

译学新论丛书

主编 张柏然 许 钧

平行语料库与积极型汉英词典的研编

李德俊 著

上海译文出版社

图书在版编目(CIP)数据

平行语料库与积极型汉英词典的研编:基于平行语料库
的词典编纂系统 CpsDict 的研制与应用/李德俊著.

—上海:上海译文出版社,2008.1

(译学新论丛书)

ISBN 978-7-5327-4405-3

I. 平… II. 李… III. ①语言学-研究②英语-汉语-
词典编纂法-研究 IV. H0 H316

中国版本图书馆 CIP 数据核字(2007)第 166821 号

本书中文简体字专有出版权
归本社独家所有,非经本社同意不得连载、摘编或复制

平行语料库与积极型汉英词典的研编
——基于平行语料库的词典编纂系统 CpsDict 的研制与应用
李德俊 著

上海世纪出版股份有限公司

译文出版社出版、发行

网址:www.yiwen.com.cn

200001 上海福建中路 193 号 www.ewen.cc

全 国 新 华 书 店 经 销

上海锦佳装璜印刷发展公司印刷

开本 890×1240 1/32 印张 8.25 插页 2 字数 172,000

2008 年 1 月第 1 版 2008 年 1 月第 1 次印刷

印数:0,001—3,250 册

ISBN 978-7-5327-4405-3/H·804

定价:19.00 元

如有质量问题,请与承印厂质量科联系。T:021-56401196

总 序

谈翻译,我们首先注意到这样一个现象:翻译,作为一种实践活动,在人类的文化交流中一直在广泛地进行着,但在历史中却很少得到世人的关注;丰富的翻译活动,一直被实践者认为是充满障碍的工作,但在相当长的历史时期内,却很少有学者对其进行深入而系统的研究。这两个方面情况的长期存在,或者说翻译领域被历史学界、被理论界长期忽视的状况造成的直接影响便是,翻译一直被当作一种“雕虫小技”。在很长一个历史时期内,语言学家对翻译问题不予重视,历史学家对翻译活动熟视无睹,知识界对翻译的认识几乎是零。这在很大程度上使翻译活动在两个方面被遮蔽:一是翻译在人类文明发展史上的作用得不到足够的承认;二是对翻译的种种问题得不到科学、系统的研究。尤为耐人寻味的是,对翻译的这种轻视态度不仅仅来自翻译界的外部,而且还来自翻译界的内部。翻译界内部的这种自我定位也在很大程度上影响了其他学科对翻译的看法,渐渐地也在有关的学界形成一种偏见:翻译为雕虫小技,不登大雅之堂。因此,翻译的作用得不到应有的承认,对翻译的理论研究得不到学界的关注和支持。这种状况一直延续到上个世纪五十年代,才逐渐开始改变。

长期以来翻译家本身专注于翻译实践,忽视了对翻译问题的理性思考

与系统研究,这不能不说是个历史的误会。但是,丰富的实践与贫乏的理论之间所产生的这道深深的裂痕却不能完全掩盖在数千年的翻译历史中。翻译活动的特殊性提出了种种问题和困难,而面对这些问题和困难,翻译家们也不可能视而不见,无动于衷,因为它们是无法回避的客观存在。从这个意义上说,翻译家们对翻译理论思考的忽视,并不说明翻译就没有理论研究的必要,也并不意味着翻译的种种问题因为翻译家的忽略便不存在。

事实上,在漫长的翻译历史中,翻译家们在不同的历史阶段进行着形式多样的丰富实践,虽然对实践中所出现的问题,还没有以一种清醒的理论意识去加以关注,但他们针对这些问题所采取的种种手法、策略,他们在实践中积累的一些经验、体会,尤其是他们从中悟出的一些道理或原则,是一笔笔非常珍贵的遗产。然而,令人遗憾的是,这些弥足珍贵的译事经验,由于学界长期以来没有予以重视与关注,像一颗颗散落的珍珠,被历史所尘封,无法发出其耀眼的光芒。

从上个世纪 50 年代开始,一批具有强烈的探索精神和清醒理论意识的学者,如前苏联的费道罗夫、英国的卡特福德、加拿大的维纳与达尔贝勒内、法国的乔治·穆南等,试图以语言学为指导,打开翻译研究的大门,将数千年来一直处在技艺层面的翻译经验纳入理论的思考、系统的分析与科学的探索范围。到了 70 年代,出现了以美国尤金·奈达为代表的一批翻译理论家,他们不断拓展翻译研究领域,将翻译研究提高到一个新的高度,为翻译学的建立打下了坚实的基础。

