



管新潮 陶友兰 著

# 语料库与翻译

Corpus  
and Translation

 复旦大学出版社

管新潮 陶友兰 著

# 语料库与翻译

Corpus  
and Translation

 復旦大學 出版社

图书在版编目(CIP)数据

语料库与翻译/管新潮,陶友兰著. —上海:复旦大学出版社,2017.3  
ISBN 978-7-309-12882-6

I. 语… II. ①管…②陶… III. 语料库-翻译学-研究 IV. H059

中国版本图书馆CIP数据核字(2017)第043641号

语料库与翻译

管新潮 陶友兰 著

责任编辑/施胜今

复旦大学出版社有限公司出版发行

上海市国权路579号 邮编:200433

网址:fupnet@fudanpress.com http://www.fudanpress.com

门市零售:86-21-65642857 团体订购:86-21-65118853

外埠邮购:86-21-65109143

上海浦东东北联印刷厂

开本 787×960 1/16 印张 13 字数 189千

2017年3月第1版第1次印刷

印数 1—5 100

ISBN 978-7-309-12882-6/H·2700

定价:32.00元

---

如有印装质量问题,请向复旦大学出版社有限公司发行部调换。

版权所有 侵权必究

## 内 容 提 要

本书秉持“以问题为导向、以决策为引导、以客户为终端”的写作原则,介绍了语料库的起源和不同类型及其创建方法和应用工具,指出语料库创建和应用的关键在于其质量,而质量则表现为多维性,因用途而异。不同的语料库可运用不同的方法进行分析,以获得不同的应用结果。语料库在计算机辅助翻译实践中则转化为记忆库,以此助力于翻译实践的专业化并提升效率。与翻译相关的语料库学术研究主要表现在翻译文体、翻译共性、翻译规范、词语搭配等方面,而翻译实践同样也需要进行相关的研究。语料库在翻译教学中的应用已拓展了相关的教学途径和方法。伴随着语料库而出现的“语料库语言学”和“语料库翻译学”为语料库的发展提供了理论指导。

本书旨在帮助读者了解并习得语料库的相关知识、方法、技术和工具以及语料库与翻译的关系,既可作为感兴趣者的参考用书,亦可作为教材使用。

## 序 | PREFACE

自 Mona Baker 于 1996 年在其“Corpus-based Translation Studies: The Challenges that Lie Ahead”论文中提出语料库翻译学以来,这一学科的发展异常迅猛。尤其在我国的语料库翻译学的研究成果已层出不穷,令人目不暇接。这与当今大数据时代的语料库所能发挥的作用息息相关。语料库无论其为何种形式,如双语平行语料库、单语语料库、双语可比语料库等,均为翻译学研究、翻译教学和翻译实践等搭建起了极好的学术研究、教学与实践应用平台。语料库的学术研究价值已经得到国家社科项目和教育部社科项目的立项验证,其教学应用价值已有诸多教学机构为之提供极富成效的案例,而语料库在翻译实践中的应用价值已在诸如华为等公司的“走出去”实践中得到证明。

正如我在《语料库翻译学概论》一书中所言,语料库在翻译实践中的应用主要有四大方面:①利用平行语料库和单语语料库,译者可在分析众多翻译实例的基础上,确定源语词汇的目的语对应词;②平行语料库可用于分析不同语言语句之间的对应关系及其转换规律,帮助译者提高翻译效率和翻译质量;③借助于平行语料库,可以分析一些翻译名家或翻译名著所采用的翻译策略和方法,这对于译者的翻译实践能够发挥示范作用;④可比语料库的应用可以帮助译者判断译文是否地道、自然。除了语料库的学术研究和教学研究外,语料库的翻译实践应用也需要有学术研究,尤其是在企业级的翻译实践之中。

