

EFL Error Patterns in China

英语错误型式分析

杨达复 著



陕西人民出版社

(陕)新登字 001 号

EFL Error Patterns in China

英语错误型式分析

杨达复 著

陕西人民出版社出版发行

(西安北大街 131 号)

西安外国语学院印刷厂印刷

850×1168 毫米 32 开本 11.125 印张 270 千字

2000 年 12 月第 1 版 2000 年 12 月第 1 次印刷

印数:1-1200

ISBN 7-224-05550-6/H·200

定价:14.50 元

前 言

本书是作者负责国家哲学社会科学研究规划“九五”基金资助课题“基于语料库的中国英语学习者语际语特征分析”(98BY013)的成果之一。该课题研究的是中国学生英语写作方面的语言特征。课题的意义在于和近年来的国际同类研究接轨,探索基于学习者语料库的第二语言习得研究途径。这一应用型研究的先决条件是:(1)应用语料库语言学的研究方法;(2)了解学生英语写作的语言特点。课题研究采用数据驱动的方法,以中国学生英语写作中出现频率相对较高的错误类型为切入点,在英汉双语对比的基础上侧重于从汉语知识影响的角度对其写作错误产生的原因进行较为系统的分析,得出的结论对英语写作教学和英语学习均有一定的参考价值。作者的这一选择基于本人对跨语言影响研究在外语教学中的必要性的认识及对应用语料库语言学的方法进行第二语言习得研究这一国际潮流的关注。原因有五:

1. 语言学家已经意识到语料库在语言研究中的重要地位、语料库研究的可证明性及语言事实的可提取性等优势。基于语料库的语言研究正在不断地显示着它在揭示人类语言本质方面的潜势,已经成为国际语言学研究领域的一个热点。我们应该顺应这一潮流,充分利用计算机技术,在外语教学与研究领域开展语料库语言学研究。

2. 作为一种方法,基于语料库的语言研究不同于基于内省的语言研究但与其形成优势互补。前者以来自语言行为的真实数据为依据,通过计算机的辅助实现以统计方法为主要手段的精确定量分析。这种分析虽然有其优势,但在定性分析方面却有很大的局限性。这主要是因为定性分析在智能上的要求超越了目前计算机的实际能力,因此需要研究人员的介入。由于定性分析和定量分析一样重要,基于语料库的系统性语言研究应当是在定量分析的基础上进而进行定性分析。这种定性分析可在某一语言学理论框架中进行,也可采用基于经验的描写方法。本课题的研究选择的是描写的方法。

3. 绝大多数英语语料库是为语言学的一般研究而建立的。为了进行基于语料库的语际语(interlanguage)研究,可以通过搜集学生英语写作等语言材料来建立适合这一用途的语料库。国外于九十年代初期开始了学生英语语料库的创建。我们也应在创建中国学生英语语料库的基础上进行同类研究,从而缩小和国际同行间的距离。

4. 在过去的几十年间,国内第二语言习得研究规模较小,样本有限。共时性和历时性对比研究都显得不足。除了缺乏系统性之外,研究方法和语言数据类型方面的失衡也非常突出(见第一章)。本书作者已意识到国内在第二语言习得研究方面的这些问题并力图在自己的研究中避免之。

5. 基于学生语料库的英语写作错误分析目前尚不多见。而跨语言影响作为第二语言习得的一个重要研究方面在经历了数十年的“失宠”之后又重新受到普遍重视。这两方面都有许多研究工作要做。本课题将它们结合起来,分析力求有一定的广度和深度,从而使其对国内的第二语言习得研究和英语教学有一定

的借鉴作用。

本书从对国外近几十年来在对比分析、错误分析、跨语言影响及语料库语言学方面所做研究的回顾入手,继而提出该课题研究的理论基础、任务、方法和有待验证的假设。书的后半部分为学生语料库中英语写作错误的定量分析以及基于定量分析的定性分析,内容涉及错误的多维分布、有关假设的验证、汉语知识在英语写作方面对中国学生的潜在影响及对本书整个研究的总结。各章概要如下:

第一章:绪论。本章简要论述基于语料库的语言研究的重要性、必要性和可能性以及本研究的动机和课题的整体设计。

第二章:语际语研究回顾。本章在简要介绍语际语及其特征之后即对其三个研究方面(对比分析、错误分析与跨语言影响)在近几十年间的状况进行回顾。因为基于学生语料库的英语写作错误定性分析是本课题的核心,语际语研究回顾也以本课题的研究内容为中心进行。

