DOI:10.19789/j.1004-9398.2017.06.005

国内计算机辅助翻译综述

姚运磊[1] 赵小兵[2]

(1. 中央民族大学信息工程学院,北京 100081; 2. 国家语言资源监测与研究少数民族语言中心,北京 100081)

摘 要

本文依据收录于 CNKI 的 1988—2016 年间 631 篇"计算机辅助翻译"相关文章,通过相关软件绘制关键词共现网络等方法进行统计分析. 对教学、翻译技术和翻译软件及其应用三个研究主题做了归纳梳理,并在此基础上对计算机辅助翻译今后的研究方向提出了几点建议.

关键词: 计算机辅助翻译,CAT 技术工具,CAT 教学.

中图分类号: TP319

1 计算机辅助翻译发展历程

计算机辅助翻译(computer-aided translation,以下简称 CAT)的研究在国外起步较早. 1946 年,美国科学家 Weaver 发表了一份备忘录,首先提出利用计算机进行翻译的构想,紧接着大批机器翻译研究机构在西方一些国家的支持下纷纷涌现,机器翻译方面的学术研究蓬勃发展. 但是,1966 年美国语言自动处理咨询委员会公布了一份报告,认为应停止对全自动机器翻译研究的资金投入,转而建议开发译员使用的机器辅助工具,CAT 才成为另一个研究的分支. 1980 年,Kay 提出了"译员工作站/台"的概念和"译员抄写员"的概念[1]. 同一时期,Melby 提出了设计多层次双向交互式的计算机辅助翻译系统的思想[2]. 二十世纪八十年代以后,随着语料库等新兴方法的出现以及翻译市场对于翻译效率的要求越来越高,CAT 借鉴机器翻译方法加强翻译记忆库和术语库管理,进入了一个新的时期.

我国最早在二十世纪初由各大学进行研发构建了翻译语料库. 早在 1957 年,机器翻译就已经被列入了国家《科学发展纲要》,然而直到二十世纪最后十几年,随着计算机和互联网才进入中国大众的视野以及我国不断深入全球化浪潮,本地化翻译需求越来越多,计算机辅助翻译才真正开始发展起来.

收稿日期: 2017-03-16

2 计算机辅助翻译介绍

计算机辅助翻译是一套人机协作的系统,计算机作为辅助译员的工具,可以协助译员在处理翻译重复率较高的翻译工作时提高翻译效率,保证翻译质量、优化翻译流程. 其中翻译记忆是计算机辅助翻译的最核心部分. 翻译记忆库在开始之初为零,随着不断学习和存储,变得越来越"聪明",效率越来越高. 此外,翻译记忆库还能够实现共享,并将语言技术和系统相结合,对于工程性翻译的翻译项目管理等模块发挥着重要作用.

简单的翻译流程如图 1 所示: 用户利用已有的原文和译文,以句子为单位进行双语对齐,建立起一个翻译记忆库,在翻译过程中,系统将自动检索翻译记忆库的翻译资源,如果记忆库中有与原文完全匹配的译文则直接输出译文,如果记忆库中没有与原文完全匹配的译文,则会自动匹配相似程度较高的译文给出参考译文,译者可以直接使用也可以修改参考译文.

