



中国学习者 花园幽径句的量化研究

杜家利 于屏方 李明琳 著



中国社会科学出版社



中国学习者 花园幽径句的量化研究

杜家利 于屏方 李明琳 著



中国社会科学出版社

图书在版编目 (CIP) 数据

中国学习者花园幽径句的量化研究 / 杜家利, 于屏方, 李明琳著.
—北京: 中国社会科学出版社, 2018. 12
ISBN 978 - 7 - 5203 - 3856 - 1

I. ①中… II. ①杜…②于…③李… III. ①英语—翻译—教学研究—高等学校 IV. ①H315.9

中国版本图书馆 CIP 数据核字 (2018) 第 292226 号

出 版 人 赵剑英
责任编辑 刘志兵
特约编辑 张翠萍等
责任校对 李 斌
责任印制 李寡寡

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010 - 84083685
门 市 部 010 - 84029450
经 销 新华书店及其他书店

印 刷 北京明恒达印务有限公司
装 订 廊坊市广阳区广增装订厂
版 次 2018 年 12 月第 1 版
印 次 2018 年 12 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 20.5
字 数 295 千字
定 价 89.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社营销中心联系调换
电话: 010 - 84083683
版权所有 侵权必究

国家语委语言文字科研项目优秀成果后期资助计划（HQ135-20）、
国家自然科学基金项目“面向汉译英口语测试中自动评测方法的
研究”（61877013）资助

序

魏向清

2018年3月28日，杜家利博士发来电邮，嘱我为其团队的新作《中国学习者花园幽径句的量化研究》写个序言。对此，我一开始比较犹豫，主要是因为这本学术著作标题中的三个关键词，除了“中国学习者”外，“花园幽径句”和“量化研究”于我而言都是较为陌生的研究对象或学术领域。按理我应婉言谢辞，以免贻笑大方。然而，恰恰又是这两个相对小众的研究关键词勾起了我的阅读兴趣，因为它们与我较为熟悉的“中国学习者”研究关联。从题目来看，这应该是一个与中国英语学习者有关的计算语言学应用研究课题，它所涉及的具体研究对象是花园幽径句的认知问题，采用了计算语言学的新视角和新方法。这让我回想起杜家利博士在南京大学外国语学院外国语言学与应用语言学博士后流动站驻研期间所做的一项非常有意义的教学工作。当时，杜家利博士给南大外院的硕博硕士生集中讲授了“计算语言学”课程，他的讲解深入浅出而且生动活泼，很快就打消了大家的“畏难”情绪，大家学习热情很高，收获很大。作为那次扫盲课程的受益者之一，虽然我对计算语言学的了解至今仍止步于入门水平，但杜家利博士的讲解加深了我对不同语言研究路径的认识，也引发了我诸多后续的思考，视野得以开阔。这次杜家利博士的团队将计算语言学的理论研究与二语学习的实践问题结合起来，应该是我作为

应用语言学研究者进一步拓展学习的好机会。加之杜家利博士盛情邀请,我想,作为一个外行读者,我的阅读反馈或许会给他的这项开拓性研究提供些后续深化的另类参考。于是便有了以下这些感悟式“絮言”。

作为语言处理过程中的一种局部歧义的复杂结构,“花园幽径句”(garden path sentences)的研究自20世纪70年代起,至今已有近半个世纪的历史。从“花园幽径句”研究的源流来看,西方学者最早开始探索,而且一直在持续开展相关研究,形成了较为丰富的理论成果。就早期研究而言,研究者关注的焦点大多在于“花园幽径效应”产生背后的理论机制,他们从各自熟悉的理论视域去解释和讨论,理论研究的跨学科性很突显。进入21世纪以来,有关“花园幽径句”的实证研究成为新的研究趋势与热点。相比之下,根据中国知网(CNKI)的数据统计与分析结果,我国的同类研究起步较晚且研究较少。具体来看,以“花园幽径句”为主题的研究文献,自2003年起至今仅26条记录,而且以计算语言学领域的理论研究为主。如果添加“(中国)学习者”这一新的研究主题检索条件,则仅有4篇文献符合条件,而其中最新的研究也是杜家利博士团队的成果。不过,值得注意的是,这4篇结合中国学习者英语学习的“花园幽径句”研究都是实证类型论文,说明中国学者的研究在这方面已经有国际化的前沿意识。而从研究的系统性和深度来看,作者团队的这部专著无疑具有代表性。这种基于中国学习者语言学习特殊性展开的应用研究可谓对长期以来西方学者理论导向为中心的研究现状的积极反映,有很好的互补价值,值得肯定。

通读《中国学习者花园幽径句的量化研究》一书的书稿,我有三个方面的主要感受,现与大家分享。

首先,该研究的问题取向和实证路径结合紧密。众所周知,实证研究的难点首先在于研究问题的提出,其次则是面向理论或应用问题的实证研究设计。对此,作者们在该书的引言部分开门见山,鲜明地提出了他们的研究问题。更为重要的是,他们提问的方式并非纯粹自