第三章:语料库语言学。本章讨论其发展过程、研究方法、研究领域、语料库设计、标注赋码与信息提取等。

第四章:课题设计。本课题为基于语料库的语际语研究,因此遵循语料库语言学的一般方法来进行。本章详述课题的整体规划:原则、任务、方法、假设以及学生英语写作语料库的建立、错误标注系统的设计、错误赋码与信息提取、学生错误的一些普遍特征及其分析与处理等。

第五章:学生语料库错误的定量分析。本章报告 LOB 语料库(抽样)与学生语料库的语言特征比较结果(内容涉及词长、句长、词汇密度、词类比例等),学生英语写作错误在各维度上的分布状况及其与学生写作质量之间的关系。本章报告的研究结果

为以下四章的定性分析提供了依据。

第六—九章：学生语料库错误的定性分析。这四章分析定量分析时出现频率较高的英语写作错误。每章各有侧重但步骤大致相同（见目录）。因对学生错误产生原因的定性分析侧重于汉语知识的影响，故每章都有一节专门对比英语和汉语的相关语言特征。这四章的内容分别为：实词错误与汉语影响，措词错误于汉语影响，汉语知识对英语词汇组合的影响，汉语知识对英语句式的影响。

第十章：本课题的研究结果及其启示。本章总结本课题的全部研究结果，讨论与对比分析、错误分析、跨语言影响相关的一些语言习得问题以及本课题的研究对英语教学与学习的启示。

本书涉及的课题研究得到广东外语外贸大学桂诗春教授的悉心指导；该校许罗迈教授提供了中国四所高校英语专业学生作文（属桂诗春教授主持创建的大型中国学生英语语料库中的部分语料）和部分自然语言处理软件；华南理工大学秦秀白教授曾对本书初稿提出修改意见；M. McCausland, A. Giles 和 K. Holburn 评析过书中提到的所有错误语句；L. Godfrey 和 G. Byrne 阅读过本英语书稿并就其文字表达提出过一些意见。作者对他们的指导和帮助表示衷心的感谢。

本书的出版得到西安外国语学院和陕西人民出版社、西安外语音像出版社的热情支持，谨致谢意。

作者学识有限，书中论述不当之处在所难免。敬请学界前辈、专家学者、同行及读者不吝赐教。

作者

2000年10月

目 录

略语表

数据目录

图表目录

1 绪论	1
1.1 基于语料库的语言研究:重要性、必要性和可能性	1
1.2 基于语料库的语际语研究	4
1.3 本书各章概要	6
2 语际语研究回顾	9
2.1 语际语及其特征	9
2.2 对比分析	15
2.3 错误分析	20
2.3.1 错误类型分析	21
2.3.2 错误原因分析	26
2.3.3 错误分析:重要意义与不足之处	28
2.4 语际语研究:从迁移到跨语言影响	33
2.4.1 迁移的定义	33
2.4.2 从迁移研究到跨语言影响研究	35
2.4.3 迁移研究回顾	39
2.4.4 可迁移性	59
2.5 语际语研究方面的进展、存在问题与新的动态	61
3 语料库语言学:发展、方法与研究领域	65
3.1 语料库建设是基于语料库的语言研究的基础	65
3.1.1 语料库建设:发展、分类与原则	65

3.1.2 语料库建设:标注赋码与信息提取	69
3.2 语料库语言学的研究方法:定量分析 与定性分析	82
3.3 基于语料库的语言研究	85
3.3.1 语料库在语言研究中的应用	85
3.3.2 基于语料库的语际语研究	87
4 课题设计	93
4.1 研究的基础、任务与方法	93
4.2 假设的验证	97
4.2.1 定量验证的假设	97
4.2.2 定性验证的假设	100
4.3 本研究课题中的语料库建设	102
4.3.1 英语专业学生语料库的建设	102
4.3.2 LOB 语料库的抽样	105
4.4 语际语错误:定义与标注系统	106
4.4.1 错误的定义	106
4.4.2 错误标注系统的设计	107
4.5 错误赋码与数据提取	116
4.5.1 错误赋码	116
4.5.2 数据提取	117
4.6 学生写作错误的一些特征及其处理	119
5 学生语料库错误的定量分析	125
5.1 LOB 语料库(抽样)与学生语料库的 语言特征比较	125
5.2 学生写作错误的分布	136
5.2.1 学生写作错误的范围分布	136
5.2.2 学生写作错误的层次分布	139
5.2.3 学生写作错误的词类与结构分布	141
5.2.4 学生写作错误的类型分布	147
5.2.5 学生写作错误的原因分布	150