企业翻译不同于常规翻译,在翻译风格、翻译规划、赞助商、翻译目的、译者(员)教育以及译者主体性等方面具有自身独特的属性,应纳入翻译学研究。如若忽略该领域研究,翻译研究的理论框架往往不完善。任何企业但凡走出国门,都需要凭借翻译开道。没有翻译,任何企业均难以拓展国外市场。因而,研究企业翻译,尤其是大型跨国公司的翻译现象与翻译文本,能够丰富翻译学研究的内涵,完善翻译研究的理论构建,对于推动中国企业“走出去”具有重要的借鉴意义。

《语料库与翻译》一书是对翻译实践进行的一次有益而有效的研究尝试,其最大特点是基于著者 20 多年来在翻译实践中所积累的大量实践经验以及对此进行具有针对性的理论总结和归纳。我相信,这本书会对 MTI 和 BTI 翻译实践教学以及与企业翻译相关的诸多方面产生积极的影响,同时也会对语料库翻译学的翻译实践研究做出积极的贡献。在我看来,这本书具有三大特点:一是前瞻性,提出了诸多有待学界研究的问题并对未来进行了展望;二是应用性,对翻译实践具有直接的帮助作用;三是实用性,书中呈现的许多“小技巧”是一种有效翻译的技巧。《语料库与翻译》的主要作者管新潮在上海交通大学出版社从业时,我已与其相识。他对学术的独到理解令人难忘,他对“语料库翻译学”的追求令人感动。他曾专门策划并编辑了“语料库翻译学文库”,至今其中已有四种图书荣获教育部或省级人文社科奖。在调至上海交通大学外国语学院工作后,他为 MTI 学生开设了“语料库与翻译”、“翻译管理”、“翻译项目基础与实践”等课程,颇受学生的欢迎。课堂上的互动性使他的翻译实践经验得到了充分发挥。在学术研究上,他喜获教育部人文社科规划基金项目“基于语料库的马克思《资本论》汉译研究”等,可谓实践经验与理论研究的有效结合在他身上得到了真实体现。《语料库与翻译》一书的写作,现有复旦大学陶友兰老师的加盟,真是锦上添花,相信会让更多的师生受益匪浅。

今受管新潮老师之邀,乐为之序。在此,真心祝愿他在今后的学术道路上顺顺利利、成果多多。

胡开宝

2016年10月6日于上海

## 前言 | FOREWORD

语料库翻译学是指以语料库为基础,以真实的双语语料或翻译语料为研究对象,以数据统计和理论分析为研究方法,依据语言学、文学和文化理论及翻译学理论,系统分析翻译本质、翻译过程和翻译现象等内容(胡开宝 2011: 1)。由此定义,我们可以知晓语料库翻译学不仅仅是一门可以应用于学术研究的学科,也是符合翻译实践研究之需要的学科。语料库翻译学的研究内容如翻译共性、译者风格、翻译规范等,正是如何高效地实施翻译实践活动所需要进行研究的内容,其完全可以成为指导翻译实践活动的一门学科理论。有关语料库翻译学的学术研究,请参加胡开宝的《语料库翻译学概论》和《基于语料库的莎士比亚戏剧汉译研究》、王克非的《语料库翻译学探索》、肖忠华的《英汉翻译中的汉语译文语料库研究》、黄立波的《基于语料库的翻译文体研究》以及戴光荣的《译文源语透过效应研究》等专著,本书对此不再赘述。本书更多关注的是语料库与翻译实践之间的互动关系和相互影响,希冀能够在语料库翻译学的理论指导下,从根本上提升翻译实践尤其是非文学翻译实践的理论意义。这是我多年来追随语料库翻译学的动力之所在,虽然真正实现翻译实践理论意义的路途可能尚属遥远,但随之而来的乐趣已经成为了我写作本书的一个动因。

