

面向汉语作为第二语言教学的 语法点知识库构建研究

The Construction of Grammar Points Knowledge Base for TCSL

谭晓平 著

上海三联书店

面向汉语作为第二语言教学的 语法点知识库构建研究

The Construction of Grammar Points Knowledge Base for TCSL

谭晓平 著



上海三联书店

图书在版编目(CIP)数据

面向汉语作为第二语言教学的语法点知识库构建研究/谭晓平著. —
上海:上海三联书店,2018.10

ISBN 978-7-5426-6503-4

I. ①面… II. ①谭… III. ①对外汉语教学—语法—教学研究
IV. ①H195.3

中国版本图书馆 CIP 数据核字(2018)第 223616 号

面向汉语作为第二语言教学的语法点知识库构建研究

著 者 / 谭晓平

责任编辑 / 杜 鹃

特约编辑 / 周治华

装帧设计 / 一本好书

监 制 / 姚 军

责任校对 / 张大伟

出版发行 / 上海三联书店

(200030)中国上海市漕溪北路 331 号 A 座 6 楼

邮购电话 / 021-22895540

印 刷 / 上海惠敦科技印务有限公司

版 次 / 2018 年 10 月第 1 版

印 次 / 2018 年 10 月第 1 次印刷

开 本 / 890 × 1240 1/32

字 数 / 230 千字

印 张 / 9.25

书 号 / ISBN 978-7-5426-6503-4/H · 76

定 价 / 49.00 元

敬启读者,如发现本书有印装质量问题,请与印刷厂联系 021-63779028

本研究获得

“上海师范大学第九期重点学科（文科）建设计划（汉语国际教育）”

资助

前 言

构建面向汉语作为第二语言教学的语法点知识库是语言资源建设的一项重要内容,也是提高汉语作为第二语言教学信息化水平的重要途径。

本研究以三个平面理论、对外汉语教学语法理论为基础,对语法点知识库的描述对象、语法点描述框架、语法点标注体系进行了专门研究,构建了面向汉语作为第二语言教学的语法点知识库及应用平台,探讨了语法点知识库在语法点句法语义接口研究、汉语作为第二语言教学与教材研究等领域的应用。

面向汉语作为第二语言教学的语法点知识库主要由语法点描写知识库及与之配套的语法点标注语料库构成。该知识库的构建研究主要包括五个方面的内容:

第一,语法资源的需求分析。通过文献调研法及问卷调查法分析了教师及学习者对语法资源的需求,认为面向汉语作为第二语言教学的语法资源应具有综合性、关联性、便捷性的特点,并提供相关的频率信息。

第二,语法点选取研究。首先,通过对理论语法和对外汉语教学语法研究对象和范围的梳理,将“语法点”界定为:根据汉语作为第二语言教学的实际需求,对系统的语法知识进行切分、

选择后获得的语法教学项目,可以包括语素、词、短语、格式、句子(单句、复句)、句群、篇章等方面的内容。然后,根据针对性、高频性、可描述性、必要性原则,确定了 152 个语法点作为语法点知识库描述与标注的对象。

第三,语法点描述框架及语法点描写知识库构建研究。首先,以三个平面理论、对外汉语教学语法理论为基础,提出了语法点描述框架,认为语法点的描述应包括:语义知识、句法知识、句法语义关系知识、语用知识、教学指导知识;然后,在语法点描述框架的基础上,构建了语法点描写知识库,制定了知识库填写规范,通过 25 个属性项对语法点进行了结构化、数据化、形式化的描述。

第四,语法点标注体系及语法点标注语料库构建研究。首先,对语法点标注体系进行了研究,确定了标注对象、标注维度、标注颗粒度,并为 152 个语法点制定了标注规范(包括标注框架、标注说明);其次,在汉语国际教育动态语料库(141464 条语料,约 350 万字)中进行了标注实践,对标注体系,特别是语法点的标注规范进行了验证和修订;最后,在标注语料的基础上形成了语法点标注语料库。

第五,语法点知识库的数据分析及应用研究。在标注语料的基础上,研究了语法点句法语义关系,获取了语法点使用上的频率信息,为语法点的教学与研究,对外汉语教材研究、语言测试提供了数据参考和语料基础。此外,论述了语法点知识库应用平台的功能、实现方式,以及其在语法点教学与研究领域的作用。

