

翻译学

核心话题系列丛书

Key Topics in
Translation Studies



外语学科核心话题
前沿研究文库

- 语料库翻译研究
Corpus-based
Translation Studies

语料库文体统计学方法与应用



Corpus Stylostatistics: Methods and Applications

胡显耀 著

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

“十三五”国家重点出版物出版规划项目

KEY TOPICS
TK

翻译学

核心话题系列丛书

Key Topics in

Translation Studies

外语学科核心话题
前沿研究文库

● 语料库翻译研究

Corpus-based

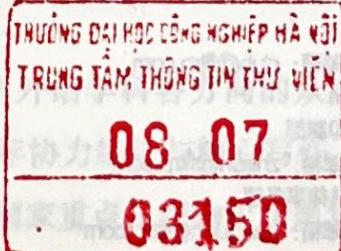
Translation Studies

语料库文体统计学方法与应用

*

Corpus Stylostatistics: Methods and Applications

胡显耀 著



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

出版前言

随着中国特色社会主义进入新时代，国家对外开放、信息技术发展、语言产业繁荣与教育领域改革等对我国外语教育发展和外语学科建设产生了深远影响，也有力推动了我国的外语学术出版。为梳理学科发展脉络，展现前沿研究成果，外语教学与研究出版社汇聚国内外语学界各相关领域专家学者，精心策划出版了“外语学科核心话题前沿研究文库”（下文简称“文库”）。

“文库”精选语言学、应用语言学、翻译学、外国文学研究和跨文化研究五大方向共25个重要领域100余个核心话题，按一个话题一本书撰写。每本书深入探讨该话题在国内外的研究脉络、研究方法和前沿成果，精选经典研究及原创研究案例，并对未来研究趋势进行展望。“文库”在整体上具有学术性、系统性、前沿性与引领性，力求做到点面结合、经典与创新结合、国外与国内结合，既有全面的宏观视野，又有深入的细致分析。

“文库”项目邀请国内外语学科各方向的众多专家学者担任总主编、子系列主编和作者，经三年协力组织与精心写作，自2018年底陆续推出。“文库”已获批“十三五”国家重点出版物出版规划项目，作为一个开放性大型书系，将在未来数年内持续出版。我们计划对“文库”进行不定期修订，使之成为外语学科的经典著作。

我们希望“文库”能够为外语学科及其他相关学科的研究生、教师及研究者提供有益参考，帮助读者清晰、全面地了解各核心话题的发展脉络，并有望开展进一步深入研究。期待“文库”为我国外语学科研究的创新发展与成果传播作出更多积极贡献。

