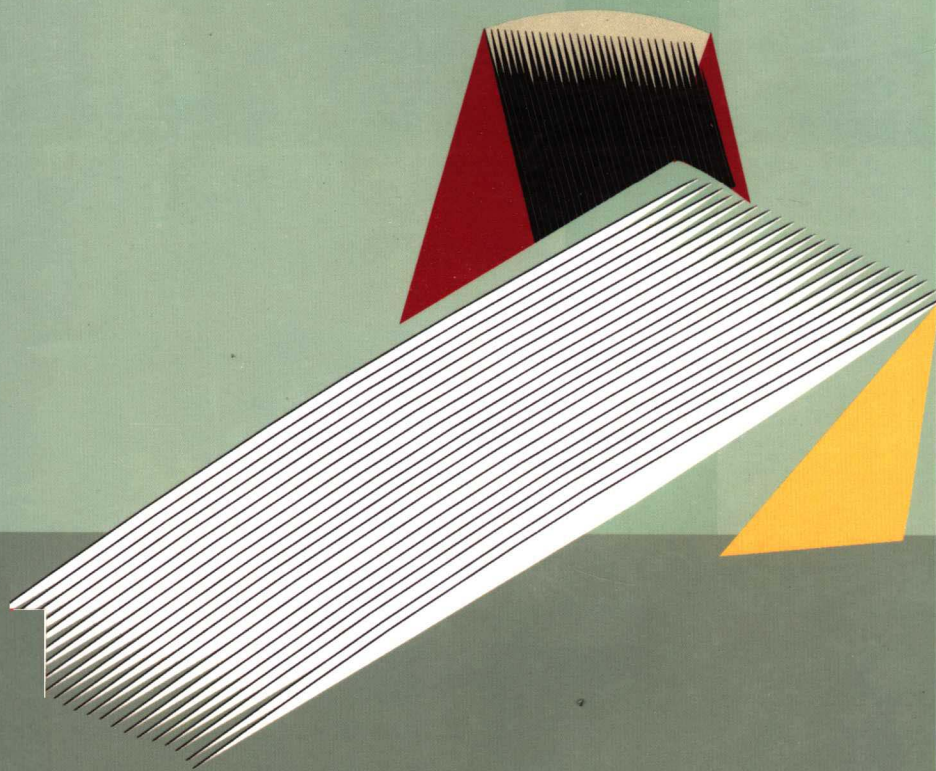


教育部人文社会科学重点研究基地
北京外国语大学中国外语教育研究中心·学术著作系列

双语对应语料库 研制与应用

王克非等 著



外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

教育部人文社会科学重点研究基地
北京

心·学术著作系列

H085.3
W241



郑州大学 *04010164057S*

双语对应语料库 研制与应用

王克非等 著

-60
4本



H085.3
W241

2005.20/24

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS
北京 BEIJING

图书在版编目(CIP)数据

双语对应语料库: 研制与应用/王克非等著 —北京: 外语教学与研究出版社, 2004 6

ISBN 7-5600-4236-8

I 双… II 王… III 一对一翻译(机器翻译)-数据库 IV ①H085.3 ②IP391.2

中国版本图书馆 CIP 数据核字(2004)第 065868 号

双语对应语料库 研制与应用

王克非等 著

x * *

责任编辑 赵岩华

出版发行 外语教学与研究出版社

社 址 北京市西三环北路 19 号 (100089)

网 址 [http //www fltrp com](http://www.fltrp.com)

印 刷 北京外国语大学印刷厂

开 本 850×1168 1/32

印 张 8.75

版 次 2004 年 9 月第 1 版 2004 年 9 月第 1 次印刷

书 号 ISBN 7-5600-4236-8/H·2044

定 价 11.90 元

x * *

如有印刷、装订质量问题出版社负责调换

制售盗版必究 举报查实奖励

版权保护办公室举报电话: (010)88817519

总 序

中国外语教育研究中心决定将其专/兼职研究员近几年的个人科研成果及课题成果组成学术著作系列，由外语教学与研究出版社（外研社）出版。外研社社长李朋义同志请我讲讲这套系列学术专著的来龙去脉及其性质、特点，为丛书做个总序。