提笔写作本书的目的在于对自己 20 多年来从事翻译实践经验的总结,具体说是对翻译实践与翻译技术、语料库、翻译教学等

相结合的实践经验的总结,期望能够通过经验总结进而在翻译实践理论上有所长进。至于能否真正如己所愿,自己说了不算,还得有待同行专家或领域专家的首肯与认可。从总结经验到推出理论是一个充满不确定性的过程,虽然期间困难重重,但每每稍有收获之时,便喜悦如泉涌。在写作过程中,感觉最为关键的是文字的提炼。翻译经验毕竟只是翻译经验,它可能只对某一方面的应用最为合适,若要稍作适用性拓展,则必须对文字表述进行多番锤炼和升华。相信,现在文中所呈现的经总结归纳后的文字表述定能对翻译实践和MTI教学实践有所帮助。本人自认为本书已经具备了以下三大特点:

- 对翻译实践的可应用性;
- 对MTI教学的可适应性;
- 对翻译理论的可拓展性。

第一,对翻译实践的可应用性。本书所呈现的知识均以本人多年来所做的大量翻译实践为依据,可以说有近80%的内容是完全依赖于经验的总结、归纳、提炼和升华。这是可应用性特点的一种面向翻译实践的体现。另一体现是将经过总结、归纳、提炼和升华后的文字表述交由业界专家审定,确认其对具体翻译实践具有指导意义后才予以定稿。在此,要感谢对文字进行业界技术相符性审定的徐军先生。他以其复旦软件工程专业的教育背景、计算机辅助翻译行业的从业背景和欧洲具体翻译项目的管理背景来确保本书在涉及翻译技术的知识方面具有可靠性和可应用性。本书对翻译实践的可应用性还体现在散落于书中不同章节的各种有关翻译或语料库技术应用的“小技巧”。这些小技巧或配以流程图、或通过文字进行描述,均源自于具体翻译项目的翻译实践,具有很强的可操作性,可直接进行实践性运用。其目的在于加快读者学习翻译技术或语料库技术的进度,完善读者理解翻译技术或语料库技术的效果,提高读者运用翻译技术或语料库技术的效率。

第二,对MTI教学的可适应性。MTI教学中的计算机辅助翻译是翻译技术的一部分或其中的一个环节,其并不能代表整个翻译技术。因而,翻译技术的MTI教学有其复杂性,即如何将不同的翻译技术知识糅合在一起并有效地传授给学生,以及如何教会学生以习得的翻译技术知识去解决实际问题。



因为技术在不断更新和进步,而且更新和进步的速度在不断加快,现有的技术可能会在若干年后不见其踪影。由此可见,翻译技术或语料库技术教学的关键并不在于教会某项技术,而是如何使学生习得应用技术去解决问题的能力。对此,一个更为有效的教学途径是以翻译项目为导向的技术应用教学方法。该方法立足于:①彻底解读具体的翻译项目对技术或功能的具体要求;②将最合适的技术或功能应用于具体的项目要求;③优化设计将技术或功能应用于具体项目的流程;④将技术与学生的实际翻译能力相结合。总之,融合运用能力才是这一方法的关键所在。如果能以上述思路来阅读本书,相信一定会领悟到技术运用的妙处以及收获一份随技术应用而产生的心得。如若确实如此,那就不枉这里所说的本书对MTI教学的可适应性这一特点了。

第三,对翻译理论的可拓展性。一种理论的产生在其初始阶段必然是进行实验,然后是总结归纳,最后才上升为理论层面。翻译技术理论的产生也不例外。虽然迄今为止尚未见到翻译技术方面的具有指导意义的理论,但也不妨碍其实际应用。不过,如果能够将翻译实践活动纳入到理论的框架之下,这样的实践就不再是一种盲从,而是一种理智的行为。创建理论的意义就在于此,在于其对具体实践的指导意义。写作本书的初衷也在于此,但写作的过程却使我感受到了从经验总结到想要对翻译技术进行理论呈现的困难重重。无论如何,这次写作体验算是一次尝试,一次设想对现有的翻译理论稍作拓展的尝试。完成本书写作后,觉得翻译技术有多面内容可供学术研究,如人机互动翻译的边界研究、技术环境下的翻译策略、技术背景下的译者行为研究、企业翻译研究,等等。如果本书的若干章节可以被视为是对翻译技术的粗浅研究,那么我已经深感荣幸。我会以此为动力,继续这种“研究式”的写作体验。