5.3 学生写作错误的出现频率与作文质量之间的关系	152
5.3.1 写作质量与错误识别范围之间的关系	153
5.3.2 写作质量与错误层次之间的关系	154
5.3.3 写作质量与词类之间的关系	156
5.3.4 写作质量与错误类型之间的关系	158
5.3.5 写作质量与错误原因之间的关系	160
5.4 结 论	163
6 实词错误与汉语影响	169
6.1 引言	169
6.2 英语和汉语中的实词特征比较	171
6.3 动词错误分析	178
6.3.1 及物动词与不及物动词的混淆	178
6.3.2 其他动词类型误用	183
6.3.3 语态错误	187
6.4 名词错误分析	190
6.4.1 可数名词与不可数名词的混淆	190
6.4.2 可数名词错误	193
6.4.3 名词类型误用	195
6.5 形容词错误分析	197
6.5.1 形容词形式与位置错误	197
6.5.2 形容词类型误用	199
6.6 结 论	200
7 措词错误与汉语影响	205
7.1 引 言	205
7.2 学生措词错误的语言学思考	206
7.3 词类错误分析	213
7.4 替代型错误分析	219
7.5 省略型错误分析	223
7.5.1 冠词省略引起的错误	224

7.5.2 其他结构词省略引起的错误	228
7.6 冗余型错误分析	231
7.7 结 论	234
8 汉语知识对英语词汇组合的影响	237
8.1 引 言	237
8.2 英语和汉语中的词汇组合特征比较	239
8.3 错误分析:数的一致问题	244
8.4 词汇搭配方面的错误	253
8.4.1 实词之间的搭配错误	253
8.4.2 实词与介词之间的搭配错误	257
8.5 结 论	259
9 汉语知识对英语句式的影响	263
9.1 引 言	263
9.2 英语句法与汉语句法	264
9.3 词序方面的错误分析	273
9.4 <i>there be</i> 句型使用错误分析	278
9.5 结 论	284
10 本课题的研究结果及其启示	289
10.1 本课题的研究结果	289
10.2 语际语研究方面的一些思考	297
10.2.1 对比分析与错误分析在语际语	
研究方面的作用	297
10.2.2 母语在外语学习中的作用	299
10.2.3 本课题的研究对外语教学的启示	302
附录	305
参考文献	324

Contents

List of abbreviations	
List of tables	
List of figures	
1 Introduction	1
1.1 Corpus-based studies of language: importance, necessity and possibility	1
1.2 The motivation for a corpus-based approach to IL studies	4
1.3 An outline of the present research project	6
2 IL studies in retrospect	9
2.1 IL and its characteristics	9
2.2 Contrastive analysis	15
2.3 Error analysis	20
2.3.1 Studies on error types	21
2.3.2 Studies on error sources	26
2.3.3 EA: its significance and weaknesses	28
2.4 IL studies: from transfer to cross-linguistic influence	33
2.4.1 Transfer defined	33
2.4.2 From transfer to cross-linguistic influence	35
2.4.3 Transfer studies: a general view	39
2.4.4 Transferability	59
2.5 Achievements, problems and new trends in IL studies	61

3 Corpus linguistics: development, methodology and exploration	65
3. 1 Corpora building as the basis of corpus-based studies of language	65
3. 1. 1 Corpora building: development, classification and principles	65
3. 1. 2 Corpora building: annotation and information retrieving	69
3. 2 Methodolgy for corpus-based studies: qualitative and quantitative analyses	82
3. 3 Corpus-based language studies	85
3. 3. 1 The general use of corpora in language studies	85
3. 3. 2 Corpus-based IL studies	87
4 Project design	93
4. 1 Research foundation, tasks and methods	93
4. 2 Hypotheses for testing	97
4. 2. 1 Hypotheses to be tested by the quantitative approach	97
4. 2. 2 Hypotheses to be tested by the qualitative approach	100
4. 3 Corpora building for the present study	102
4. 3. 1 The building of the English majors' IL corpora	102
4. 3. 2 The building of a sample LOB corpus	105
4. 4 ILC errors: their definition and marking system	106
4. 4. 1 Error definition	106
4. 4. 2 EMS designing	107
4. 5 Error tagging and data retrieval	116
4. 5. 1 Error tagging	116
4. 5. 2 Data retrieval	117