外语教学与研究出版社

2018年11月



目录

总序	王克非	xi
----------	-----	----

前言	胡显耀	xvi
----------	-----	-----

第一章 语料库文体统计学简史		1
----------------------	--	---

1.1 什么是语料库文体统计学?		1
------------------------	--	---

1.2 现代文体统计学的发展		3
----------------------	--	---

1.2.1 起源时期		3
------------------	--	---

1.2.2 突破时期		5
------------------	--	---

1.2.3 挣扎时期		6
------------------	--	---

1.2.4 多变量时期		7
-------------------	--	---

1.2.5 累积和时期		9
-------------------	--	---

1.2.6 人工智能时期		11
--------------------	--	----

1.3 语料库文体统计学发展简史		12
------------------------	--	----

1.3.1 语料库与文体统计学		12
-----------------------	--	----

1.3.2 语料库文体统计学的发展		13
-------------------------	--	----

推荐阅读		22
------------	--	----

第二章 语料库文体统计学方法		23
----------------------	--	----

2.1 描述文体特征的层次和变量		23
------------------------	--	----

2.1.1 形态特征		25
------------------	--	----

2.1.2 词汇特征		29
------------------	--	----

2.1.3	搭配特征	32
2.1.4	句法特征	34
2.2	语料库和语料库分析方法	36
2.2.1	语料库的类型	38
2.2.2	代表性语料库	40
2.2.3	语料库分析方法	51
2.3	文体统计分析方法	64
2.3.1	统计分析的基本概念	64
2.3.2	对文体集中趋势的描述统计	68
2.3.3	对文体离中趋势的描述统计	69
2.3.4	文体差异显著性的假设检验	71
2.3.5	多元统计分析方法	77
2.4	小结	80
	推荐阅读	81
第三章 语料库文体统计学与作者辨别		82
3.1	作者辨别研究概述	82
3.1.1	传统作者辨别研究	83
3.1.2	现代作者辨别研究	85
3.2	统计分析方法解析	92
3.2.1	判别分析	92
3.2.2	对应分析	96
3.3	语料库文体统计学的应用	99
3.3.1	判别分析在作者辨别中的应用	100
3.3.2	对应分析在期刊文章作者辨别中的应用	105
3.4	小结	111

推荐阅读	112
第四章 语料库文体统计学与文学文体学	113
4.1 文学文体学发展概述	113
4.2 统计方法解析	118
4.2.1 聚类分析	118
4.2.2 主成分分析	121
4.3 语料库文体统计学的应用	124
4.3.1 《黑暗的心》语言风格量化研究	124
4.3.2 詹姆斯小说风格量化研究	133
4.4 小结	146
推荐阅读	146
第五章 语料库文体统计学与翻译学	148
5.1 语料库翻译学概述	148
5.1.1 简化	150
5.1.2 显化	151
5.1.3 规范化	152
5.1.4 匀质化	154
5.1.5 译入语独特项减少	155
5.1.6 源语干扰	156
5.1.7 异常搭配	156
5.2 统计分析方法解析	157
5.2.1 方差分析	157
5.2.2 因子分析	162

5.3	语料库文体统计学的应用	166
5.3.1	翻译英语的源语干扰现象	166
5.3.2	翻译英语的变体特征研究	182
5.4	小结	197
	推荐阅读	197

第六章 语料库文体统计学与刑侦语言学 198

6.1	刑侦语言学的发展概述	198
6.1.1	刑侦语言学的起源与发展	199
6.1.2	刑侦语言学的主要研究领域和方法	201
6.2	统计方法解析	205
6.2.1	非参数检验	205
6.2.2	贝叶斯统计方法	206
6.3	语料库文体统计学方法的应用	211
6.3.1	伊万斯谋杀案供述语言风格研究	211
6.3.2	《联邦党人文集》作者争议研究	217
6.4	小结	224
	推荐阅读	225

第七章 语料库文体统计学与数据挖掘 226

7.1	数据挖掘概述	227
7.1.1	什么是数据挖掘?	227
7.1.2	数据挖掘的任务和方法	229
7.1.3	数据挖掘的应用和发展趋势	232
7.1.4	数据挖掘与文体分析	234

总序	7.2 数据挖掘方法解析	236
	7.2.1 神经网络	236
	7.2.2 决策树	243
	7.3 数据挖掘方法在文体研究中的应用	245
	7.3.1 基于神经网络的莎士比亚与弗莱彻文体风格研究	245
	7.3.2 基于机器学习的翻译语言特征研究	255
	7.4 小结	261
	推荐阅读	261
参考文献		263
附录一: 第5章研究案例2采用的96个英语语言特征总表		288
附录二:《文体统计分析方法操作指南》网址		291
索引		292

季羡林先生曾从事非常重要的文化翻译工作,对翻译有着深切的体会和很高的评价。他谈及五千年中华文明虽历经波折却始终绵延不绝、蓬勃发展的原因时,曾这样说:

倘若拿河流来作比,中华文化这一条长河,有水满的时候,也有水少的时候,但却从未枯竭。原因就是有新水注入。注入的虽数大大小小是很多的,最大的有两次,一次是从印度来的水,一次是从西方来的水。而这两次的大注入依靠的都是翻译。中华文化之所以能长盛不衰,万般是因就是翻译。翻译之为用大矣哉!

翻译是翻译对中华文化的演进、对人类各民族文化交流发展的巨大作用。

翻译活动最基本的三个要素是语言、文化和译者。语言离不开民族、社会、文化,并浸润其中;双语文本转换离不开执行转换的人,即译者;