本研究的特色及创新主要有三个方面:一是拓展了语言知识库的应用领域,深化了语料库的标注加工研究。语法点描写

知识库通过 25 个属性项对语法点的频率、句法语义关系、语用等信息进行了结构化、数据化、形式化的描述。构建了语法点标注体系,制定了语法点标注规范,提出了标注框架覆盖率的检测方法,在一定规模的语料中进行了标注实践,形成了语法点标注语料库。二是将语法点语义和句法之间的关系分析为“一对一”“一对多”“多对一”三种类型,并获取了语义与句法关系之间的频率信息,为语法点的教学与研究提供了参考数据。三是在语法点知识库基础上构建的应用平台为用户提供了综合性的语法知识,展现了语法点使用上的频率信息,实现了语法点与语料之间、语义类别与语料之间、句法形式与语料之间、不同语法点之间的关联性查询。

本研究为教师、研究者、学习者提供了备课、研究、测试、自学所需的科学的数据、丰富的用例、便捷的查询工具,有助于丰富汉语国际教育领域语言资源的类型,提升汉语作为第二语言教学的信息化水平,促进汉语国际教育事业的发展。

目 录

前 言 / 1

0 绪论 / 1

0.1 选题的缘起 / 1

0.1.1 汉语国际教育事业的发展对语言资源建设的需求 / 1

0.1.2 汉语作为第二语言教学的需求 / 2

0.1.3 汉语国际教育动态语料库建设的需求 / 4

0.2 研究价值 / 5

0.2.1 理论价值 / 5

0.2.2 应用价值 / 5

0.3 文献综述 / 6

0.3.1 知识库的相关研究 / 7

0.3.2 语料库的标注加工研究 / 12

0.3.3 汉语教学资源及应用平台的建设研究 / 21

0.4 理论基础 / 23

0.4.1 三个平面理论 / 23

0.4.2 对外汉语教学语法理论 / 25

0.4.3 语料库语言学理论 / 28

0.5 研究目标、内容和重点 / 28

0.5.1 研究目标 / 28

0.5.2 研究内容 / 29

0.5.3 研究重点 / 29

0.6 术语界定 / 29

0.6.1 汉语作为第二语言教学与对外汉语教学 / 29

0.6.2 语法点 / 30

0.6.3 语法点知识库 / 30

0.7 研究材料与方法 / 31

0.7.1 研究材料 / 31

0.7.2 研究方法 / 31

0.8 本书结构 / 33

1 语法教学资源概况及用户需求 / 35

1.1 语法教学资源的概况 / 35

1.1.1 资源的类型 / 35

1.1.2 资源的优势与不足 / 39

1.2 用户需求 / 40

1.2.1 综合性 / 40

1.2.2 频率信息 / 41

1.2.3 关联性 / 42

1.2.4 便捷性 / 42

1.3 语法点知识库的概况 / 43

1.3.1 语法点知识库的定位 / 43

1.3.2 语法点知识库的构成 / 44

1.4	本章小结 / 45
2	语法点的选取 / 46
2.1	语法研究的对象与范围 / 46
2.1.1	理论语法的研究对象与范围 / 46
2.1.2	教学语法的研究对象与范围 / 50
2.2	语法点的界定 / 58
2.3	语法点选取的原则与方法 / 60
2.3.1	语法点选取的原则 / 61
2.3.2	语法点选取的方法与步骤 / 67
2.4	语法点选取的结果 / 72
2.5	本章小结 / 75
3	语法点描写知识库的构建 / 76
3.1	相关理论 / 76
3.1.1	三个平面理论在教学中的应用 / 76
3.1.2	语法点讲解的原则与方法 / 79
3.2	语法点的描述框架 / 81
3.2.1	语义知识 / 81
3.2.2	句法知识 / 84
3.2.3	句法语义关系知识 / 94
3.2.4	语用知识 / 97
3.2.5	教学指导知识 / 100
3.3	知识库属性项的设置及填写规范 / 105
3.3.1	属性项的设置 / 106

