

语言服务背景下油脂双语平行语料库的研制

刘瑾¹,高盼²,何东平²

(1. 武汉轻工大学 外国语学院,武汉 430023; 2. 武汉轻工大学 食品学院,武汉 430023)

随着经济一体化发展,中国粮农国际化需求越来越迫切,关注中国油脂的科技发展、加快中国油脂行业发展的国际化进程、引领我国油脂行业的现代化与国际化进程对中国粮食与农业的发展影响重大。如今国际粮食市场动荡,世界粮食结构性短缺成为我国粮食安全亟待解决的突出问题,构建国际农业发展共同体,传播中国特色的粮食农业文化,携手创建国际粮食与农业供应链的稳定发展,加快建设国际化粮食与农业的相关平台迫在眉睫。

通过查阅文献发现,国内最早的语料库研究起源于20世纪90年代,经过40多年的发展,目前建成的大型双语平行语料库有中科院汉英平行语料库、南京大学双语词典研究中心约200万对的英汉双语平行语料库、北京大学计算语言研究所约3000万对的双语平行语料库等,但对于油脂双语平行语料库的研究还较为缺乏。

本研究尝试建设油脂双语服务平台,构建国内与国际油脂发展的桥梁,助力油脂行业国际化发展,打造国内首个特色型油脂双语在线平台,为培养油脂国际化人才提供方法,建造中国特色油脂文化双语库,为油脂国际化研究人员提供学习的工具,促进油脂行业与文化“走出去”。

1 油脂英汉平行语料库建设内容

1.1 油脂英汉双语平行语料库构建流程

构建该语料库的总目标是为了建设油脂双语平台,建立双语文化长廊,建立双语专业词库。构建流程分为5个阶段:第一个阶段是语料选择与收集,第二个阶段是语料加工,第三个阶段是语料导入机器,第四个阶段是语料检查,第五个阶段是双语语料库导入平台。

1.2 油脂英汉平行语料库的研制

1.2.1 语料库材料选择

《贝雷油脂化学与工艺学(第六版)》(第五卷)

(Fereidoon Shahidi,2016)共有13章节,介绍了油脂的加工技术,具体内容包括:油脂加工工艺概述,油脂制取,油籽和脂肪原料中油脂的获取,油脂的储存、处理、运输、包装,油脂的吸附分离、脱色、脱臭、氢化加工技术,植物油深加工超临界技术,油脂的膜处理,人造奶油加工厂和设备,油料饼粕挤压生产食品和饲料。该译本翻译语言通俗易懂,用词严谨,忠实于原文的意群,语意清晰,英译汉术语翻译准确,为科技类文本翻译提供了较好的参考。

1.2.2 语料库设计

语料库的设计是指关于语料库的总体建设规划。研究人员需要明确语料库建库目的、意义和语料库的类型,分析语料库的规模与代表性及语料加工的程度等。油脂双语语料库主要从语料术语来源设计、语料平台的功能设计、语料的加工设计进行探讨。语料的术语来源主要选取国内各大粮食类词典,如《粮油术语标准汇编》《粮食英语词典》《油脂英语词典》等。句段对齐语料库收集粮油的专业书籍,如《贝雷油脂化学与工艺学(第六版)》(第五卷)。语料平台的功能设计为查词版块与句子对齐两大版块。语料的加工需要对语料进行采集提取、清洗以及机器导入与人工校对等步骤。

1.2.3 语料库制作过程

(1)语料的选择。语料库的建设关键在于语料库的精准度选择,因为其影响整个语料库建设的意义与效果。根据D. Summer提出的书面语料选择的七大原则,本语料库选择语料时遵循了学术价值或影响力原则、档案馆中文本的可访问性原则以及人们阅读习惯的统计采样原则。在油脂行业具有权威影响力的语料有《粮油标准汇编》系列书籍以及《油脂词典》,具体内容包括粮油加工、粮油特性、油脂产品、油脂物流技术、油脂加工技术、油脂储藏科学与工程及仓库建筑工程、粮油食品质量与安全检验与控制技术,以及粮油贸易业务等方面的专业术语。具有学术价值的书籍会选择本行业权威专家的专著或教材译本等,重点以《贝雷油脂化学与工艺学(第六版)》(第五卷)的译本为主创建双语语料。

(2)语料加工。语料加工分为语料的提取、转

基金项目:2021年教育部协同育人项目“面向语言服务的食品英语语料库建设”(202102483032)

