

论法律语言平行语料库的构建

蒋 婷

(重庆大学 外国语学院, 重庆 400030)

摘要:语料库技术的广泛应用为语言对比研究和翻译研究开辟了又一广阔的发展领域。相比于其他语言学分支,法律语言学者大多仍然固守传统的直觉和内省法对语言进行定性研究,而利用海量语料、采用经验主义研究方法对法律文本进行的研究却寥寥无几。因此,构建法律语言学平行语料库已成为法律语言研究必须直面的一个紧迫任务。本文针对并结合平行语料库和法律语言的各自特点,从法律词汇学、法律句法学、法律语义学等六方面探讨平行语料库在法律语言研究方面广阔的应用前景。

关键词:平行语料库;法律语言学;翻译;应用

中图分类号:H313.9

文献标识码:A

文章编号:1008-5831(2005)04-0094-04

On Construction of Forensic Parallel Corpus

JIANG Ting

(College of Foreign Languages, Chongqing University, Chongqing 400030, China)

Abstract: The wide use of corpus technology has opened a door to contrastive language research and translation study. Compared with other linguistics disciplines, forensic linguistics researchers still adhere to intuition and retrospection methods; consequently, there are few studies based on abundant corpora and empirical methods. Therefore, the construction of forensic parallel corpus becomes an urgent task faced with forensic researchers. This paper combines the characteristics of parallel corpus and forensic linguistics, and probes into the promising application of parallel corpus to forensic linguistics on six aspects.

Key words: parallel corpus; forensic linguistics; translation; application

20世纪70年代以来,语料库语言学迅猛发展,单语语料库建设展露出勃勃生机。至90年代,双语平行语料库开始受到语言学界和自然语言处理学界的青睐。它在两种语言对比研究、翻译和机器翻译、翻译教学、双语词典编撰等方面都具有极大的应用价值。然而,在语言学的各个领域都积极利用语料库这个强有力的工具时,法律语言平行语料库的建设和开发却在国内仍属空白。因此,在法律语言研究的领域内,是否有必要建立法律语言平行语料库,建立平行语料库是否切实可行,以及如何建立平行语料库,这一系列问题是法律语言研究者必须直面的重大课题。

一、平行语料库的建设和应用

语料库(Corpus)是一个将真实情况下使用的海量语言信息经过科学收集和组织而集成的专供研究使用的资料库。语料库语言学就是在文本语料的基础上进行语言研究的一门学科。它的基本特征是以真实使用中的语言事实为研究对象,采用概率法对语言事实进行分析,目的是探寻语言交际规律和进一步理解人类语言。自20世纪60年代初,第一代大型电子语料库——布朗(BROWN)建立以来,语料库语言

学经过了40年的发展历程,日臻成熟,目前正朝着不断扩大库容、深化加工和不断拓展新的领域等方向继续发展。其中一个重要的表象就是平行语料库的建设和研究,它代表了当今基于语料库方法进行语言对比研究、翻译研究和外语教学研究的发展趋势(甄凤超等,2004:73)。

“平行语料”(Parallel Texts)是指使用不同语言撰写、相互间具有“翻译关系”的文本。在计算语言学界,它有别于“对比语料”(Comparable Texts)。后者也使用不同的语言撰写,并且针对同一主题,但相互之间却不存在直接的“翻译关系”。平行语料库是把两种语言完全对应的文本收入计算机,通过分析对比,找出两者对应关系,用于探索“同一内容是如何用两种语言表达的”(肖忠华,2004:37)。它将源文本与其译语文本进行对照检索并显示,较之词典和单语语料库有独特优势。平行语料库为翻译和语言对比研究提供了重要资源,它的研究范围包括:(1)从语义、句法、语用等方面进行语言对比,深入了解所对比的语言;(2)通过系列比较,揭示双语的共性和特性,以及文化差异;(3)揭示原文与译文、母语与第二语言或外语的

收稿日期:2005-04-26

基金项目:重庆大学语言认知及信息处理研究所2004年基金项目

作者简介:蒋婷(1974-),女,四川乐山人,重庆大学外国语学院讲师,主要从事语料库语言学、专门用途英语研究。

差异;(4)以大量的双语对照语言实例、广泛的语域、较强的时效性、丰富的语境等特点,广泛用于双语对比研究、翻译、机器翻译、双语词典编撰。同时教学中采用数据驱动的教学法,学习者从平行语料库中能自主发现母语与目的语在语法结构、词汇和语言使用上的差别,进而学好外语。随着平行语料库的不断研制和完善,基于语料库的方法论体系必将成为翻译研究和语言对比研究的主流。