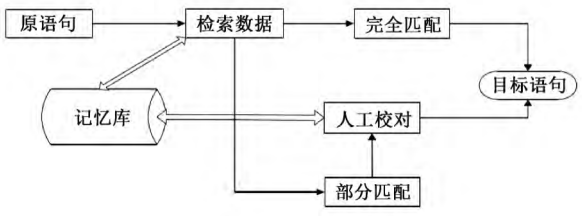


图 1 简单的翻译流程图

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

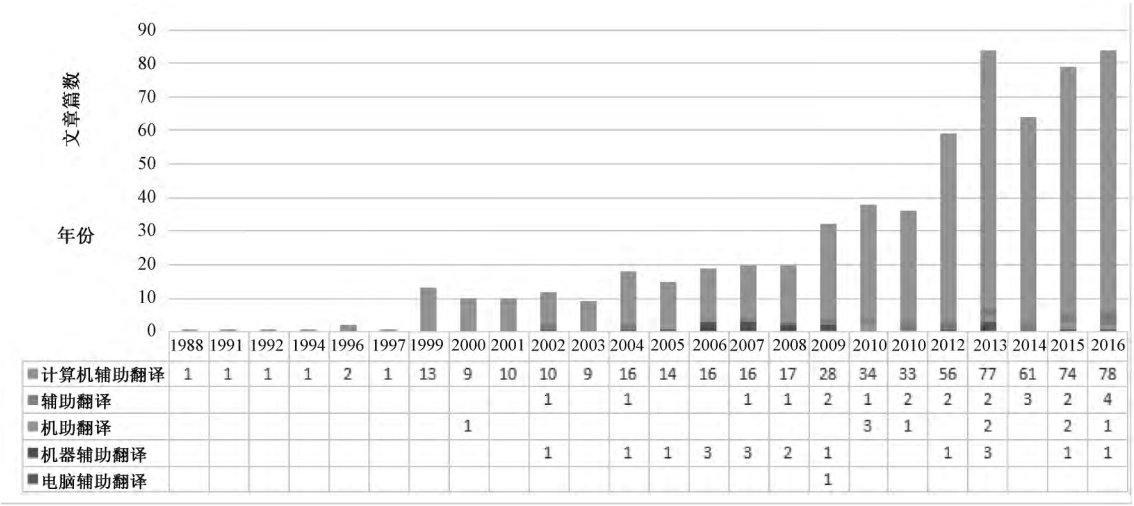
3　计算机辅助翻译的研究现状

关键词是为了文献标引工作从报告、论文中选取出来的用以表示全文主题内容信息款目的单词或术语,能够鲜明而直观地表述文献论述或体现文章主题. 因此,从关键词入手是快速跟踪某一领域研究现状的一个捷径.

笔者分别将关键词设为“电脑辅助翻译”“机辅翻译”“机器辅助翻译”“辅助翻译”“计算机辅助翻译”进行检索,在 CNKI 精确匹配检索出 1988—2016 年间 631 篇文章,并借鉴 CNKI 的计量可视化分析等工具做了进一步统计分析结果. 检索结果统计如表 1 所示,具体文章篇数和年份统计如图 2 所示. 笔者发现 CAT 主要受到国家社会科学基金、国家自然科学基金和全国教育科学规划的支持,其文章数量从 2012 年开始呈现飞跃式增长,2012 年—2016 年的研究总量占据了总研究的 58. 6%. 从研究层次来看,基础性的研究占 52. 2%,主要集中于中国语言文字、外国语言文字和计算机软件及其应用三大学科. 从作者分布来看,北京师范大学的王华树、山东师范大学的徐彬、北京大学的俞敬松以及厦门理工学院的余军较为活跃,文章产量都在 5 篇以上.

表 1　检索结果统计

关键字	电脑辅助翻译	机辅翻译	机器辅助翻译	辅助翻译	计算机辅助翻译	合计
文章篇数	1 篇	10 篇	18 篇	23 篇	579 篇	631 篇



	1988	1991	1992	1994	1996	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2010	2012	2013	2014	2015	2016
计算机辅助翻译	1	1	1	1	2	1	13	9	10	10	9	16	14	16	16	17	28	34	33	56	77	61	74	78
辅助翻译										1		1			1	1	2	1	2	2	2	3	2	4
机助翻译								1										3	1		2		2	1
机器辅助翻译										1		1	1	3	3	2	1			1	3		1	1
电脑辅助翻译																	1							

图 2　文章篇数和年份统计图

“计算机辅助翻译”的关键词共现网络(图 3)可以反映关键词之间的相关性,进而展现文章主题之间的关系,揭示文章的内在联系,能够帮助研究者从整体上把握 CAT 的研究概况. 图中的节点代表关键词,节点越大代表相关研究的数量也就越多. 连线代表节点之间的相关度,数字的数值越大,线条越粗,相关度越高. 从图中我们可以发现,“机器翻译”和“翻译教学”是 CAT 研究的中心节点. 其中,“翻译教学”还有“翻译硕士”和“课程设置”的专门研究. “机器翻译”又延伸出“翻译记忆”和“语料库”两个相关度较高的重要节点,而这两者却又是 CAT 技术的核心. 两大中心节点又有“翻译软件”“翻译技术”等交叉结点,这些交叉节点又扩展至诸如“本地化”“翻译能力”“术语库”等具体应用领域的外围节点. 笔者按照教学、翻译技术和翻译软件及其应用三个主题,去除无法归类的“计算机辅助翻译”“CAT”“翻译”三个关键词,对文章中出现频数为 10 次以上的关键词做了归纳统计(表 2),并分别就这三个方面做了评析,并在此基础上对 CAT 的未来研究方向提出了几点建议.