我先简要介绍一下中国外语教育研究中心的基本情况。北京外国语大学中国外语教育研究中心成立于2000年3月。学校投入百万余元装修了近300平方米的办公用房和图书资料室，累计购买了数十台计算机和其他各种科研设备。当年7月，中心通过了教育部专家组的检查验收，9月教育部正式下文，批准中心为全国高校人文社会科学100个重点研究基地之一。

中心的研究人员构成主要是北外外国语言研究所、英语学院、高级翻译学院和应用英语学院的多名教授、副教授，中心同时还聘请了本校日本语研究中心和俄语学院的两位教授，北京大学、社科院语言所、解放军外国语学院、华南理工大学、南京大学等单位的几位教授为兼职研究员。大部分研究员具有博士学位，其中几位是香港中文大学、剑桥大学或其他国外知名大学毕业的博士。研究人员实行流动管理，带科研项目进入中心，完成项目后可以离开，有新的项目可以延聘。专职和兼职研究员各保持在8-10人左右。

该中心的总体任务是，除语言基础研究外，还系统研究中国外语教育中的各种问题：从外语政策、规划到发展策略，从外语教育理论到教育实践，从教材编写到课程设置，从课堂教学到测试评估，从中小学外语教学到高校的专业外语和公共外语教学。中心的发展目标是建成外语教育的科研攻关基地，外语研究人才的培养基地（硕士生、博士生和博士后），资料图书和信息交流基地，为有关部门提供相关咨询的基地。中心力图成为全国第一的外语教育问题研究机构。

目前，中心承担一项国家社会科学基金项目 and 六项教育部的重大项目，与加拿大合作开展一项国际研究，其中包括中国高校外语教学改革现状与发展策略研究、英汉平行语料库的创建与应用研究、英语能力测试的理论研究和试题开发、高校英语教师教育与发展研究、二语习得理论研究、网络技术在外语教育中的应用、英语学习词典的研编等。这些项目，有的已经发表；

有的已经完成，成果不久即可问世；有的将写出咨询报告。目前中心招收硕士研究生 40 余名，博士生 20 余名，博士后 1 名。这些学生在校期间必须参加中心的部分研究工作，必须有论文发表；博士生和博士后的研究成果必须达到出版水平。

从建立之日起，中心就与外研社有着密切的合作关系。外研社给予中心大力支持，中心通过外研社的研究发展中心为社里提供咨询服务、项目策划、作者联系、稿件终审等。三年来，中心协助外研社举办了多期大学英语教师的暑期培训班（共约 3 万人），中心研究员每期都提供几个学术报告，同时也对教师进行问卷调查和小组访谈，双方都收到良好效果。更值得一提的是，中心与外研社合作，设立“中国英语教育研究基金”，外研社提供 50 万元科研经费，每年都组织课题的招标活动，课题内容与中心重大项目相联系，中心负责审查批准立项的项目，培训主持人，最后对报告进行终审等。项目的研究报告将与中心的主题报告同集出版。同时，中心与外研社达成协议，中心的一切科研成果，由外研社独家出版。

这套学术著作系列包括四种科研成果形式。首先是中心承担的重大科研项目的成果，或者以专著形式，或者以系列论文形式，总之是围绕一个课题而展开的。这种研究的优点是：全部为实证研究，是有组织、有计划、规模较大的调查研究；全部是有学术意义的课题，有极大的探索性；这类成果也包括给教育主管部门的咨询报告，为决策者出谋划策。其次就是专/兼职研究员们自选课题的科研成果。这些项目虽然没有在科研主管部门立项，但都是研究员们最感兴趣的课题，有的是多年积累沉淀的最后结晶，有的是把当年的博士论文继续修改、完善，使之与当今研究接轨的成果。第三类是我们自己培养的博士生和博士后的成果，经过修改之后，通过评审，达到出版水平，也被纳入这个系列。第四类是学术会议论文集。中心每年举行一次学术研讨会，通过评审的优秀论文由中心编辑加工并加上适当的评介后结集出版。因此，这套系列专著的内容将十分广阔丰富，既有理论性很强的成果，也有突出实用性的成果，形式多样。我们的学术著作丛书是个开放的系列，首批已经成形的有 9 种，但随着中心研究工作的深入和发展，将会有更多有价值的专著问世。