特别需要指出的是,上个世纪 70 年代以后,翻译界的一批有识之士将目光投向被历史尘封的一笔笔珍贵的翻译遗产。他们一方面追踪历史上的重大翻译事件,将翻译家的实践置于宽阔的历史空间加以认识与定位;另一方面对伴随着翻译活动而产生的点滴思考与经验体会加以梳理与探讨。以现代学理对传统的翻译经验加以阐释,为我们开辟了一条深

化翻译理论研究的重要途径。同时,客观上也促使被历史遗忘或忽视的翻译活动得到了越来越多学者的关注。如果说翻译在历史上曾长期受到冷落,处于历史边缘的话,那么正是借助现代学理,借助哲学、美学、语言学、符号学、文艺学等学科的最新研究成果,丰富并加深了人们对翻译的认识,不断拓展翻译研究的领域,使翻译活动在历史的舞台上逐渐占据本应属于自己的位置,从历史的边缘开始走向中心。

在把翻译从边缘推向中心的历史进程中,语言学家们充当了先锋的角色。如费道罗夫、卡特福德、维纳与达尔贝勒内、乔治·穆南等从语言学的角度,对翻译进行了系统而深入的研究,其贡献是多方面的。是翻译的语言学研究把翻译从经验主义中解放出来,为翻译研究打开了科学的大门,历史上第一次赋予了翻译研究科学的性质,将过去近两千年来一直在经验层面讨论或争论不休的问题置在科学的层面进行探讨。但翻译活动十分复杂,涉及的因素多,范围广,有关翻译的许多问题,如翻译者的主观因素、语言转换中的文化移植、影响翻译的社会、政治因素等问题,在语言学层面难以展开系统和深入的分析,得不出令人信服的答案,翻译历史上的众多现象也无法得到辩证的解释。翻译的语言学研究途径暴露出的这些局限,不仅使其他学科理论的介入显得非常必要,更为这些学科自身的发展提供了崭新的探索空间。

当回过头去,对自上个世纪 50 年代以来翻译研究所走过的路作以回顾性的审视时,我们不难看到以下几点:一是翻译研究在近半个世纪以来得到了突破性的发展,其广度和深度都是在过去任何一个时期所未能达到的;二是翻译研究的途径得到不断开拓,各种翻译研究流派纷呈,出现了一大批具有代表性的研究成果;三是翻译理论研究的不断深入与发展越来越受到其他学界的关注与承认。在国外,从上个世纪 80 年代末起,就有学者开始对 50 年代以后的翻译理论研究状况进行分析与总结,如加

拿大的罗贝尔·拉罗兹、英国的埃德温·根茨勒,他们分别于1989年与1993年发表了同名著作《当代翻译理论》。前者以翻译所涉及的基本问题为核心,对上个世纪50年代至80年代在翻译理论研究领域比较活跃的代表人物的观点与思想进行评述;后者则根据自第二次世界大战至20世纪90年代初西方翻译理论研究的发展状况,以不同的观点和重要著作作为依据,将当代的翻译理论分为“美国翻译培训派”、“翻译科学派”、“早期翻译研究派”、“多元体系派”和“解构主义派”等五大流派。香港的陈德鸿与张南峰编写的《西方翻译理论精选》收录了西方20位重要翻译理论家主要著作的部分章节的译文,这20位译学理论家中,除德莱顿、泰特勒、施莱尔马赫3位之外,其余17位均是当代的。根据编者的划分,西方译学研究界的这20位代表人物被列入6大学派:语文学派、诠释学派、语言学派、目的学派、文化学派、解构学派。除传统的语文学派,其余的5个流派都是近50年来发展起来的。在《西方翻译理论精选》的绪论中,两位编者这样说道:“西方的翻译理论,除了语言学派和传统的语文学派之外,还有近一二十年才兴起或盛行的翻译研究学派,以及解构主义、女性主义、后殖民主义等学派,可谓百花齐放。”如果再把我们的视野扩大一些,还可列举出符号学派、交际学派、语言哲学派、文艺学派等翻译流派。从历史上长期以来“不入流”的翻译经验之谈到20世纪末令人眼花缭乱的翻译流派的形成,我们可以看到,翻译的理论研究开始或已经进入了一个全面发展的时代。透过这些被冠以各种名称的翻译思想或观点,我们不难看到相同的一点,那就是借助其他学科的理论成果,对翻译进行研究。以语言学的理论指导产生的研究成果被统称为“语言学派”,以女性主义理论为参照的研究,被冠以“女性主义翻译流派”,总之,一种理论的介入,从积极的角度讲,都会给人们认识与研究翻译打开一条新的通道。