3. 1　计算机辅助翻译教学

CAT 的研究涵盖硕士、本科和高职院校三个培养层次. 硕士层次的研究较多,如钱多秀从 CAT 发展史、CAT 原理、CAT 工具等八个方面对 CAT 教学作了系统性分析和思考[3]. 严杰和李婷比较了国内外高校的翻译硕士(master of translation and interpreting,以下简称 MTI)培养方案在培养目标、培养方式、实践环节等方面的共性与差异,主张重视现代化技术在翻译中的应用[4],为深化翻译硕士改革以提高人才质量提供了参考. 此外,北京大学设立

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

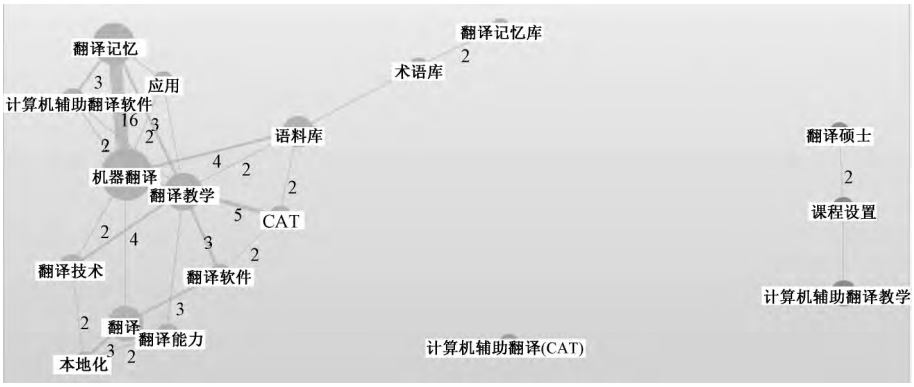


图3 关键词共现网络

表2 不同研究主题关键词频数统计表

研究主题	关键词	出现频数	小计	百分比
教学	CAT 教学	18	90	28.40%
	课程设置	14		
	翻译教学	34		
	翻译硕士	11		
	翻译能力	13		
技术	翻译记忆	53	167	52.80%
	术语库	11		
	语料库	23		
	机器翻译	61		
	翻译技术	19		
软件应用	CAT 软件	14	59	18.70%
	应用	12		
	信息技术	10		
	本地化	11		
	翻译软件	12		
总计		316	总计	100%

语言信息工程系,率先招收 CAT 专业硕士,俞敬松详细介绍了 CAT 方向工程硕士培养计划,培养面向新世纪的语言服务人才[5]. 本科培养上主要是大学英语教育的改革,如王玉钰认为翻译人才的培养不只是外语专业或翻译专业的专利,大学英语教学同样可以承担起培养翻译人才,尤其是技术翻译人才的任务[6]. 大学英语教学也要纳入 CAT 技术课程,扩大 CAT 课程的学生覆盖面. 高职院校 CAT 人才培养的研究寥寥无几,王雪飞(2015)从业界市场需求角度出发,提出高职院校要注重教授 CAT 翻译软件的实际操作,为社会输送专业化、职业化以及技能化的翻译人才.

从课程设置来看,国外以及中国台湾、香港的不少大学都很重视教授翻译技术的使用,但我国内地的高校所开设的翻译课程很少涉及翻译新技术的使用[7]. 在教育部公布的 MTI 培养课程设置中 CAT 课程只是选修课程[8],课程重点都放在文学翻译的教学上. 而在欧美高校中 CAT 课程早已纳入必修课,且开设了与术语管理工具、本地化等相关的课程完善学科建设,构成 CAT 课程体系[9]. 内地高校结合翻译技术趋势和借鉴国外发展经验对口译、笔译课程设置作了研究思考. 如宋新克(2012)把综合考虑学科类和活动类两类 CAT 课程,介绍了郑州成功财经学院基于应用型翻译人才的课程目标的 CAT 课程设置. 杨振刚(2015)主张本科 CAT 课程应翻译技术与翻译实践并重,课程设置上要以实用性为导向,根据学生实际情况,开设适合本专业的有特色的 CAT 课程和项目实践.