我非常高兴地向读者推荐这套学术著作系列，我相信它对语言研究者、外语教育研究者、外语教师、语言学硕士生和博士生及对我国外语教学有兴趣者，都有一定的参考价值。

我谨代表中国外语教育研究中心向外研社表示诚挚的谢意。

刘润清

2003 年 3 月于北京

前 言

本书是教育部人文社科重点研究基地北京外国语大学中国外语教育研究中心的“双语平行语料库的创建及其应用研究”课题研究成果。

语料库及语料库语言学近二三十年在国际上发展迅速，其主要目标是以计算机储存大量的真实语料，对语料做各种带有研究目的的加工标注，利用研制的检索工具对标注语料进行快捷的搜寻和分类，以发现并分析以往因条件所限而未能注意的语言现象。

三年多以前，我们申请教育部重点研究基地时，认识到外语教育研究应当既有通盘的规划、细致的调研，又有长远的基础项目。这个基础须同外语教学与研究密切相关和互补，而创建大型双语平行对应语料库，以它收集、积累并赋以适当标注的大量语言素材，使我们能更好地观察和分析语言、翻译现象，探讨外语教学诸多方面的问题，正是这样一个合适的基础研究项目。

于是在接下来的三年里，为了这一既定目标，我们在对国内外语料库研制与应用现状做了全面调研的基础上，精心设计，精细选材，经过长期艰苦的工作，前后处理了上亿词次的语料，终于初步建成目前世界上最大的双语平行对应语料库（含汉英、汉日两个双语对应语料库），也是国际上涉及汉语的最大的双语语料库。

两个语料库（汉英语料库由中国外语教育中心研制，汉日语料库由北京日本学研究中心研制）分别在资金上获得教育部重大科研项目和国家社科基金项目的支持，在技术上得到北京大学计算语言学研究所和日本日立中央研究所、奈良先端科学技术大学院大学等机构的支持。研制过程中，课题组研究和解决了一系列对应语料库难题：1）设计方案和规模上兼顾语言与翻译研究，2）语料取样的均衡性和代表性，3）汉英语句自动对齐（对齐率80—90%左右），4）对齐语句的自动链接式检索，5）基本的语法标注和词、词频、短语、句型、搭配等的检索，6）所有语料均给予文本类型、分类、文体、语体、时代、作者/译者、字数等参数标记，7）应用双语库开展相关的语言、翻译、教学及词典编纂等研究。

我们的双语语料库的特点在于：1）在设计上将全库分为若干子库，如百科语料库、专科语料库、翻译文本库、对应语句库等。2）整个语料库可

分可合：分可开展单项研究，如百科语料库、对应语句库利于双语词典研编，专科语料库利于自动翻译研究，翻译文本库利于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。3) 此语料库区分4种语料：外语（英/日语）原文、汉语原文、外语（英/日语）译文、汉语译文，可分别进行单语研究、双语对比研究、原文与译文语言对比研究。4) 利用此双语语料库可搜寻大量的对应词语、短语，丰富英/日-汉、汉-英/日双语词典的编纂。

几年中，研制人员尽管都是教学科研任务缠身，但仍然对语料库工作非常投入，而且边研制，边使用，边探讨，因此才能在建成初步的大型语料库的同时，提交自己的研究成果。关于汉日对译语料库，去年9月已出版《中日对译语料库的研制与应用研究论文集》；在汉英对应语料库建成的今天，我们又将新的探索与发现呈现给大家。全书分四部分、24节，分别讨论语料库的研制、检索问题，以及应用双语库开展的相关语言、翻译等研究，由研制人员分工撰写，总撰稿人修订、统合。我们希望这本书对于想了解双语语料库的读者会有所帮助，也希望有读者因此而有兴趣进入这一领域，成为我们的同道。书中多有不成熟甚至幼稚之处，也许这正代表了我们的双语语料库研究工作的起步，是一个阶段的见证和答卷，仅供学界参考，并恳请同人批评指正。

王克非

2003年12月20日

另：双语的对应语料库英文为 parallel corpus，又称平行语料库，意为库中的两种语言呈平行对应或对译关系。国内使用较多的术语是平行语料库，直译自英文术语，似不大能反映双语一一对应的关系，故部分学者使用“对应”、“对照”或“对译”。本书为求一致，就全部改为“对应”，敬请使用“平行”等其他术语的作者和读者体谅之。