- 3.3.2 知识库填写规范 / 107
- 3.4 本章小结 / 117
- 4 语法点标注语料库的构建 / 119
 - 4.1 语料基础 / 120
 - 4.1.1 语料的规模 / 120
 - 4.1.2 语料的构成 / 120
 - 4.1.3 语料的存储 / 123
 - 4.2 语法点标注体系 / 126
 - 4.2.1 标注对象与标注维度 / 127
 - 4.2.2 标注颗粒度 / 134
 - 4.2.3 标注规范 / 135
 - 4.3 标注难点及其标注规范 / 141
 - 4.3.1 述补结构的标注规范 / 141
 - 4.3.2 “把”字句的标注规范 / 160
 - 4.3.3 助词“了”的标注规范 / 168
 - 4.3.4 比较句的标注规范 / 183
 - 4.4 语法点的标注实践 / 187
 - 4.4.1 标注者与标注工具 / 187
 - 4.4.2 标注方法与流程 / 188
 - 4.5 语法点标注的一致性及准确率 / 194
 - 4.6 本章小结 / 195
- 5 语法点知识库的数据分析及应用 / 196
 - 5.1 数据分析与句法-语义接口研究 / 196

5.1.1	语法点句法和语义之间的映射关系 / 197
5.1.2	基于句法语义关系的教学 / 201
5.2	数据分析与教材及教学 / 204
5.2.1	语法点频率与教材及教学 / 205
5.2.2	语义频率与教材及教学 / 207
5.2.3	句法频率与教材及教学 / 208
5.3	数据分析与语法点编排 / 211
5.3.1	述补结构的编排 / 212
5.3.2	“把”字句的编排 / 224
5.3.3	比较句的编排 / 228
5.4	应用平台的主要功能及实现 / 231
5.4.1	应用平台的功能 / 232
5.4.2	应用平台功能的实现 / 241
5.5	语法点自动识别 / 243
5.5.1	语法点自动识别的任务 / 244
5.5.2	语法点自动识别的探索 / 245
5.6	本章小结 / 246
6	总结与展望 / 249
6.1	本研究的主要工作 / 249
6.2	不足之处和下一步工作计划 / 251
参考文献 / 253	
附录 / 265	
附录 1: 语言资源使用情况及需求分析调查问卷 / 265	

附录 2: 语法点描写知识库中对“一……就……”的描写 / 271

附录 3: 语法点的频率统计 / 273

附录 4: 助词“了”句法语义关系及相关统计 / 280

0 绪论

语言知识库的构建是自然语言处理领域资源建设的一项重要内容,其质量和规模关系到自然语言处理系统的优劣。语言知识库可以分为两类:第一类是包含词典、句法规则和语义概念的知识库;第二类是由真实文本及其所包含语言知识组成的语料库。(俞士汶等,2013)

本研究所论及的面向汉语作为第二语言教学的语法点知识库主要包括两方面的内容:一是语法点描写知识库,通过 25 个属性对语法点的句法、语义、语用等信息进行多维度、数据化、结构化、形式化的描述;二是与语法点描写知识库配套的语法点标注语料库,对语法点在语料中所体现的具体的句法、语义信息进行综合标注。

0.1 选题的缘起

0.1.1 汉语国际教育事业的发展对语言资源建设的需求

2011 年孔子学院大会标志着对外汉语教学转型为汉语国际教育(许嘉璐,2012)。截至 2015 年 12 月 1 日,全球有 134 个国家(地区)建立了 500 所孔子学院、1000 个孔子课堂。汉语国

际教育事业的迅速发展对语言资源的建设提出了新的需求,而中文信息处理技术的发展也为语言资源建设提供了技术上的支持。张普(1991)指出:“汉语信息处理与对外汉语教学是科学与技术的结合,是新科学与高技术的结合。科学与技术一经结合就会产生无穷的力量,就会发生飞跃。”^①詹卫东(2013)认为:“汉语语言学研究应该更加注重语言工程的研究和开发,以提高汉语大规模语言资源的数量、类型及易获得性,因此,汉语语言资源建设应努力实现语言范畴形式化、语言数据专项化和语言知识可视化。”^②