从教育活动来看,CAT 的研究第一是对教育者的研究,如王华树、俞敬松(2010)介绍了北大软件与微电子学院、外国语学院和北大计算语言研究所对创建 CAT 教学环境的支持还有 CAT 教师面临的压力和挑战. 第二是对受教育者的研究,如蒋春香(2016)发现 CAT 词库和记忆库在提升学生外语阅读能力方面独特效果,可以大大缩减学生的英语阅读时间,并让学生掌握了更多的阅读技巧. CAT 技术对译者提供了帮助和参考,但是译者在翻译过程中仍然居于主体地位,要充分发挥能动性和为我性[10]. 第三是对教育影响的研究,如张婷婷、崔建明(2012)等人提出了网络实验室教学、课外网上学习组、建设在线翻译资料库等新的教学模式的构想. 运用网络化的翻译教学工具和教学平台,通过专门的可实时共享的作业库和翻译案例库辅助教学. 让翻译课堂教学从以教师为中心转变为以学生为中心,学生由被动的知识接受者转变成为积极主动的学习者和具有批判精神的思想者[11].

3.2 计算机辅助翻译技术

CAT 起源于机器翻译,与机器翻译有着密不可分的联系. 所以 CAT 技术类的研究首先集中在 CAT 与机器翻译的理论研究. 如邵艳秋介绍了机器翻译

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的发展历程,以及同机器翻译相关的一些重要术语概念,包括基于规则的机器翻译,基于实例的机器翻译,统计机器翻译以及计算机辅助机器翻译. 翻译技术研究的未来发展方向应该是综合多种研究方法,优势互补、混合建模[12]. 陈永智对比了机器翻译和计算机辅助翻译在设计理念、应用范围和翻译引擎上的不同[13]. 人工翻译译者可以利用机器翻译作为辅助手段,从而大大提高翻译的效率,提高翻译质量. 然而机器翻译从目前的技术角度上讲,还只能是一种辅助手段,而不能完全替代人工翻译,机器翻译技术还有待进一步提高.

针对 CAT 的具体研究,主要聚焦于翻译记忆、语料库和术语库三个点上. 如赵善祥、刘万军介绍了多种翻译记忆中数据筛选方法,使用基于《知网》和《同义词词林》的本体论、基于统计以及语法驱动的语义相似度等评价方法,最后通过实验数据对其中四种典型方法进行比较分析,表明融合多目标评价可以提高翻译记忆中数据筛选结果的实际有效性[14]. 奥香兰(2012)通过项目实例,归纳了翻译记忆库最常见的三类问题,包括双语对齐语料译文存在质量缺陷、文本类型混杂以及更新维护不到位,并就翻译记忆库的建设和维护给出了具体建议. 王正、孙东云将翻译记忆系统引入双语平行语料库的建设中来,利用翻译记忆系统的文本对齐功能,结合汉语分词软件制作出功能齐备的双语平行语料库,有效地提高语料库建设的效率和质量[15]. 吴雪颖(2011)介绍了利用 TradosWinalign 组件,并结合常用办公软件的简单操作进行汉英文本分离,构建出能够应用于英语语言课程和自主学习的句子层级双语平行教学语料库,提供了不必使用高难的语料库技术构建简单的小型语料库的新思路. 蓝瞻瞻(2013)从双语网站获取,根据 URL 命名规律获取网页的双语材料,到段落层次的对齐,再到句子的对齐等步骤,实验证实了双语语料库的研建的可操作性. 术语翻译和术语管理是翻译流程中重要的环节,康宁、尉莹、曲乐从术语库创建、应用术语库及进行术语管理等实际问题论述了 CAT 环境下术语库建设对提高术语翻译效率和翻译准确性的重要作用[16].

3.3　计算机辅助翻译工具及其应用

CAT 工具集成了多种技术和插件,基本满足用户在翻译过程中的多样化翻译需求. 整体而言,有学者对 CAT 工具做分类研究,如俞敬松把 CAT 工具分为有成熟理论和实现的狭义 CAT 工具、面向翻译全过程的广义的 CAT 工具以及与翻译相关的其他 CAT 工具[17]. 有 CAT 工具的比较研究,如张剑从用户角度出发对 Déjàvu、SDLX 和 STAR Transit & TermStar 三款主流 CAT 软件在术语管理工具、翻译记忆库、翻译对齐工具和翻译项目管理工具四个核心模块的功能异同,为译员根据特定翻译项目和个体需要选择合适的 CAT 工具提供了参考[18]. 有对 CAT 工具的测评研究,如高志军分析借鉴《软件工程产品评价》国家标准,提出构建 CAT 工具质量模型—构建用户需求模型—准备评估—执行评估的 CAT 工具测评框架[19].