本书共计6章,其中的第6章由陶友兰教授负责完成,其余章节均由我负责完成。第2章2.3.1节的“撰写系统”和“桌面排版系统”以及第3章3.5.3节的“案例3:某大型客车制造企业的语料质量提升”均由徐军先生完成初稿,由我最后定稿。特别是其中的“桌面排版应遵循的规则”完全由徐军先生根据多年的项目管理经验总结而成。此外,他还为本次写作贡献了许多富有价值的业界资料。在此特表谢意,同时感谢他参与本次写作讨论过程中所表现出

的热情和专业。还有两个群体是在无形之中参与了我的写作过程,一是来自上海交通大学、复旦大学和南京农业大学等学校的 MTI 学生,二是参与 2015 年和 2016 年上海翻译技术沙龙夏季翻译技术培训的老师和学员们。他们在课堂上提出了各种各样富有创意的问题,是他们给予了我机会,让我去深入思考如何才能使翻译技术的教学更有成效。感谢他们对翻译技术的热情以及所提出的多姿多彩的问题。在此,还要感谢复旦大学出版社的施胜今老师,是她的策划思想给予我们出版本书的机会。本书是 2015 年复旦大学研究生院批准的 2015—2016 年度研究生课程建设立项项目“翻译项目基础与实践”的部分成果。

写作本书的又一个体验是,在完成基于更多翻译实践经验的章节时,觉得下笔如有神、非常顺畅,但在那些经验积累稍有不足的方面,则深感写作的艰难。所以,对由此可能产生的缺点与不足先行说声道歉。希望领域专家和业界同仁多提宝贵意见,使我有机会在本书再版时予以更新。

管新潮

2016 年 9 月 30 日于上海

# 图示目录

图 2.1	待转换成 WORD 格式的 PDF 文档(文字与图形混排) .....	017
图 2.2	1953 年出版的繁体字版《资本论》单页 PDF 文档(上)以及 转换成 WORD 后的文字(下) .....	019
图 2.3	WORD 软件中用于降噪的“特殊格式” .....	021
图 2.4	消除语料文本中段落之间留有一个空行的操作方法 .....	021
图 2.5	待消除“<seg id = "4">”、“</seg>”和空格的文档 .....	022
图 2.6	消除格式后的文档 .....	022
图 2.7	含有 seg 符和标注词性的文档 .....	023
图 2.8	使用 EmEditor 进行降噪处理的界面 .....	024
图 2.9	可用于 SDL Trados 2011 的 WinAlign 对齐的文档格式 .....	024
图 2.10	可用于 ABBYY Aligner 2.0 对齐的文档格式 .....	025
图 2.11	“断开所有单元(D)”和“标准连接(S)” .....	026
图 2.12	ABBYY Aligner 的对齐界面 .....	028
图 2.13	WordSmith 工具词表(WordList)功能所统计的语料库文本 特征值 .....	029
图 2.14	WordSmith 工具(上)和 AntConc 工具(下)的词汇检索使用 界面 .....	030
图 2.15	“影响”一词在英汉法律平行语料库样例中的检索结果 .....	031
图 2.16	法律双语平行语料库的词频统计结果 .....	032
图 2.17	使用 SDL Trados 翻译记忆库的检索界面 .....	033
图 2.18	通过插件 S-Tagger for Framemaker 程序生成 .txt 双语 文件 .....	041