随着从事语言对比研究和翻译研究的学者对平行语料库重要性的逐渐认识,英国、挪威等国的许多研究机构和大学已相继建立各种双语甚至多语平行语料库,如 The English - Norwegian Parallel Corpus (ENPC), The English - Swedish Parallel Corpus (ES-PC) 和 The Portuguese - English Parallel Corpus (PEPC) 等。我国也大力开发和建设平行语料库,始建于 2001 年的 Babel 汉英平行语料库最初是服务于一个面向新闻领域的机器翻译系统,为基于其中存储的引擎提供翻译实例。目前,该语料库的规模已达 20 万句对。中国外语教育研究中心的“北外通用汉英平行语料库”(PCCE)(王克非主持研制)是目前国内最大的双语平行语料库(含汉英和汉日两个库),也是国际上涉及汉语的最大双语语料库。Sinclair 认为上世纪 90 年代末大型语料库建设的势头趋缓,取而代之的是大批小型语料库的兴起(甄凤超,2004:74)。现代信息技术的发展为语料库的建设提供了物质基础,使许多研究者根据自己的兴趣和研究方向建立形式多样的平行语料库成为可能,如医疗保健领域的英汉平行语料库、文学作品原著和译文平行语料库等等。

在平行语料库的应用领域中,翻译对比研究是一个重要分支。目前,许多研究者纷纷利用平行语料库进行翻译研究,如秦洪武、王克非运用英汉平行语料库,对 so... that 结构和它的汉语对应结构进行描述和分析(秦洪武、王克非,2004:41)。在《“一……就”的英译——兼谈英汉平行语料库辅助汉译英调查》一文中,作者利用个人英汉平行语料库,对“一……就”这一汉语常用词的种种英译进行了再调查(谢家成,2004:19)。张美芳在《利用语料库调查译者的文体——贝克研究新法译介》一文中介绍了贝克利用语料库调查译者文体的尝试,以此探讨影响译者的文体与风格的社会文化因素(张美芳,2002:54)。在《双语库:翻译研究新途径》一文中,作者从四方面详细阐述了基于语料库进行翻译研究的意义和方法(柯飞,2002:36)。

尽管运用语料库进行翻译研究已经取得一定成果,但这些为数有限的成果往往局限于基于语料库对通用英语进行翻译研究,却在科技翻译研究和语言对比研究方面,鲜有涉足。

二、法律语言学及其研究现状

法律语言学 (forensic linguistics) 是在语言学和法学交叉点上产生的新学科。它主要采用语言学的理

论和研究方法,以法学领域的语言为研究中心,专门研究法律和语言关系的一种专门用途语言学。法律语言学与其他语言学不同之处在于它不是旨在寻找语言的一般规律,而是根据语言的一般规律研究法律语言的一般规律和特殊性(社金榜,2000:211)。法律语言学包括法律语言本体研究和应用研究,本体研究由法律语音学、法律词汇学、法律语义学和法律句法学构成,应用研究分为法律修辞学、法律语用学、法律语篇学、法律语言心理学和法律翻译。法律语言作为具有权威性和约束力的法律载体,除具有一般语言的特征外,还有以下特点:(1)概括性与准确性;(2)一致性与灵活性;(3)明晰性与隐含性;(4)专门性与通用性;(5)简约性和繁杂性;(6)主观性和客观性。(社金榜,2000:211)

纵观法律语言研究历程,我们发现它与语言学整体的发展历史一脉相承,即法律语言特征的静态研究(句法特征词汇的研究)——法律语言作为过程的动态研究(应用研究、话语分析)——法律语言作为工具的研究(社会语言学研究)(廖美珍,2004:23)。从目前国内文献看,法律语言本体研究基本上还停留在法律语言的表象层面,如词汇、句法等的研究。很多研究者从词汇、语法、句法和文体特征各个角度,初步探讨了法律语言学的特点,如《法律英语的文体特点及英译技巧》(季益广,1999:31)、《法律语言学、法律语言——兼谈法律英语的特点》(马庆林,2003:12)、《法律英语中词语的特色》(凤智,2004:29)、《法律英语语句结构探析》(唐震熙,2000:17)等。