具体而言,有针对某一款 CAT 工具的使用的研究,其中以 Trados 的使用最为广泛,如汪晓婧(2016)从记忆、匹配、术语管理、多种编辑环境和自动排版功能说明了 Trados 在促进译者快速学习和提高翻译产出效率方面所发挥的积极作用. 有的针对翻译过程的某一阶段的具体问题做出讨论,如魏长宏(2010)对可移植文档格式提出了手动转换和自动转换的方法. 崔启亮等总结了译后编辑的发展动力,研究了适用于译后编辑的题材类型,提出了利用译后编辑工具提高译后编辑质量和效率的实践准则[20]. 由于 CAT 工具对应用类文本的处理效率远远高于对文学文本的处理,因此大量研究探讨了 CAT 工具在某一特定应用领域文本类型翻译中的应用,如医药化工类文本、商务英语、法律等.

需要指出的是,由于 CAT 工具的出现和使用源于翻译市场对于翻译效率和翻译质量需求的不断提高,翻译行业的工作者对 CAT 工具抱着极大的热忱,拥有大量实际经验和感悟. 因此,从文章的来源看,CAT 工具和应用的研究不再局限于高校,有大量的文章来自于一些本地化公司和专门从事翻译工作的业界人士,他们对软件行业的发展有着深入理解,为 CAT 工具的推广使用及其发展趋势提供了更广阔的视角.

4　对今后计算机辅助翻译研究的建议

(1)利用 CAT 软件辅助教学已成为翻译教学的发展趋势. 教学方面应该在以下领域做进一步研究: 一是 CAT 教材的研究. CAT 教材在开设 CAT 课程的必备要素中处于中心地位. 教师对于教学内容的把握和自身能力的提升,学生对知识技能的掌握,教学方法的创新都源于教材的编写的好坏. 但是 CAT 教材研究却很薄弱. 因此,加强 CAT 教材研究有着

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

重要的现实意义和长远发展意义. 二是 CAT 教师的培养. 既掌握计算机相关技术,又具备翻译专业的基本知识和技能的 CAT 师资严重缺乏. 加紧培养复合型 CAT 教师,改变翻译教师不懂技术,计算机教师不懂翻译的现状任务紧迫. 三是校企合作联合培养方面的探索. 如西安翻译学院与思迪软件科技有限公司合作共建了 Trados 翻译实验室. 学校和企业开展深层次合作,拓展订单式培养、合作课程开发、共建教学团队等机制方式能够拉近教学与就业的距离,以实现双方在现代化专业型翻译人才培养上的共同诉求.

(2) CAT 技术与工具的研究经历了基于规则、基于语料库、基于统计的研究方法已经越来越成熟,但是在一些问题上还不够深入. 一是汉语语法和语义构成较之英语更加复杂,套用英语研究的相关成果解决不了所有问题. 研究者还需考虑汉语的特殊性做进一步尝试. 二是随着我国社会主义市场经济的发展,我国内地之间、与其他国家的交流日益频繁,少数民族语言的服务需求急剧增长. 少数民族语言双语和多语 CAT 翻译平台的建设日趋成为各大民族高校的研究对象,但是由于社会、历史等原因,该类语料相对缺乏,面向少数民族语言的 CAT 语料库建设和完善将是一个长期的过程. 三是 CAT 技术和工具主要立足于文本的处理,语音方面的研究少之又少,目前科大讯飞在语言识别领域取得了重大突破,相信利用 CAT 语音工具改善现有的口译教学,语音云翻译技术服务在不久的将来会迎来一片春天.

此外,各省市相关部门应该加大对 CAT 研究的重视程度. CAT 研究的经费支持一半来自于国家层面,地方层面只有辽宁、河南少数省份的规划有对 CAT 研究的经费支持; 扩大跨学科研究的广度与深度对 CAT 发展至关重要. CAT 作为一门计算机技术和翻译的交叉型应用型学科,任何单一学科背景的研究者都不可能深入到 CAT 的各个方面; 翻译行业作为 CAT 技术的使用者和受益者,更应该在 CAT 的研究和发展上做出努力. 中国标准化协会、中国翻译协会、中科院科技翻译协会等组织应该积极组织业界本地化翻译公司、专业翻译人士的交流,制定 CAT 的相关标准,实施 CAT 职业技能的专业认证等.