近年来,面向汉语作为第二语言教学领域的语言资源建设主要集中在中介语语料库上,例如北京语言大学的 HSK 动态作文语料库、中山大学的汉字偏误标注的汉语连续性中介语语料库等。此外,现代汉语语料库也在一定程度上促进了汉语作为第二语言的教学与研究,例如国家语委现代汉语平衡语料库、北京大学 CCL 语料库、北京语言大学 BCC 语料库等。但现有的语言资源建设尚不能满足汉语国际教育事业迅速发展的需求。目前,与中文信息处理技术相结合的、专门面向汉语国际教育领域的新型语言资源正在酝酿、发展,例如北京大学的现代汉语句式数据库、北京师范大学的语法资源库、汉语国际教育动态语料库等。

0.1.2 汉语作为第二语言教学的需求

汉语国际教育要从单纯的汉语教学转变为全面的文化交

① 张普.论汉语信息处理技术与对外汉语教学[J].语言教学与研究,1991,01:111—129.

② 詹卫东.大数据时代的汉语语言学研究[J].山西大学学报(哲学社会科学版),2013,05:70—77.

流,而实现全面交流的媒介是语言(许嘉璐,2012)。

语言的三要素包括:语音、词汇、语法。语法是语言的结构规则,“通过语法教学可以使学生准确地理解、准确地表达,尽量减少表达(包括口头表达和书面表达)中的语法毛病”^①。“从语言教学和习得的角度看,语法比语音和词汇更加难教,不容易学习,经常会出错;但习得一定的语法规则后,则能显著提升语言使用的效率。”^②

汉语作为第二语言的语法教学主要分为两种形式:第一种是系统的语法知识的教学,部分院校在中高级阶段开设有专门的语法课;第二种是以语法点为单位的语法教学。这种方式是以语言的应用为目标,将系统的语法知识分散在不同课型、不同阶段的教学中,强调对语法点的讲解和操练。受习得规律和学习者汉语水平的影响,第二种方式更为普遍。

教师对语法点的了解程度和教学上的处理方式直接影响着教学的效果。要想充分、全面了解一个语法点并在教学上做合适的安排是一件耗时、费力的事。通常需要通过教材中语法点的释义了解当课学习的重点,通过翻阅语法理论著作对语法点作全面的了解,再通过语料库寻找适合的、足量的例句以便在课堂中对语法点进行操练。而对于一些“新手”教师来说,更是难上加难。“新手”教师常常困惑于哪些内容属于语法点?语法点应该从哪些方面去教学?什么样的例句更适合于课堂教学?用例对于学生来说是不是太难?而“新手”教师所占的比例并不少,据统计,2010年—2012年派往海外进行汉语教学的3682名

① 陆俭明.“对外汉语教学”中的语法教学[J].语言教学与研究,2000,03:1—8.

② 袁毓林,詹卫东,施春宏.汉语“词库—构式”互动的语法描写体系及其教学应用[J].语言教学与研究,2014,02:17—25.

学生志愿者当中,非汉语国际教育专业的本科生占 57.3%,硕士生占 36.1%。^①

本研究正是着眼于这一需求,希望研究并建设一个综合教材、语法工具书、语料库各自优势的语言资源,提高教师备课的质量与效率。研究成果不仅能帮助教师备课,还可以为教材及试题编写提供数据参考和丰富的用例,帮助具有一定汉语水平的学习者进行自主学习,并在一定程度上推动中文信息处理技术在汉语国际教育领域的应用,最终促进汉语国际教育事业的发展。

0.1.3 汉语国际教育动态语料库建设的需求

“汉语国际教育动态语料库”是北京师范大学中文信息处理研究所承担的“十二五”国家 863 课题“海量文本多层次知识表示及中文文本理解应用系统研制”的研究成果之一。该语料库是面向汉语国际教育领域构建的语言资源,规模约为 350 万字。项目组从汉语作为第二语言教学的需求出发,对语料库展开了多义词、语法点、话题、交际功能的标注加工研究,并在标注语料的基础上开发了服务于汉语作为第二语言教学的应用平台,提供有基本检索、字词关联检索、多义词检索、语法点查询、话题查询、交际功能查询等多项功能。

本书的研究内容源自于“汉语国际教育动态语料库”构建过程中本人所承担的研究任务——语法点标注研究,而“面向汉语作为第二语言教学的语法点知识库构建研究”是对该项研究的深化和拓展。

^① 朱瑞平,钱多.汉语教师志愿者背景、动机与志愿者项目的可持续发展研究[J].国际汉语教学研究,2015,01: 63—68.