借助其他学科的研究成果,客观上确实为翻译研究拓展了巨大的空

间,为翻译研究注入了科学的活力,渐渐地从边缘开始走向中心。翻译,作为一种复杂的活动,涉及面广,若仅仅局限在一个领域对之进行研究,无法揭示其性质及活动规律。在这个意义上,翻译研究必定具有综合性。但是,当各种学科的理论介入翻译研究领域之后,当我们在为翻译研究由此进入全面发展而欣喜的同时,不能不看到在种种理论指导下取得的研究成果存在着一个致命的弱点,那就是如同“盲人摸象”,每一种理论流派所认识的翻译在很大程度上具有片面性,揭示的只是翻译活动的一个方面,难以深刻地反映翻译活动的全貌。此外,在理论的层面上,从目前翻译理论研究的现状看,还出现了“理论+翻译”的两张皮现象,有的理论只浮在表面,难以真正起到指导翻译研究的作用。最为值得注意的是,翻译研究在引进各种理论的同时,有一种被其吞食、并吞的趋向,翻译研究的领域看似不断扩大,但在翻译从边缘走向中心的路途中,却潜伏着又一步步失去自己位置的危险。面对这一危险,我们不能不清醒地保持独立的翻译学科意识,从翻译学建设的高度去系统地探索翻译理论问题,而在上海译文出版社支持下主编的这套《译学新论丛书》正是向这一方向努力的具体体现。

《译学新论丛书》有着明确的追求:一是入选的课题力求具有相当的理论深度和原创性,能为翻译学科的理论建设和发展起到推动作用;二是研究力求具有系统性,以强烈的问题意识、科学的研究方法、扎实的论证和翔实的资料保证研究的质量;三是研究力求开放性,其开放性要求研究者既要有宽阔的理论视野,又要把握国际翻译理论研究前沿的进展状况,特别要在研究中具有探索的精神,力求有所创新。但愿在翻译界同仁的支持下,在各位作者的努力下,我们的追求能一步步得以实现。

主编

2007年8月18日于南京大学

第一章	引论	/ 1
1.1	语料库与词典编纂	/ 3
1.2	汉英双语词典的研编现状	/ 7
1.3	汉英积极型双语词典研编的核心问题	/ 9
1.4	汉英词汇语义对比研究的新方法	/ 11
1.5	CpsDict 的研制目的	/ 13
1.6	本书结构	/ 17
第二章	语料库语言学的复苏和语料库词典学的兴起	/ 23
2.1	语料库语言学的复苏	/ 25
2.2	我国语料库建设的兴起与繁荣	/ 35
2.3	国内外平行语料库的建设及研究现状	/ 37
2.4	语料库词典学的兴起及特色	/ 41
第三章	积极型汉英词典研编的不足及原因	/ 45
3.1	我国汉英词典编纂的两个阶段	/ 47
3.2	现有汉英词典存在的主要问题	/ 49
3.3	汉英词典不足的主要原因	/ 54

目 录 Contents

第四章	词义研究的理论和方法	/ 57
4.1	意义理论和意义的分类	/ 59
4.2	词义研究的理论	/ 71
4.3	词义研究的方法	/ 101
4.4	积极型汉英词典词义研究的特点	/ 106
第五章	CpsDict 系统的研制	/ 111
5.1	CpsDict 系统的研制目的	/ 114
5.2	系统的结构组成	/ 115
第六章	CpsDict 系统在汉英积极型词典研编中的应用	/ 141
6.1	CpsDict 系统在汉英积极型词典研编主要环节上的应用	/ 143
6.2	CpsDict 系统在积极型词典词义研究中的运用	/ 158
第七章	关于 CpsDict 系统研编和运用的思考	/ 201
7.1	语料库本身的代表性	/ 203
7.2	配例的科学性	/ 208
7.3	关于高频词	/ 210

目 录

Contents

7.4	关于语料的对齐	/ 212
7.5	关于语料的深加工和检索	/ 214
7.6	结语	/ 218
主要参考文献		/ 220
主要参考词典		/ 242
附录一 示范性词条		/ 244
附录二 重要术语		/ 247
后记		/ 250