另一方面,使用研究主要集中在法律翻译方面。作为法律交流的桥梁,法律翻译处于法律和翻译的交叉点,在法律活动中起到了举足轻重的作用。随着法律全球化的进展,法律翻译日益重要(社金榜,2004)。从事法律翻译的研究者和实践人员,从英汉法律术语翻译、法律法规翻译、法律文献翻译、合同及法律文书翻译等的实例中,总结归纳出法律语言翻译的一些规律,如《中国法律法规英译的问题和解决》(社金榜、张福,2004:72)、《法律英语词汇翻译切忌望文生义》(陆文逊,1998:32)、《“撇开”原文写译稿——浅谈法律合同文件的翻译》(孙喜民,1998:14)等。虽然近年来,法律翻译硕果累累,成绩喜人,但仍然存在不少问题,如拼写语法错误、法律术语翻译错误、文体不当、对原文理解不透、法律文化差异等诸多问题。

另外,从法律语用学、法律修辞学、法律语言文体学等角度进行的英汉对比研究仍处于起步阶段。在这些领域,大多数研究者仍然固守直觉和内省法对语言进行定性研究,而利用海量语料、采用经验主义研究方法对法律文本进行的研究仍然寥寥无几。显然,在法律翻译及其对比研究方面,平行语料库建设的严重滞后已极大限制了它自身的发展,制约了其从定性向定量研究方法、从理性主义向经验主义研究理念的

转变。同时,平行语料库在法律翻译领域建设的空白状态,不仅显现该领域研究资源的局限性,而且这种局限性也反过来导致研究者更加依赖于传统的定性研究方法,二者相互依赖又相互制约。

鉴于此,建立法律语言平行语料库毫无疑问是法律语言对比研究的一个新起点,它提供了大量的真实语料和可靠的量化数据,促使研究人员用可比语料库对法律语言作进一步分析和研究。

三、平行语料库——法律语言研究的新途径

法律语言研究离不开对法律语言材料的充分收集、整理和分析。翻译及语言对比研究至少涉及两种语言,其研究自然更离不开大量的语言材料和翻译实例。建立在现代计算机技术基础上的法律语言平行语料库,代表了法律语言研究的新手段,将在法律语言研究方面起到不可估量的作用。

杜金榜把法律语料库定义为:“用于法律活动和法律语言学研究的语言信息库”(杜金榜,2004:211)。这就明确了法律语言平行语料库的基本用途体现在两个方面:用于对比研究法律语言特点和法律实践。同时,整个语料库可分可合:分开可开展单项研究,如双语词典研编,翻译文体和翻译教学研究;合则可在大规模语料基础上进行词频、搭配、对应词、句型、语体、文体等方面研究。以下结合法律语言学本体研究和使用的角度,提出平行语料库应用于法律语言学的初步构想。

(一)法律语言平行语料库和法律词汇学

基于语料库的法律词汇学,主要探讨某个词的出现频率、出现的语境和邻近搭配。例如,在文体特征中,频率特征是最重要的特征之一。在法律英语中大量使用了专业术语(technical terms)、古体词(archaic words)、外来语(foreign words)、词语并列(words of conjunction)和带有指令性和强制性的shall,来体现法律正规、严肃、古板的文体特点。语料库可以通过KWIC(Key Word in Context)的索引工具来进行这些词汇的搭配词统计、词语型式(pattern)统计。也可用WordSmith Tools进行词丛统计(word cluster)、主题词提取(key word list)与词图(plot)。为研究法律词汇特征提供了可靠的量化依据。由于语料库中的语料提供了大量与原文有关的信息,如作者、时代、体裁、词类附码等,所以很容易确定某个词或词组在特定体裁、变体中的用法。

法律平行语料库的另一直接用途就是为编纂法律双语词典提供了大量真实例句。在语料库问世以前,词典的编纂一直依靠经验来收集语料。语料库的出现和运用能使语言研究者从几百万的机器可读语料中检索出一个词或者词组的例句,因此,凭借大型双语平行语料库编纂一部法律翻译参考词典一定非常实用。