4. 6 Some traits of ILC errors and their treatment	119
5 Quantitative analysis of the ILC errors	125
5. 1 Analysis of the basic linguistic features between S-LOB and ILC	125
5. 2 Dimensional distribution of the ILC errors	136
5. 2. 1 Error distribution at RANGE	136
5. 2. 2 Error distribution at LEVEL	139
5. 2. 3 Error distribution at CLASS	141
5. 2. 4 Error distribution at TYPE	147
5. 2. 5 Error distribution at SOURCE	150
5. 3 The relationship between error occurrences and group performance	152
5. 3. 1 Relation at RANGE	153
5. 3. 2 Relation at LEVEL	154
5. 3. 3 Relation at CLASS	156
5. 3. 4 Relation at TYPE	158
5. 3. 5 Relation at SOURCE	160
5. 4 Conclusion	163
6 Notional words and L1 influence	169
6. 1 Introduction	169
6. 2 Notional words in English and Chinese	171
6. 3 Analysis of verb errors	178
6. 3. 1 Errors in transitivity	178
6. 3. 2 Errors in verb patterns	183
6. 3. 3 Errors in voice	187
6. 4 Analysis of noun errors	190
6. 4. 1 Errors in countability	190
6. 4. 2 Errors in number	193
6. 4. 3 Errors in noun patterns	195
6. 5 Analysis of adjective errors	197

6.5.1 Errors in form and position	197
6.5.2 Errors in adjective patterns	199
6.6 Conclusion	200
7 Wording and L1 influence	205
7.1 Introduction	205
7.2 Linguistic considerations of errors	
in wording	206
7.3 Analysis of errors in parts of speech	213
7.4 Analysis of errors in substitution	219
7.5 Analysis of errors in omission	223
7.5.1 Errors in article omission	224
7.5.2 Errors in the omission of other	
structural words	228
7.6 Analysis of errors in redundancy	231
7.7 Conclusion	234
8 L1 influence in concordance	237
8.1 Introduction	237
8.2 Features of concordance in English and Chinese	239
8.3 Analysis of errors in agreement	244
8.4 Analysis of errors in collocation	253
8.4.1 Errors in collocation nn	253
8.4.2 Errors in collocation np	257
8.5 Conclusion	259
9 L1 influence in syntax	263
9.1 Introduction	263
9.2 English and Chinese syntax compared	264
9.3 Analysis of errors in word order	273
9.4 Analysis of errors in pivotal and existential	
constructions	278
9.5 Conclusion	284

10 Major findings, discussion and general considerations	289
10.1 Discussion on the major findings by the present study	289
10.2 General considerations	297
10.2.1 The function of CA and EA in IL studies	297
10.2.2 The role of L1 in L2 learning	299
10.2.3 The suggestion of the present study to L2teaching	302
Appendices	305
References	324

List of abbreviations

ACH	The Association for Computers and Humanities
ACL	The Association for Computational Linguistics
AL	applied linguistics
ALD	Oxford Advanced Learner's Dictionary
ALLC	The Association for Literary and Linguistic Computing
CA	contrastive analysis
CFL	Chinese as a foreign language
CLEC	Chinese learner English corpus
CLI	cross-linguistic influence
EA	error analysis
EMS	error marking system
EFL	English as a foreign language
ESL	English as a second language
ICE	international corpus of English
ICLE	international corpus of learner English
IL	interlanguage
ILC	interlanguage corpus
L1/L2	first/second language
NL	native language(=L1)
OED	The Oxford English Dictionary
SL	source language
SLA	second language acquisition
TACT	the textual analysis computing tools
TL	target language
UG	universal grammar

List of tables

Table 2. 1 Dimensions and development of lexical knowledge	43
Table 2. 2 Grammatical collocations	47
Table 2. 3 Lexical collocations	47
Table 2. 4 Attested cases of basic word order transfer in L2 acquisition	50
Table 2. 5 Attested cases of basic word order transfer in language attrition due to bilingualism	51
Table 2. 6 Illustrative language examples	53
Table 2. 7 Contrastive string arrangements	53
Table 4. 1 Information of the sub-corpora built by the number of words	104
Table 4. 2 Information of the sub-corpora built by the number of errors	104
Table 4. 3 Comparison of genres between ILC and S-LOB	105
Table 4. 4 TYPE errors	113
Table 5. 1 Basic linguistic statistics of S-LOB and ILC	127
Table 5. 2 Basic linguistic statistics of the ILC sub-corpora (W_1 - W_8)	130
Table 5. 3 Parts-of-speech frequencies of S-LOB, ILC and its sub-corpora	130
Table 5. 4 Correlation and chi-square statistics of parts of speech between S-LOB and ILC	131
Table 5. 5 One way ANOVA and correlations for parts of speech (W_1 - W_8)	131
Table 5. 6 The first 10 high frequency structural words	