图 2.19	SDL Trados 2007 版及之前版本的翻译编辑器界面	045
图 2.20	SDL Trados 2015 版翻译编辑器界面	046
图 2.21	黑马校对工具的校对数据选项	050
图 2.22	WORD 软件的校对功能	050
图 2.23	验证功能内所包含的各项质量保证参数	053
图 2.24	经质量保证处理的部分译文	054
图 2.25	适用于翻译记忆工具的 .tmx 格式文件	056
图 2.26	MultiTerm Convert 可转换的术语格式	057
图 2.27	英汉对照部分之间未留有空格的术语	058
图 2.28	使用 Text Tools 软件转换格式	059
图 2.29	MultiTerm 功能模块所显示的术语	060
图 3.1	语料库创建流程简图	067
图 3.2	语料采集流程图	070
图 3.3	语料降噪处理流程图	075
图 3.4	句对关联编号双文件保存格式	078
图 3.5	TMX 句对保存格式	079
图 3.6	TMX-ParaConV 软件操作界面	080
图 3.7	将指定格式的文件转换为 TMX	081
图 3.8	记忆库格式选择	087
图 3.9	Xbench 工具的整理优化界面	088
图 4.1	语料库质量的构成	098
图 4.2	House 翻译质量评估新模式(刘毅、鲍德旺 2016)	102
图 4.3	译文错误类型与翻译质量关系模型	103
图 4.4	ISO 17100  workflow (ISO 17100: 2015: 12)	108
图 4.5	美国的翻译质量保证各环节 (ASTM 2014: 6)	109
图 4.6	SDL Trados 的长度验证功能	113
图 5.1	使用 WinAlign 生成 TMX 格式的功能	123
图 5.2	使用 ABBYY Aligner 生成 TMX 格式的功能 (Export to TMX)	123

图 5.3	使用 Tmxmall 生成 TMX 格式的功能(导出)	123
图 5.4	记忆库维护的基本检查项目	126
图 5.5	记忆库中源文相同、译文不同的情形	126
图 5.6	SDL Trados 的“翻译记忆库”	127
图 5.7	TMX Editor 的维护功能	128
图 5.8	不同工具的匹配统计报表	130
图 5.9	SDL Trados 翻译编辑器中的匹配显示	131
图 5.10	SDL Trados 中最低匹配率的设置	132
图 5.11	SDL Trados 中的自动沿用功能	133
图 5.12	PerfectMatch 功能所呈现的匹配结果	135
图 5.13	从 SDL MultiTerm 中导出含有制表符的术语词条	137
图 5.14	术语库去重流程	138
图 5.15	AutoSugget 词典自动提示单词或短语	138
图 5.16	双文件 ParaConc 格式转换为 .tmx 格式流程	140
图 6.1	基于语料库的翻译教学路线(刘泽权、刘鼎甲,2011: 50)	169
图 6.2	“调解”一词英译统一化过程	178
图 6.3	动态翻译教学语料库建设流程图	182

# 表格目录

表 1.1	代表性语料库所属类型 .....	009
表 3.1	语料库库容对比 .....	071
表 3.2	语言对标记 .....	080
表 3.3	专用语块库 .....	082
表 4.1	译文错误类型分类 .....	101
表 4.2	通用汉英对应语料库英译汉部分(非文学)与英汉医学平行语料 库英译汉部分的翻译文字量对比 .....	111
表 4.3	2015 年度国家社科项目 .....	115
表 4.4	2016 年度国家社科项目 .....	116
表 5.1	李克兴和张兴红提供的专业通用词表 .....	145
表 5.2	MTI 专业学生以课程论文形式呈现的法律专业通用词 .....	147
表 5.3	英汉医学平行语料库的部分英至汉专业通用词 .....	152
表 5.4	英汉医学平行语料库的部分汉至英多译词汇对比 .....	153
表 6.1	原文词汇统计 .....	176
表 6.2	搜索语料库“调解”在不同用法下的英译 .....	177
表 6.3	“调解”的译法 .....	178

