍了新型高强韧锆合金的设计与制备,获取了可在空间探测、深海探测以及高速铁路等特殊领域中使用的锆系合金的最佳成分,并详细阐述了新型高强韧锆合金的强化机理及其在核电、化工及航空领域中的应用。

百度译文: Zirconium and its alloys have good dimensional stability, radiation resistance and corrosion resistance, so it has an important application prospect in aerospace, navigation, nuclear reactor and biomedicine. The tensile strength of zirconium and the widely used zirconium alloy is low, and the mechanical properties of the zirconium and its alloys become its energy. Work should be used as a key link in structural parts. This paper briefly outlines the development status of zirconium and its alloys in nuclear industry, chemical industry, medical treatment, aviation and navigation. The design and preparation of new high strength and tough zirconium alloys are reviewed. The best components of zirconium alloys used in special fields such as space exploration, deep sea exploration and high speed railway are obtained. The strengthening mechanism of new high strength and toughened zirconium alloys and their applications in nuclear power, chemical and aviation are described in detail.

中英文比较发现,除“提升锆及其合金的力学性能成为其能成功应用于结构件的关键环节”由于句子结构复杂,译文有误差,其余部分的翻译准确性较高,与原文表达的思想基本一致。

实例 2:以《材料科学与工艺》2018 年第 26 卷第 2 期刊载的另一篇论文为例。

原文: DP590 点焊接头最优的焊接工艺参数为: 焊接时间 70 ms, 焊接电流 15.0 kA, 电极压力 6.5 kN, 在此参数下接头的抗拉强度为 9521.4 N。

百度译文: The best welding parameters of DP590 spot welding joint are 70 ms, 15 kA welding current, 6.5 kN electrode pressure, and the parameters are connected under this parameter. The tensile strength of the head is 9521.4 N.

由于原文“在此参数下接头的抗拉强度为 9521.4 N”表意不清,存在语病,因此英文翻译表达不准确。

上述实例表明百度翻译系统可以较为准确地将中文科技论文在线翻译成英文,使非中文读者领会论文的学术思想。

2 提高期刊在线翻译系统准确性的措施

目前,机器翻译技术正在逐步被人们所接受,在线翻译平台已经成为各大网络公司利润的增长点,但是与专业的人工翻译相比,结构歧义和语义歧义依然是在线翻译的主要问题^[11],可从以下 3 个方面提高在线翻译的准确性。

(1) 开发适于专业科技期刊的在线翻译平台和平行语料库。目前,机器翻译的主要工作原理之一是基于实例进行机器翻译,首先给定源语言输入句,然后在双语语料库中匹配查找一个相近的句子,该句子即为原句的译文^[12],这非常适合语法结构相对简单、科技词汇重复率较大、学科相对集中的科技期刊,特别是借助机器翻译的记忆功能和质量控制功能。与专业的机器翻译技术研究机构合作,深度开发适于期刊专业特点的平行语料库,必将有效地提高译文的准确性。目前,CNKI 总库已经形成涵盖自然科学和社会科学各个领域的海量双语平行语料库,这为中文科技期刊开发专业的在线翻译系统奠定了良好的基础。

(2) 借助深度学习、神经网络、云计算和大数据等新兴技术,打造人工智能翻译平台。机器翻译技术从 20 世纪 30 年代诞生至今,其理论和方法不断发展,技术不断升级换代,但是科技论文长文翻译的连贯性和严谨性问题一直没有得到解决。近年来,随着深度学习、神经网络、云计算和大数据等新兴技术在机器翻译领域的成功应用,基于人工智能的在线翻译平台取得重要突破^[13]。周海林等^[14]选取 Google 公司的神经机器翻译系统与英语专业学生进行对比实验,发现两者翻译质量均存在一定的问题,基于人工智能的机器翻译还无法取代人工翻译,但已经成为人机结合翻译的有力保障。Facebook 公司推出基于深度学习的机器翻译系统,经过不断训练,译文的精准性和流畅性得到显著提升。国内百度、科大讯飞等互联网公司也在人工智能翻译领域发力。中文科技期刊编辑部如果能够结合自身专业领域和读者反馈意见,利用基于人工智能的翻译系统进行深度学习和训练,人工智能翻译系统必将成为中文科技期刊走向国际化的利器。

(3) 中文论文再加工。为提升中文科技论文的

在线翻译质量,编辑部需要考虑机器翻译技术的特点,结合英文科技论文特定的结构和风格,在不改变原文语义的条件下,利用标点符号、调整句子结构、消除语义歧义等方法对门户网站发布的科技论文进行再加工。

①完善无主语或隐含主语句子结构,保证中文句子的完整性。英语语法和汉语语法都是以拉丁语法为基本架构建立和发展起来的,虽然两者同源,但在语法结构上还存在一些不同:汉语主要属于意合,句子中主语省略或者隐含的情况较多;而英语主要属于形合,对于句子的完整性要求较高。为了提高机器翻译的准确性,编辑需要修改或完善论文中无主语或者隐含主语的句子。