目 录

前言	1
第一部分 双语对应语料库及其研制	
第一节 语料库和双语语料库	3
第二节 国内外双语对应语料库研制现状	17
第三节 对应语料库的对齐、标注与应用	21
第四节 汉英对应语料库及其多重研究用途	28
第五节 汉英对应语料库的设计	36
第六节 汉日对应语料库的设计	54
第二部分 双语对应语料库检索研究	
第七节 双语库检索平台研究	61
第八节 汉英近义句翻译检索	69
第九节 英汉对应词的自动提取	80
第十节 日汉对译词的自动提取	89
第三部分 基于对应语料库的语言研究	
第十一节 英汉语句对应考察	99
第十二节 英汉语中“体”的分布考察	108
第十三节 英汉语致使结构对比	119
第十四节 汉语把字句特点、分布及英译问题	128
第十五节 英汉语被动语句的特点与机器翻译处理	140
第十六节 英汉语动结结构对比与语料考察	151
第十七节 日汉对比的语料考察与研究模式	166
第四部分 基于对应语料库的翻译研究	
第十八节 翻译及其语言特点	181
第十九节 翻译语言个案考察与分析	187
第二十节 异化与归化现象个案语料考察	200
第二十一节 语料与翻译教学	207

第二十二节 语料与机助翻译	215
第二十三节 一项翻译实验分析和语料库分析	223
第二十四节 对译语料与双语词典的研编	232
术语	237
相关语料库及软件网站	246
参考书目	250
本书撰稿人	269

第一部分

双语对应语料库及其研制

第一节 语料库和双语语料库

本节从语料与语言研究的关系,谈及语料特别是双语语料库的研发背景和历史,以及由此而产生的语料库语言学。文中详细描述了对应语料库、类比语料库和翻译语料库等双语/多语语料库类型及其彼此之间的区分,探讨了对应语料库的加工及检索方法。文中重点描述了对应语料库的研制与应用,特别是它在翻译理论与实践、对比语言学上以及翻译教学上的应用。本节也分析了这种研究的优势与局限。

1. 语料与语言研究

语言研究,无论古今中外,从来都重视语言材料的充分收集、整理和分析。在20世纪前的语言研究中,重实证的历史研究法占主导地位,虽然也出现了《保尔·罗瓦雅尔语法》(1660,普遍唯理语法),但根本性的变化是在本世纪,即索绪尔《普通语言学教程》中提出“历时”、“共时”概念之后,出现了重语言史的历时研究和重语言态的共时研究并重甚至更重后者的局面。但不论历时研究或共时研究,语言材料都是至关重要的。以英语语法研究来说,20世纪著名的叶斯泊森《现代英语语法》(7大卷,1909—1949)及夸克等人合编的《当代英语语法》(1972)和《英语语法大全》(1985)分别代表了广收例证的历时研究成果和共时研究成果之最。叶氏成书时代尚无计算机的帮助,他辛勤选取自乔叟以后历代文学作品1200余种,引录多达12万条,并将作者、书名、页码等一一注明。叶氏的工作为语言变迁等历时研究提供了丰富的语料。夸克等人同样认识到,大规模语料对于准确了解和描述语法非常重要,他们于50年代末在伦敦大学率先建立了现代意义上的语料库,即“英语用法调查”(Survey of English Usage),广泛收集当代的而不是历代的、社会的而不只是名家的、通用的(包括口头的)而不只是经典的英语语篇材料,并在此百万英语词(以后逐渐扩大)的语料库基础上编出上述语法巨著,对当代使用中的英语作出了令人信服的描述和解释。此后由于计算机技术的飞速发展,语料库的规模更大,范围更广,功能更强,如著名的LOB(兰卡斯特-奥斯陆-卑尔根)语料库、伦敦-隆德语料库、朗文-