(二)法律语言平行语料库和法律语义学

法律语义学是法律语言特点研究的基础,它的研究最贴近法律语言本体的研究(杜金榜,2004:92)。法律语义学研究法律词语的意义和意义变化,特别是从社会和历史角度探讨词语变化的原因和规律,如词义的扩大、缩小,升格和降格,词义转移、语义和句法结构的关系等等。利用语料库,用机器自动分析和人工分析相结合来研究法律词汇和语法结构之间的联结模式,即法律词汇和他们语法环境之间的关系,和语法结构与它们法律词汇之间的结构,并以此区分同义或近义的语法结构。也可对语料库里某些特定词汇进行赋码,对比分析英汉法律词汇的模糊性或冗余性。

(三)法律语言平行语料库和法律句法学

语料统计说明语法特征在不同语域中有系统的分布特点,因此利用语料库,可在不同层次上研究法律语言的句法,包括形态特征、语类和句法结构,即在有了语法标注的平行语料库里研究各种词性标记的组合模式和对各类句型的量化分析。法律语言研究者可在语料库里面调查立法语言、司法语言和其他法律语言的结构频率分布,调查句法结构与语言其他分析层次的关系,以及语言因素和非语言因素之间的关系,用实证方式对句法资源的各种使用模式作出定性定量分析。也可具体到动词的时态、语态,情态动词或者分句的语法特征。

(四)法律语言平行语料库和法律篇章分析

语篇分析的目的是描述超出句子范围的语言特征。确定和分析法律语篇特征比确定和分析词汇和句法特征难得多,但对描写法律语言十分重要,因为许多词汇和句法特征只有通过它们在语篇中的功能才能得到完美解释。以前的语篇分析一般使用文本的片断作为例证,但这种分析一般不使用定量分析方法,也不研究不同文本之间以及不同文本之间、不同语域之间篇章特征的异同(Biber,2000:68)。语料库的方法可以处理大量文本,精确地描写所选语域的语篇特征,以及文本符合其语域的一般语篇模式程度,这就弥补了以前语篇分析的不足。法律类平行语料库可从语域理论入手,用互动式计算机技术分析方法,从社会学角度来分析法律语言的三个变项:语场、语旨和语式。也可应用Biber介绍的以语料库和计算机技术为基础的多维度分析法(multi-dimensional analysis)来分析法律语言不同语域之间的相同点和不同点,确定有不同功能的大量语言特征,归纳出它们的语法范畴,用“因素分析法”来识别和统计每种形式的出现情况,总结出不同变体的共现模式和变异维度(variation dimension),为研究法律语域变体提供多层次的视角。

(五)法律语言平行语料库和法律翻译

在世界国际化进程不断加快的今天,法律文化交流越来越频繁,一个能满足法律语言研究者和翻译工作实践人员需求要的法律语言类平行语料库已成为

一种迫切需要。利用平行语料库可以进行翻译转换研究。双语语料库收有平行对应的语言材料,据其进行翻译转换研究最合适。翻译转换不仅涉及词语、修辞手法,还包括各种句式,如存在句、无主句、被动句、把字句、省略句等(柯飞,2002:36)。

以往对译作的研究、评论涉及文体、风格的不多,主要是经验、感受型的,因为很难对长篇译作进行文体风格的量化研究。利用双语语料库则有可能比较便利地研究译文的语言特性和译者的文体(柯飞,2002:36)。利用平行语料库,可以对法律语料进行附码标注,并标注许多超语言信息,如对译者情况(包括译者姓名、性别、民族、职业、翻译方向等)、翻译方式、翻译类型、源语、原书情况、出版社等等均予以标注。这些都是考察法律翻译及翻译文体/风格的重要信息,因为译者对所译文本类型的选择、翻译策略的选择,以及他在前言、后记、注释中的表述,都可能表现出他的翻译动机、风格或取向。此外,语料库还可开展对某一译者的多译本分析,对某一原作的多译者的译本分析,对男女译者的译本分析,更可以根据大量语料对译者个人偏爱的语言表达形式(如词类/标记比率、句子长度、词频、句型、搭配方式、叙事结构等)加以分析,发现更有说服力的翻译文体/风格表征。