# 目录 | CONTENTS

<b>第 1 章 绪论</b> .....	001
1.1 语料库的起源与发展 .....	001
1.1.1 语料库的定义 .....	001
1.1.2 代表性语料库及其特征 .....	002
1.2 语料库的类型 .....	007
1.2.1 单语与双语(多语)类型 .....	007
1.2.2 平行与可比类型 .....	008
1.2.3 通用与专用类型 .....	008
1.2.4 单向与双向(多向)类型 .....	009
1.2.5 小结 .....	010
1.3 语料库语言学与翻译学 .....	010
1.3.1 语料库语言学 .....	011
1.3.2 语料库翻译学 .....	011
1.4 本书概要 .....	013
<b>第 2 章 语料库技术和翻译技术</b> .....	015
2.1 引言 .....	015
2.2 语料库技术 .....	016
2.2.1 语料格式转换工具 .....	016
2.2.2 语料降噪工具 .....	020
2.2.3 语料对齐工具 .....	024
2.2.4 语料库检索和分析工具 .....	027

2.2.5	问题与展望 .....	034
2.3	翻译技术 .....	035
2.3.1	文档管理类工具 .....	036
2.3.2	源文准备类工具 .....	040
2.3.3	译文处理类工具 .....	042
2.3.4	翻译管理类工具 .....	051
2.3.5	语料管理类工具 .....	056
2.3.6	本地化工具 .....	061
2.3.7	问题与展望 .....	062
<b>第3章</b>	<b>语料库创建流程 .....</b>	<b>066</b>
3.1	引言 .....	066
3.2	语料库的设计 .....	067
3.3	语料采集和降噪 .....	068
3.3.1	语料采集 .....	068
3.3.2	语料降噪 .....	074
3.4	语料对齐及其保存格式 .....	076
3.4.1	语料对齐 .....	076
3.4.2	语料对齐保存格式 .....	077
3.5	语料库应用 .....	081
3.5.1	案例1: 计算机辅助翻译项目应用 .....	081
3.5.2	案例2: 以组合语料库训练翻译能力 .....	083
3.5.3	案例3: 某大型客车制造企业的语料质量提升 .....	085
3.5.4	案例4: 对法律法规术语外译的建议 .....	089
3.5.5	案例5: 语料库证据在美国的法庭应用 .....	090
3.5.6	案例6: 两岸三地英汉科普历时平行语料库 .....	091
3.5.7	案例7: 《资本论》德汉平行语料库 .....	093



<b>第 4 章 语料库质量</b> .....	097
4.1 引言 .....	097
4.2 语料库质量的多维性 .....	098
4.2.1 语料库的设计质量 .....	098
4.2.2 语料的翻译质量 .....	099
4.2.3 语料的采集与降噪质量 .....	105
4.2.4 平行语料库的对齐质量 .....	106
4.2.5 应用于具体项目的语料组合质量 .....	106
4.3 源自于翻译实践的语料库质量 .....	107
4.3.1 相关标准的翻译流程设置 .....	108
4.3.2 质量保证理念与工具 .....	110
4.3.3 句对库中的高质量句对 .....	114
4.4 应用于译学研究的语料库质量 .....	114
<b>第 5 章 语料库与翻译实践</b> .....	119
5.1 引言 .....	119
5.1.1 计算机辅助翻译的特点 .....	120
5.1.2 计算机辅助翻译的分类 .....	121
5.2 语料库与记忆库 .....	121
5.2.1 语料库与记忆库的区别 .....	121
5.2.2 记忆库的创建与维护 .....	122
5.2.3 记忆库的匹配 .....	125
5.2.4 记忆库的更新与组合 .....	134
5.2.5 记忆库的其他作用 .....	136
5.3 翻译项目中语料库的使用 .....	139
5.3.1 双语平行语料库的翻译应用 .....	140
5.3.2 可比语料库的翻译应用 .....	141