兰卡斯特语料库和英国国家语料库 (BNC) 等收词量上亿的超大型语料库 (Aijmer & Altenberg, 1991; Leech, 1991; Rundell, 1992; Biber et. al, 1998; 谢应光, 1996; 王建新, 1996; 顾曰国, 1998), 近年在大型语料库基础上又编出了《英语口语和书面语语法》(Biber et. al., 1999) 这样大部头的语法书。英语词典的编纂同样是在大量语料的基础上进行的。最著名的《牛津英语词典》即是从 19 世纪中叶开始, 积 500 余学者 71 年采集的几百万条引语, 始于 1928 年编成 (参看许国璋, 1991: 343-345)。1995 年牛津、剑桥、柯林斯、朗文 4 家出版社不约而同推出了自己品牌型的学习词典, 它们无一不是以语料库为基础编纂而成。可以说, 现在语言研究、辞书编纂甚至教材、试题的编写或研制都已经越来越多地依靠大型语料库了。

2. 语料库语言学的发展

语料库的兴建, 带动了基于语料库的研究, 语料库语言学这门新的分支学科发展起来, 成为近几十年里新的语言研究热点。语料库语言学的主要目标是以计算机储存大量的真实语料, 对语料做各种带有研究目的的加工标注, 利用研制的检索工具对标注语料进行快捷的搜寻和分类, 以发现并分析以往因条件所限而未能注意的语言现象。

关于语料库语言学的特点, Halliday (1993: 24) 曾指出: “语料库语言学将数据收集与理论论述有机地结合在一起, 使我们对语言的理解发生了质的变化。” Tognini-Bonell (2001: 3-4) 认为, 尽管语料库语言学属于应用语言学领域, 但与该领域其他学科不同, 它被视为一种应用前的研究方法 (a pre-application methodology)。它首先确立自己的一套规则系统和知识系统, 然后将它们运用于语言描述之中。因此, 语料库语言学有其应有的理论地位。在应用方面, 它特别有助于词典研编、翻译、语言教学、文体、语言对比以及计算语言学等研究领域。Hunston (2002: 1) 指出: “可以毫不夸张地说, 在过去几十年里, 语料库以及对语料库的研究对语言研究以及语言应用研究进行了一场革新。”

基于语料库的研究主要通过以下三种途径展开:

- 1) 运用实证方法对语言使用进行描述;
- 2) 从语境和功能的角度构建意义理论框架;
- 3) 使用新技术。

这种研究具有三大优点:

- 1) 运用计算机的强大功能, 进行快速、准确和复杂的分析;

2) 语料库规模大、语域宽、语言信息范围广;

3) 既有定量分析, 又有定性解释功能, 对语言的描写比较全面。

作为一个新兴研究领域, 一种新的探寻语言本质的哲学方法, 80 年代以来, 语料库语言学发展迅速, 对语料库的规模、类型、功能等也提出了更高的要求。在当今世界, 英语语料库的研制仍占主导地位, 但其他语言的语料库也陆续问世, 如法语、德语、汉语和日语, 扩展了基于语料库的语言研究范围。随着语料库语言学的发展, 词汇学、语言教学、翻译、文体学、语法学、计算语言学等研究领域将从中受益。

自 20 世纪下半叶英美开发出第一代语料库以来, 世界上许多国家和地区也陆续建立起各种单语语料库 (包括书面语和口语, 共时语料和历时语料等), 但双语语料库, 包括对应语料库或平行语料库 (parallel corpora) 和翻译语料库 (translational corpus) 的研制, 则直到 90 年代初、中期才开始 [第一代双语语料库有英语-挪威语双语语料库和 1992 年建立的英语-意大利语双语语料库, 参看 (Kenny, 1998), 又见本书第三节]。两种 (或两种以上) 语言平行对应在技术上要求很高; 此外, 人们通常认为, 翻译的语言算不上真正的或地道的母语, 而且翻译有高下之别, 有直译意译之别, 等等。诸如此类的原因使得双语对应语料库的开发相对滞后。但是正如语言学家认识到大型语料库对于语言研究的重大意义那样, 从事语言对比研究和翻译研究的学者也逐渐认识到双语对应语料库在语言对比研究、翻译教学与研究 (包括自动翻译)、双语词典编纂以及外语教学等方面的潜在价值。英国曼彻斯特大学科技学院翻译研究中心因此在 1995 年率先建立起世界上第一个翻译语料库 (Translational English Corpus, TEC)。