就法律翻译实践而言,法律语言类平行语料库为译员和翻译专业的学生提供了一个工作平台和参考工具,可用来提高语言及文化意识。在对法律各领域的理解、法律术语的准确选用及习惯表达等方面,用平行语料库辅助的译文比用传统资源(如只靠词典)的译文质量更高。研究表明,把外语翻译成母语时,参照对应库核对有关翻译问题能够帮助译者提高表达能力,而且译文中的错误也会减少。把母语译成外语时,对语料库工具的需要就成倍增加,而且语料库的用途也远远不止是核对母语中理解难点。因此,法律翻译实践者可以随时抽取语料,形成小型语料库,以提高译者理解原文,并用更流畅的目标语进行翻译的能力。

(六) 法律语言平行语料库和法律英语教学

语料库解决了大量语料存储的问题,对于翻译工作者或翻译教师,最便利的莫过于将所收集的多种译本存储于其中,并随时可调出加以对比研究(柯飞,2002:38)。在法律英语翻译教学中,教师借助大型语料库检索,为教学提供大量、详实的翻译范例,这样更利于学生形成法律语言的“意识”、“反射”和“应变”。另外,教师也可利用索引软件,分析法律语言的常用词表、非常用词表,方便、快捷地为法律语言教学大纲设计数据,也为法律英语教材编写和测试提供大量真实材料。而且,教师可通过索引,让学生自己通过实例观察词或词组的实用语境,也可开发制作“数据驱动”学习课件,对词语索引行在屏编辑,可屏蔽搜索词,用索引行直接制作成填空练习,使学习者在真实

语境中掌握某一语法结构或词语用法(卫乃兴,2001:171)。学生要带着问题自己到语料库中去发现答案。这些问题可能是学习过程中出现的问题,或者是教师根据需要预先设定的问题。学生通过自己分析语料库的检索结果最终将问题解决。这种“发现式学习”方法,更能激发学生的学习动机,牢记自己所发现的知识。

参考文献:

- [1] BIBER D, S CONRAD, R REPPEN. *Corpus Linguistics* [M]. London: Cambridge University Press, 2000.
- [2] GRAEME KENNEDY. *An Introduction to Corpus Linguistics* [M]. London: Cambridge University Press, 2000.
- [3] 杨惠中. 语料库语言学导论[M]. 上海:上海外语教育出版社, 2002.
- [4] 杜金榜. 法律语言学[M]. 上海:上海外语教育出版社, 2004.
- [5] 张美芳. 利用语料库调查译者的文体——贝克研究新法评介[J]. 解放军外国语学院院报, 2002, (3): 54-57.
- [6] 肖忠华. 平行语料库与可比语料库在语言研究中的应用[J]. 中国文字学报, 2004, (4): 21-25.
- [7] 杜金榜, 等. 中国法律法规英译的问题和解决[J]. 中国翻译, 2004, (3): 72-76.
- [8] 穆晓莉, 尹转云. 语料库在语言研究方面的应用综述[J]. 西安外国语学院院报, 2001, (3): 21-26.
- [9] 李亮. 基于语料库的机器翻译[J]. 上海科技翻译, 2004, (2): 59-62.
- [10] 谢家成. “一……就”的英译——兼谈英汉平行语料库辅助汉译英调查[J]. 中国科技翻译, 2004, (2): 27-29.
- [11] 廖美珍. 国外法律语言研究综述[J]. 当代语言学, 2004, (1): 66-76.
- [12] 秦洪武, 王克非. 基于语料库的翻译语言分析——以“so...that”的汉语对应结构为例[J]. 现代外语, 2004, (1): 41-48.
- [13] 谢家成. 小型英汉平行语料库的建立和运用[J]. 解放军外国语学院院报, 2004, (3): 27-30.
- [14] 柯飞. 双语库: 翻译研究新途径[J]. 外语与外语教学, 2002, (9): 36-39.
- [15] 马庆林. 法律语言学、法律语言——兼谈法律英语的特点[J]. 西安外国语学院院报, 2003, (3): 12-15.
- [16] 凤智. 法律英语中词语的特色[J]. 上海科技翻译, 2004, (3): 29-31.
- [17] 甄凤超, 张霞. 语料库语言学发展趋势展望——2003语料库语言学国际会议综述[J]. 外语界, 2004, (4): 74-76.
- [18] 季益广. 法律英语的文体特征及英译技巧[J]. 中国翻译, 1998, (5): 31-34.
- [19] 陆汶逊. 法律词汇翻译切戒望文生义[J]. 上海科技翻译, 1998, (2): 31-32.