实例1中“提升锆及其合金的力学性能成为其能成功应用于结构件的关键环节”是较为典型的例子,“提升力学性能”是整个句子的主语,应该予以明确,为此,对原句做突出主语处理,将原句改为“提升力学性能是锆及其锆合金能成功应用于结构件的关键环节”。利用百度翻译系统进行翻译,译文为“and the improvement of the mechanical properties is the application of zirconium and zirconium alloy to the junction”。翻译结果显示,机器翻译的准确性有所提升,译文与原文表达的意思较为接近。

②调整原文长句结构,提高机器翻译质量。汉语的长句通常是指字数较多、结构复杂和含有多层语义的句子。无论是人工翻译,还是机器翻译,结构松散、句界模糊和长句无连词都是制约其翻译质量的主要因素^[15]。为此,编辑可以合理添加标点符号和连词对论文长句进行断句或调整语序,使得句子结构清晰,翻译更准确。

以《材料科学与工艺》2018年的一篇来稿为例,原文表述为:为了研究不同缓蚀剂及复合型缓蚀剂对304不锈钢在酸性(pH=1)3.5%氯化钠溶液中的耐蚀性影响获得最佳缓蚀效果、最佳复配效果和最佳复配比例,采用失重法、电化学方法、表面形貌观察法、表面接触角和表面张力方法对咪唑、硫脲、铬酸钾三种物质的缓蚀效果及复配效果进行了研究。

该原文主要由2个长句组成,2个长句之间的关系较为混乱,机器翻译容易出现错误。如果将原文进行再加工,改为:为了提高不锈钢的抗磨损性能,对不锈钢等耐蚀合金通过低温活性盐浴表面处理技术提高其耐磨性的同时不降低其耐蚀性,使得其在具有磨损的腐蚀介质环境中能够提高其抗磨损性能而不被腐蚀,延长其使用寿命。

面形貌观察法、表面接触角和表面张力方法,对咪唑、硫脲、铬酸钾三种物质的缓蚀效果及复配效果进行了研究。

百度翻译为:In order to study the effect of different inhibitors and composite inhibitors on the corrosion resistance of 304 stainless steel in acidic (pH=1) 3.5% sodium chloride solution, and obtain the best corrosion inhibition effect, the best combination effect and the best proportion. The corrosion inhibition and compound effects of imidazole, thiourea and potassium chromate were studied by weight loss method, electrochemical method, surface morphology observation method, surface contact angle and surface tension method.

中英文对比可知,英文翻译较好地反映了原文内容,语法基本正确。

③修改存在语病的原文,消除语义歧义。将实例2的语病部分进行再加工,改为“DP590点焊接头最优的焊接工艺参数为:焊接时间70 ms,焊接电流15.0 kA,电极压力6.5 kN。采用该参数进行焊接时接头的抗拉强度为9521.4 N。”

百度译文:The optimum welding parameters for DP590 spot welding are: welding time is 70 ms, welding current is 15 kA, and electrode pressure is 6.5 kN. The tensile strength of the joint is 9521.4 N when using this parameter.

中英文对比可知,与原文表达的意思较为接近。

④结合英语从句句式的特点,明晰句子逻辑关系。英语在句式上更习惯采用从句形式表达长句和逻辑关系,如果编辑能够依据逻辑关系对论文小短句或者逻辑关系表述不清的长句进行再加工和组合,将会得到更为地道的译文。

以《材料科学与工艺》2017年第25卷第6期刊载的一篇论文为例,原文表述为:为了提高不锈钢的抗磨损性能,对不锈钢等耐蚀合金通过低温活性盐浴表面处理技术提高其耐磨性的同时不降低其耐蚀性,使得其在具有磨损的腐蚀介质环境中能够提高其抗磨损性能而不被腐蚀,延长其使用寿命。

原文包含2个长句和2个短句,并且长句结构关系混乱,逻辑表述欠清晰,如果将其进行再加工,改为:为了提高不锈钢的抗磨损性能,提出采用既提高其耐磨性又不降低其耐蚀性的低温活性盐浴表面处理技术,这将使其在具有磨损的腐蚀介质环境中

能够提高其抗磨损性能而不被腐蚀,从而延长其使用寿命。

百度译文:In order to improve the wear resistance of stainless steel, a low temperature active salt bath surface treatment technology is proposed, which can improve the wear resistance of stainless steel without reducing its corrosion resistance. This technology will make it possible to improve its wear resistance without being corroded in corrosive medium with wear, thus prolonging its service life.