可见, 继单语语料库之后, 双语语料库 (如类比的和对应的语料库等) 的研制业已成为语料库语言学研究的一个重点, 因为后者为翻译和语言对比研究提供了重要资源。正如 Aijmer & Altenberg (1991: 12) 所述, 类比及对应语料库对翻译和语言对比研究特别有用, 而且扩展了研究的范围:

- 1) 从语言对比中深入了解所对比的语言, 而这种了解往往容易在基于单语语料库的研究中被忽略;
- 2) 通过一系列比较来揭示语言的共性以及某语种所特有的、语言类型和文化上的差异;
- 3) 揭示原文与译文、母语与非母语之间的差异;
- 4) 应用范围广, 如词典编纂、外语教学和翻译。

在多语语言资源 (multilingual resources) 方面, 翻译研究者和对比语言学家可以根据他们是关心翻译过程还是翻译产品而使用不同类型的语料。作跨语言研究时, 如能同时使用或参照对应语料库和类比语料库, 则可能有更

多的发现。

简言之，语料库为语言学提供了丰富的研究资料，同时也为语料库语言学提供了工具和方法论。

下面我们着重讨论双语语料库问题。

3. 双语/多语语料库的类型、定义和功用

属于双语或多语语料库范畴的主要有对应语料库、类比语料库和翻译语料库。关于这三种双语/多语语料库的定义，不同的研究者持有不同的观点。为便于研究，我们在此作一定义，并在本书中统一使用。早于双语库的单语语料库也一并作一概述。

◆**单语语料库** (monolingual corpus)，是开发和使用最多的语料库。顾名思义，它仅收集一种语言的语料；又可再分为单原文库 (single monolingual corpus)，即只收录一种语言的原创文本的语料库 (如 BNC 英国国家语料库、CBECobuild 英语库、BROWN 布朗语料库等)，和单译文库 (single translational corpus)，即只收录一种语言的译文语料库 (如 TEC 英语翻译语料库，它包含从多种语言译成英语的文本)。一些学者认为 (参看 Kenny, 2001: 59)，单原文库还可作为描写翻译研究的“控制”对比参数。

◆**双语/多语语料库** (bilingual/multilingual corpora)，指两种或两种以上语言的文本构成的语料库。它可以有对应的、类比的和翻译的三种形式。

◆**平行/对应语料库** (parallel corpora)，由原文文本及其平行对应的译语文本构成的双语语料库，其双语对应程度可有词级、句级和段级几种。对应语料库可以是双语的或多语的，其主要形式有三种：

- 1) 单向对应语料库 (uni-directional parallel corpora)，指由一种语言的文本和将其翻译成另一种语言的文本构成的对应语料库。如，整个语料库所有文本都是由英语原文译成汉语译文的，或者都是由汉语原文译成英语译文的。
- 2) 双向对应语料库 (bi-directional parallel corpora)，指该语料库中既包括由 A 语言的文本及其被翻译成 B 语言的文本，又包括由 B 语言的文本及其被翻译成 A 语言的文本构成的对应语料库。如，整个语料库所有文本既有由英语原文译成汉语译文的，又有由汉语原文译成英语译文的。英语-挪威语对应语料库 (ENPC) 就属于这一种。
- 3) 多向对应语料库 (multi-directional parallel corpora)，指该语料库由同一种语言的原文文本及其被翻译成两种或两种以上语言的译文文本

构成的对应语料库。如, 整个语料库所有文本都是由英语原文文本及其对应的汉语、法语和德语等译语文本构成的。更复杂的多向对应语料库还包括反方向的、即两种或多种语言译成某一种语言的文本。

最初的对应语料, 如《加拿大议会会议记录》(Proceeding of the Canadian Parliament) 这样的双语准法律记录就相当于一种双语对应语料库。对齐这类语料的软件纯粹依赖两个语篇之间密切的句句对应关系。从这个意义上说, 同时以不同语言撰写的文本(如欧盟和联合国文件)也属对应库之列(Hunston, 2002: 15)。对应库中的对译词句对自动翻译的助益是显而易见的。它还有利于翻译的研究, 如 Øverås (1998) 以 ENPC (英语-挪威语对应语料库) 研究翻译中的“外显化”(explicitness) 现象。