5 结 语

我国计算机辅助翻译研究起步较晚,在二十世纪八十年代才正式开始发展. 但是在过去的几十年中我国却在该领域取得了异常丰硕的研究成果. 随着云计算、人工智能等计算机新技术的发展,跨学科研究的不断深入,以及业内人士和学术界合作的开展,相信我国计算机辅助翻译的研究在新的发展阶段会更加趋向于成熟.

参 考 文 献

[1] Kay M. The proper place of men and machines in language translation [J]. Machine Translation,1997,12(1):3 - 23.

[2] Melby A K. Computer-assisted translation system: the standard design and a multi-level design [C]. Massachusetts, USA: Conference on Applied Natural Language Proceeding,1983.

[3] 钱多秀. "计算机辅助翻译"课程教学思考 [J]. 中国翻译,2009(4):49 - 53.

[4] 严杰,李婷. 国内外翻译硕士培养方案设计的比较研究 [J]. 沈阳大学学报: 社会科学版,2013,15(6):789 - 792.

[5] 俞敬松,王华树. 计算机辅助翻译硕士专业教学探讨 [J]. 中国翻译,2010(3):38 - 42.

[6] 王玉珏,胡江. 大学英语翻译教学现状分析及改革思考 [J]. 科教文汇旬刊,2010(22):105 - 106.

[7] 吕立松,穆雷. 计算机辅助翻译技术与翻译教学 [J]. 外语界,2007(3):35 - 43.

[8] 张政. CAT 引入教学——MTI 人才应用能力培养的有效途径 [A]. 全国首届翻译硕士(MTI)教育与翻译产业研讨会论文集 [C]. 北京: 中国翻译协会,2009.

[9] 黄凤,刘文. 计算机辅助翻译在翻译专业硕士(MTI)教学中的反思 [J]. 商,2012,(20):180.

[10] 张宇浩,彭庆华. 浅析计算机辅助翻译中的译者主体性 [J]. 长春工业大学学报: 高教研究版,2014,35(1):142 - 144.

[11] 邓静,穆雷. 《象牙塔的逾越: 重思翻译教学》介绍 [J]. 外语教学与研究,2005,37(4):318 - 320.

[12] 邵艳秋. 机器翻译相关术语简介 [J]. 术语标准化与信息技术,2010(1):25 - 27 + 35.

[13] 陈永智. 机器翻译和计算机辅助翻译的对比分析 [J]. 现代交际,2016(18):154 - 155.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[14] 赵善祥,刘万军. 翻译记忆中数据筛选方法的研究 [J]. 计算机系统应用,2009(4):109 - 113.
[15] 王正,孙东云. 利用翻译记忆系统自建双语平行语料库 [J]. 外语研究,2009(5):80 - 85.
[16] 康宁,尉莹,曲乐. 计算机辅助翻译环境下术语库创建、应用与术语管理 [J]. 青岛科技大学学报: 社会科学版,2015(3):107 - 110.
[17] 俞敬松. 北京大学的翻译技术教育与翻译案例教学支持平台 [A]. 2011 年中国翻译职业交流大会论文集 [C]. 北京: 中国翻译职业交流大会,2011.
[18] 张剑. 主流计算机辅助翻译软件比较研究 [J]. 陕西学前师范学院学报,2015(4):57 - 60.
[19] 高志军. 计算机辅助翻译工具测评框架初探 [J]. 中国翻译,2013(5):70 - 76.
[20] 崔启亮. 论机器翻译的译后编辑 [J]. 中国翻译,2014(6):68 - 73.

Review of Computer-aided Translation in China

Yao Yunlei[1]　Zhao Xiaobing[2]

(1. School of Information Engineering, Minzu University of China, Beijing　100081;
2. Minority Language Center, National Language Resource Monitoring & Research Center, Beijing　100081)

Abstract

Based on 631 articles from 1988 to 2016 included in CNKI, this paper analyzes the studies by the keyword co-appearance network, and summarizes three themes of teaching, translation technology, translation software and its applications. According to the statistical analysis, the paper lists some suggestions, which hopefully could shed light on the further research of computer-aided translation.

Key words: computer-aided translation, CAT technical tools, CAT instruction.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net