译文不仅准确,而且前面部分采用从句,结构清晰,逻辑性强;最后采用 thus 结果状语从句,突出延长使用寿命,更加符合英语表述习惯。

3 结语

中文科技期刊是国家科技和文化软实力的重要体现,承担着向世界展示中国科技创新成果的责任。由于受到诸多因素的制约,科技期刊创办中英双语期刊并不是提高期刊国际显示度的唯一途径。中文科技期刊出版机构如果能够充分利用人工智能、云计算和大数据等新兴技术,在自己的英文门户网站构建在线翻译系统,并通过深度学习、开发平行语料库和中文再加工等方法提升在线翻译系统的翻译准确性,则可以为非中文读者提供一种深入阅读中文科技论文的新手段,有效提高中文科技期刊的国际显示度。随着中文科技期刊国际影响力的不断提高,必将吸引更多优秀的原创科研成果发表在中文科技期刊上。同时,该方法还可以推广到其他非英语科技期刊,对于促进非英语母语国家创办自己的优质科技期刊具有重要意义。

参考文献

- [1] 中国科学技术协会. 中国科技期刊发展蓝皮书(2018)[M]. 北京:科学出版社,2018.
- [2] 林强. 务实与求变:对中文科技期刊发展出路的思考[J]. 编

辑学报,2018,30(2):129-133.

- [3] 李殷,钱俊龙. 路在何方——中文科技期刊可持续发展[J]. 中国科技期刊研究,2014,25(9):1127-1131.
- [4] 芮海田,张伟伟,赵文. 中文学术期刊双语出版的发展困境与解决路径[J]. 中国科技期刊研究,2018,29(10):971-976.
- [5] 闻丹岩. 数字化内容服务平台对科技期刊国际化的影响[J]. 中国科技期刊研究,2014,25(1):148-149.
- [6] 杨美琴,谢明,程杰,等. 中文论文再加工对期刊国际显示度的影响[J]. 中国科技期刊研究,2017,28(11):1034-1037.
- [7] 王立松,王雪. 在线翻译工具辅助科技论文翻译的使用流程探讨[J]. 天津大学学报(社会科学版),2011,13(3):240-243.
- [8] 孙闪闪. 《网络时代的人工智能》的翻译实践报告[D]. 哈尔滨:哈尔滨工业大学,2016.
- [9] 罗华珍,潘正芹,易永忠. 人工智能翻译的发展现状与前景分析[J]. 电子世界,2017(21):21-23.
- [10] 田美娥. CNKI 翻译助手——中英文科技翻译的好帮手[J]. 西安石油大学学报(自然科学版),2008,23(5):109-110.
- [11] 潘正芹,罗华珍,易永忠. 机器翻译的困境、前景和出路[J]. 山东农业工程学院学报,2017,34(9):44-49.
- [12] 徐一灿,刘继斌. 机器翻译的现状和前景[J]. 海外英语,2017(21):117-118,148.
- [13] 张学强. 基于深度学习的机器翻译技术研究[D]. 沈阳:沈阳航空航天大学,2018.
- [14] 周海林,沈志贤. 谷歌神经机器翻译质量现状分析[J]. 科技资讯,2018(1):220,222.
- [15] 姚青. 科技论文汉英长句翻译技巧——以“高耐压 BST 薄膜介电变容管研究”翻译项目为例[D]. 成都:成都理工大学,2016.

作者贡献声明:

吕雪梅:提出研究选题,设计研究思路及论文框架,撰写与修订论文;

程利冬:收集与整理数据,负责在线翻译的实践;

张宏:参与在线翻译系统的实践和评价,提供论文修改建议,润色英文摘要;

程建霞:参与论文整体框架的讨论。

Conception of improving the international display of Chinese scientific journals based on online translation system

LYU Xuemei, CHENG Lidong, ZHANG Hong, CHENG Jianxia

Editorial Office of Journals of Harbin Institute of Technology, 92 West Dazhi Street, Nangang District, Harbin 150001, China

Abstract: [**Purposes**] In order to break through the language limitation of Chinese scientific journals, the new approach to improve the international display of Chinese scientific journals is discussed based on the online translation technology. [**Methods**] Based on the demand for Chinese-English online translation of non-Chinese readers and the technical feature of machine translation, this paper proposed a new approach to improve the international display of Chinese scientific journals. Editorial office established an online translation system on its English portal sites. Readers could translate the Chinese article they are interested in into English by searching the English information, DOI parsing links, editing Chinese papers, translating Chinese texts and rearranging English texts, which provided a new way for non-Chinese readers to read Chinese scientific papers in depth. [**Findings**] This method can translate the original Chinese scientific paper into English correctly, allowing non-Chinese readers to understand the academic idea of Chinese scientific papers relatively well. [**Conclusions**] The publishing organizations of Chinese journals can further improve the accuracy of the online translation system through methods such as using artificial intelligence, developing parallel corpus, and re-processing of original text, which is of great significance to improve the international display of Chinese scientific journals and promote non-English speaking country to establish their own high-quality scientific journals.

Keywords: Chinese scientific journal; International display; Artificial intelligence; Online translation; Parallel corpus

(本文责编:李翠霞)