◆**类比语料库** (comparable corpora), 由不同语言的文本或同一种语言不同变体的文本所构成的两个或两个以上的语料库。如国际英语语料库 (International Corpus of English, IEC) 就包含英语各种变体的、规模约 1,000,000 词的类比语料库。它可以是以子库采用相同抽样框方式建立起来的多语种语料库, 但各子库并非互为翻译, 其类比性在于相似的平衡性和代表性。

类比语料库也可再细分出单语类比库和双语/多语类比库。前者收集类似环境下的文本, 如类似的报刊文章等。Laviosa 构建的 ECC (English Comparable Corpus, TEC 的姊妹项目) 即属此种, 其中收集两种文本: 新闻文本 (报纸) 和叙事文本 (传记和小说)。瑞典哥德堡大学的类比语料库收集了 75 个小说文本, 一半是瑞典文的, 另一半是从英文译入瑞典文的。芬兰也在建设这种语料库, 收录芬兰原文撰写的文本和从英、德、俄文译成芬兰文的文本。这种单语类库便于对照本族语文本来观察本族语译文的特点 (如观察其型-例比、词汇密度、平均句长等)。而后者收集的是在内容、语域、交际环境等方面相近的不同语言的文本, 多用于对比语言学 (如 Aijmer etc., 1991), 或对比词汇学 (Teubert, 1996), 不大用于翻译描述 (参看 Zanettin, 1998)。

◆**翻译语料库** (translational corpora), 由彼此具有翻译关系的不同语言文本构成的语料库, 如上文提到的 TEC (Translational English Corpus)。翻译语料库中的双语语料不一定具有句级或某级别上的对应关系, 它可能只是篇章上的对应收录, 这是它与对应库不同之处。利用翻译库做研究并取得成果的, 以曼彻斯特大学科技学院翻译研究中心最为引人注目。

对应语料库、类比语料库和翻译语料库三者之间的异同可根据文本间是否具有翻译关系以及文本间是否可以对齐来判断, 其异同可用下面的表 1 来

表示：

语料库类型	有无翻译关系	文本间是否对齐
对应语料库	有	是
翻译语料库	有	否
类比语料库	无	否

表 1 各双语语料库之异同 (参看 Tognini-Bonell, 2001: 6-7)

Granger (2003: 21) 以更细的划分对双语库加以描述, 他的图解 (图 1) 是:

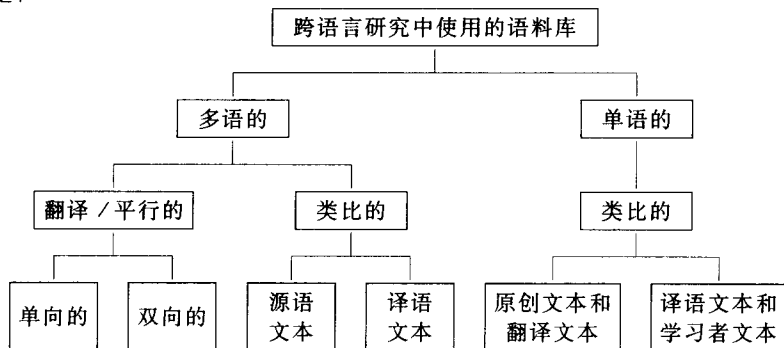


图 1 双语语料库类型与关系

由图 1 可见, 研究上的需要决定着语料库的性质, 即后者随研究类型的变化而变化。我们可以根据研究上的需要来采用或创建相应的语料库:

●由源语文本及其译语文本构成的双/多语对应语料库 (供对比和翻译研究)

●由源语文本构成的多语语料库, 各源语文本在语体、写作年代等方面具有可比性 (供对比研究)

●单语语料库, 由同一语言的原创文本和翻译文本构成 (供翻译研究)

倘若这三种类型语料库结合在同一个整体框架内, 如英语-挪威语对应语料库 (Johansson, 1998; 见图 2), 每一类型都可配合或补充另一类型。这样我们便可以在一个可称为通用型的语料库里开展语言对比和翻译研究。

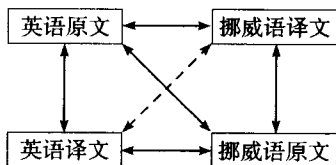


图 2 英语-挪威语对应语料库结构 (参看 Kenny, 2001: 39)