作者简介:刘瑾(1986),女,副教授,主要从事语料库翻译研究(E-mail)208962582@qq.com。

通信作者:何东平,教授(E-mail)hedp123456@163.com。

换与语料的清洗与整理,此步骤的目的是将选择的纸质或者电子资料转换成整洁的双语对齐格式。此阶段主要采用人工输入、OCE 识别软件进行扫描,运用软件 Acrobat、手机拍照识别软件、WPS 等工具,对 PDF 格式文本进行识别,并批量转换为 Word 格式、Excel 格式或者 TXT 格式进行保存。完成语料的转换后,需要对语料进行降噪(noise elimination)并以纯文本的方式保存,遵循语料文本的基本规范。降噪以消除文本中多余的空格、乱码或者是单词的拼写错误,在制作油脂双语词典过程中,作者发现转换后的文本会出现大量识别错误的单词,此时需要人工在 word 上依照原语料进行校对,并将所有的术语以一系列英语、一系列中文的方式单列排列整齐以便导入机器或者通过 Microsoft word 处理文本,也可以录制“宏”工具,录制好所有的文本编辑替换或者增删段落标记等相关执行命令,以后该工具按钮会出现在工具栏直接点击图标即可。或者结合 Excel 表格功能制作对齐形式。此外,还有相关软件可以对文本格式进行编辑,如 EmEditor 软件,找到“Find”查找功能和“Replace”替换功能,使用正则表达式查找或者替换一些常规的字符合。

(3) 语料对齐。除了人工的语料加工外,通过降噪处理后的干净文本可以直接导入机器的工具中进行对齐处理,如 Tmxmall、ABBYY Aligner、Transmate、ParaConc 软件等或者机辅翻译对齐工具进行语料对齐,转换成 TMX 或者 EXCEL 等格式保存。

1.2.4 语料库检索平台的设计

双语对齐的语料库做好后实现了语料积累的第一步,如何存储语料,如何运用语料是语料建设与运用的关键环节。目前,支持平行语料检索的工具主要有新西兰奥克兰大学的迈克尔·巴洛(Michael Barlow)设计开发的 paraConc 和北京外国语大学的 BFSU paraConc。同时也可设计网页版的检索平台,语料平台的设计根据用途进行功能设计。本平台预设英译汉、汉译英两大版块,除了选取《贝雷油脂化学与工艺学(第六版)》中英语料外,还将收录油脂类术语和油脂类双语典籍及相关油脂类双语科普知识与资讯,创建一个多维的多功能翻译传播平台。

2 面向语言服务的油脂语料库平台运用

语料库与翻译教学的结合可以改变传统教学中以“教师为中心”的突出问题,学生可以通过“发现式学习”和“数据驱动式学习训练”掌握翻译技巧,提高翻译技能,熟悉翻译职业特征。通过建立本校特色的油脂语料库中心,搜集,加工后的语料库适用于词汇教学,文献阅读教学,写作教学,油脂双语文

化教学,油脂类翻译教学等。教学方式上学生可以自由选择时间学习,学生对于某一不确定词汇用法,可以搜寻语料软件,对比两种语言的异同,更加真实直观地理解双语词汇的差异。通过双语例子的语境分析,学生更加透彻地领悟该词语的语言特点及油脂文化背景。基于语料库的翻译教学模式更加直观化、清晰化、动态化与生动化。学生的内驱动力得以挖掘,探索问题、分析问题的能力得以锻炼,带着疑问进行探索并解决问题,从而在翻译选词与造句中提升自己的翻译水平。教师的角色也由传统的主导地位变成以学生为中心,通过设计油脂文化双语平台,给学生分组,针对不同的翻译问题进行探究,如“油脂制作”类别中的英语专业术语与句型是否有固定模式,不同区域的油脂历史文化如何用英语简述,是否能从众多语料中发现其语言规律。通过项目式教学模式,引导学生解决翻译中遇到的实际问题,让课堂“活”起来,例如某一小组通过探究油脂文献的双语语料,用英语学习油脂专业知识,实现特定的沉浸式学习模式,对“翻译+专业”的特色人才培养有一定参考意义。

在翻译实践运用方面,面对如今专门类翻译语料库的稀缺,基于具有世界影响力的《贝雷油脂化学与工艺学(第六版)》作为语料资产可以丰富食品类语料库,为计算机辅助翻译软件提供翻译记忆库,为开发机器翻译引擎提供资源。学生在训练油脂类翻译素材时可以直接调用油脂类术语库及相关的翻译记忆库,在预翻译环节可以高效准确地节省时间,提高翻译质量与效率。运用油脂语料库,分析科技英语翻译的翻译策略与特点,有利于油脂类翻译人才的培养。

在油脂文化传播方面,双语或多语的油脂文化平台为海外人员和英语学习者提供了海量且生动的学习资源,用英语讲好中国油脂历史与文化,保护中国传统非遗农业文化,打造国内特有的油脂海外文化传播建设点,为国家的农业文化“走出去”提供参考与借鉴。

3 结语

建设油脂双语语料库可解决当今翻译特色教育与油脂英语跨学科教育问题,实现翻译学科与油脂学科的融合发展,同时丰富专门学科类语料库的内容,推进油脂文化国际化发展。以《贝雷油脂化学与工艺学(第六版)》双语例句建设语料库可以为油脂类翻译研究提供研究文本,丰富科技文本的翻译策略并强化翻译技巧,可为油脂类翻译文化“走出去”做好铺垫,同时可建立油脂类翻译范本规范。