目
录

Contents

第一章

引 论

1. 1

语料库与词典编纂

语料库的英文形式 corpus 来自于拉丁语,在拉丁语中的意思是“身体”、“正文”、“文集”(body)等。在语言学领域,语料库具有特殊的含义,先看下面的几条关于语料库的定义:

定义一:语料库是一些书面语文本或口语听抄稿的集合,能为语言分析和描写提供帮助。(A corpus is a body of written text or transcribed speech which can serve as a basis of linguistic analysis and description.)(Kennedy, 2000:1)

定义二:语料库是对文本进行系统收集的产物,它记录了语言变体和用法方面的特征。(A corpus is a systematic collection of texts which documents the usage features of a language or language variety.)(Hartmann and James, 2000:30)

定义三:在当代语言学领域,语料库不能简单理解为任意话语的集合,它是依据科学的抽样原则而选择的有限话语集,语料库要以计算机可处理的形式储存,要具有代表性,而且要最大限度地反映研究对象的特点。(McEnery, Wilson, 2001:32)

以上 3 个定义代表了 3 类不同的语料库:任意文本的集合(定义一);经过选择的文本的集合(定义二);经过科学抽样、精心选择的可供机器处

理的文本的集合(定义三)。我们姑且把第一和第二类语料库称为广义上的语料库,而把第三类称为狭义上的语料库。随着研究的深入,人们认识到广义上的语料库有诸多的缺点,不能作为语言研究的基础。我们认为现代意义上的适合语言研究和词典编纂的语料库至少应该具备3个条件:(1)代表性。选材不具备代表性的语料库难以保证研究的科学性,代表性是语料库选材的最基本要求。(2)目的性。语料库不论建多大,与真实的语言世界相比都是沧海一粟,语料库只是话语全域的一个代表。Teubert(2001)说过话语全域包罗万象,语料库无法包含其全部内容,“普适”的语料库是不存在的。所以要根据研究的目的决定语料库的规模。(3)机器可读性。虽然语料库不一定都要经过详细的标注,但机器可读性是必备的条件。语料库文本只有在软件环境下才能进行检索、统计等自动化处理,一个不能被机器处理的语料库在应用价值方面难以得到保障。

语料库与词典编纂的结合有着悠久的历史,早期词典编纂过程中使用的“卡片”可视为现代“语料”的前身。从18世纪中期开始,词典编纂人员开始有意识地收集真实文本,并在此基础上确定词典的词目、进行释义、提取例证等。许多词典学家把他们的工作视为在词典编纂领域使用语料库的开始(Osselton, 1983; Bejoint, 2002:97)。虽然说此时的语料库还不是真正意义上的语料库,与现代的语料库相比还有许多不科学的地方,但词典编纂者和词典学家已经注意到语料库的作用和优点。

虽然在词典编纂领域使用“语料库”已行之有年,但在20世纪60年代以前,人们(包括词典编纂者)并没有从理论的高度来看待词典编纂,词典编纂与语言学研究长期脱离。Marckwardt(1963:344)批评说:“英语词典的编纂没有反映语言学的任何成果,在词典中,词类还是按传统的方式分为名词、形容词和动词等。词典编纂者也没有考虑是否依据形式或功能而给词典一个前后一致的结构。词典在释义的时候也看不到任何结构

主义的影子。”

此时,词典编纂和词典学也长期被主流语言学家所忽视。这主要有以下原因:

(1) 语言学家们认为词典只是一种商品,从词典产生之日起,它就没有过任何的变化。词典与语言学理论无关,也没有科学性,不值得对其进行研究。词典充其量只能算是与语言学相关的一个不纯净的副产品(an impure by-product of linguistics)(Rey, 1982:17)。

(2) 因为词典是关于词汇的书,是语言中某些词的汇集,在 19 世纪以及 20 世纪 70 年代之前,词汇一直被视为语言研究中无足轻重的成分。在当时,词汇学也没有被当作语言学的一个分支来看待。Bloomfield 非常轻视词汇的研究,他认为“词汇表记录的只是一些没有规律的东西,词汇只能是语法的附庸,而语法反映的才是语言形式有意义的组合。”(Bloomfield, 1933:274)除 Bloomfield 之外,20 世纪最具影响力的转换生成语法学派在早期也对词汇研究持否定态度,该学派的两个代表人物 Chomsky 和 Halle 是这样定义词汇和认定词汇功能的:词汇如同一个垃圾桶,在语言研究中不规则的东西、没有共性的东西就会被扔到其中(Chomsky, Halle, 1968:12)。

1961 年韦氏词典第三版的出版标志着词典编纂领域一个全新时代的到来。韦氏第三版吸收了结构主义语言学的营养,崇尚描写语言学而抛弃了规定主义的历史语言学。韦氏第三版的出版引起了语言学界的广泛争论,也引起了语言学家对词典学的注意。自 20 世纪 60 年代以来,词典学引起了越来越多的语言学家的重视,词典学家也意识到语言学理论对词典学的重要意义。

从 20 世纪 70 年代起,词汇研究逐渐受到了语言学家的关注。转换生成学派发现用转换规则无法解释一些结构之间的关系,如:They de-

stroyed Pompeii 和 their destruction of Pompeii。在这种情况下,Chomsky 试图运用一种叫“词汇假设”(Lexicalist Hypothesis)的理论对词汇(像这里的 destroy 和 destruction)进行赋值。他认为这些问题靠语法(grammar)是解决不了的,应该开展词汇的研究(Chomsky, 1970)。随 Chomsky 之后,Jackendoff(1975)研究了与词汇相关的“冗余理论”(redundancy rule),后来他进一步认为词汇是语音结构(phonological structure, PS)、句法结构(syntactic structure, SS)和概念结构(conceptual structure, CS)间联系的桥梁。Hudson(1991:3)在《英语词汇语法》(*English Word Grammar*)中列举了当今语言学发展的 8 个趋势,其中“词汇主义”(lexicalism)被摆在首位(Richard Hudson, 1991:1—14)。词汇主义认为语言研究应该从语法中心转变为词汇中心。此外,词汇研究也受到了其他许多学者的重视,如 Halliday (1985, 1991, 1992), Blass(1982), Meara(1984)和 Ringbom(1998)等。

国内外对词汇研究的重视还有一个因素就是出于实际的需要,在自然语言处理领域,词汇和词义选择是难点之一。林杏光(1999:63)认为为了适合机器翻译的需要,近几年来在计算语言学界也强调词汇主义。研究重点从句单位返回到词单位,词汇的研究是描写的重点,句法研究也要落实到词语上。他还认为语言研究要搞复杂特征集,这是发展潮流,而复杂特征集的重点就是词的复杂特征集。

词汇研究与机器词典关系密切,在机器翻译方面,机器词典依赖于人们对词汇的形式化描述,机器词典的质量直接影响机器翻译的质量。在普通词典领域,以词汇为中心的各项研究,如义项和词汇语义的研究也应该成为重点。语料库方法和语料库语言学从产生之日起就和自然语言处理有着千丝万缕的联系,基于语料库的各项研究正在自然语言处理领域如火如荼地进行。在普通词典学领域,语料库也正成为国内外词典学界的一个新兴热点。今天,词典学与语言学已经相互融合、相互影响(Be-

joint, 2002:177)。词典学家和编纂者已无法继续忽视语言学领域的新方法和新理论。语料库语言学和语料库研究方法正向词典研究的各领域渗透。其实,词典与语料库的结合早在上个世纪 40 年代就已经开始。当时的重点是编写机器词典,为机器翻译服务。在普通的语文详解词典编写过程中利用电子语料库开始于 20 世纪 80 年代,由于语料库的分析研究方法有别于传统的内省的方法,基于大规模真实文本的语料库能在词典立目、义项确定、释义、例证提取、辅助释义等多方面给词典编纂带来便利,所以一门新兴的学科——语料库词典学——便应运而生。

1.2

汉英双语词典的研编现状

在国外,语料库与词典编纂的结合已取得了丰硕的成果,“牛津系列”词典和“朗文系列”词典的编写都借助了语料库,剑桥大学出版社也借助语料库编纂出版了《剑桥国际英语词典》。将语料库运用于词典编纂领域最负盛名的是英国的伯明翰大学,著名的 COBUILD 系列词典的成功标志着语料库词典学的日臻成熟。

与英语词典的研编相比,汉英双语词典的研编在理论研究和编纂的技术手段上都较为落后。自 1978 年以来,我国大陆地区出版了多本“汉英词典”,据统计,“迄今为止国内已经面世的汉英词典的绝对数目已经超过 200 部”(徐海江,2005)。彭宣维将我国 20 世纪的英语类双语词典(包