上而下的理论追问，而是结合了自下而上的问题思考。这样一种独特的提问方式，我想是与杜家利博士作为一个英语学习者、教师和研究者的三重角色以及长期较为丰富的经验积累分不开的。事实上，在国内为数不多的“花园幽径句”研究者中，杜家利博士是持续研究且成果颇丰的一位。正如他在该书引言中提出的几个问题，无一例外都是从学习者的视角（“我们”）来思考“花园幽径句”解读的实际过程中的困难或障碍，并积极尝试借助语言统计和语言计算的方法来开展有针对性的实证探究。其研究目标也非常明确，即“从计算语言学视角为中国英语学习者解读复杂结构句提供理论支撑，并为英语教学提供具有统计学意义的数据”。显然，这项研究是非常接地气的，有应用研究的现实意义与实用价值。从该书全篇的实证分析与研究思路设计来看，作者们所展开的三大块主体研究内容，即“被试成绩的偏度与相关系数研究（第四章）、花园幽径句阅读时间与效果的非独立t值研究（第五章）以及时间与能力因素与花园幽径句阅读效果的方差分析（第六章）”，较好地对接了其在引言中所提出的有关中国学习者解读花园幽径句的现实问题。

其次，该研究的规范意识与创新意识融合紧密。该书的实证研究设计，从具体研究问题的提出，到三个既相互关联又彼此完整独立的研究内容设置，充分体现出作者实证研究的良好素养和扎实基础。具体到第四章到第六章的内容设计，作者首先对被试具有英语学习综合水平代表性的三类成绩（高考总成绩、高考英语成绩以及英语专业四级成绩）进行了偏度分析，研究了其正态分布情况。这一“被试成绩的偏度与相关系数研究”为后续两章涉及中国英语学习者具体的“花园幽径句”解读相关实证分析与计算研究提供了可靠的数据支撑，做好了逻辑铺垫。这一研究步骤，一方面体现了作者的实证规范意识，确保了该研究的信度和效度，增强了研究的说服力；另一方面，也体现出作者的创新意识。对于中国英语学习者相关的实证研究而言，通常研究者对于被试的二语语言能力的考量是比较单一的，而且大多会选择常规单一标准作为评价基础。而本研究中，作者从中国

英语学习者二语学习水平考量的特殊复杂性出发,基于学习者高考成绩、高考英语成绩和英语专业四级成绩的数据,充分思考了这三类数据之间的关联性及其对于后续计算研究的影响,预先做了数据分析与计算研究,这是本土化问题意识所带来的研究方法创新,提升了研究的原创适用性,值得借鉴。

最后,该研究的专业学术性与读者友好性互补紧密。就该书的专业学术性而言,全文所涉及的专业内容与表述规范,文献回顾翔实全面,条分缕析,重点突出,易于理解。对于“花园幽径现象”的理论分析层次分明,逻辑缜密,为后续的实证研究做好了理论准备。与此同时,难能可贵的是,该书作者在全文阐述的学术性基础上,还有着很好的读者意识,行文论述的可读性较其他计算语言学研究著述或许更平实易懂。我作为计算语言学研究的外行读者,通读全书的过程体验,远没有之前预想的艰涩难懂(当然,那些计算公式仍是有难度的,但大多并不太影响理解)。这在很大程度上与作者深入浅出的写作风格密不可分。比如,该书开篇引言的写法,从实例切入,直奔主题,然后引出作者和读者共同关切的问题与思考,使读者有比较好的代入感,促进了作者与读者之间的有效互动,有助于加深读者的理解与认同。再比如,第三章有关“花园幽径现象的理论分析”内容,前四小节都是理论分析的内容,相对专业性更强,对于外行读者可能有一定难度,但第五小节,作者安排了“基于实例的理论分析”内容,有效地降低了读者理解专业内容的认知负荷,增强了可理解性。这种学术写作风格方面的读者友好性对于普及专业理论知识和应用研究推广大有裨益。当然,能够深入浅出地介绍专业性较强的研究内容,这对作者本身也是比较高的要求,是其研究水平及深度外化的一种体现。

上述读后感式的三点阅读体验分享,从理论深度方面显然是不够的,但或许可以给其他读者提供一些阅读理解的参考,尤其是给不少跟我一样有外行“顾虑”的读者增添积极探索新知的勇气,从而使更多的学习者和研究者从这本书中获益。我相信这本书的价值是多方

面的，除了对计算语言学本体及应用研究是重要拓展和补充外，还会给国内外关注中国英语学习者语言学习的相关研究带来启发和思考，开阔视野抑或借鉴方法。正如我读此书一般，受益良多。

当然，对于这部新作，我作为读者有一点期待。具体一点说，这部著作主要是从计算语言学研究视角对中国英语学习者解读“花园幽径句”的认知机制进行了实证分析和研究，数据分析和计算研究见长，取得了积极的成果。但在此基础上，一方面，今后可否再对影响中国学习者解读“花园幽径句”的相关认知因素做数据基础上的深度探讨，可否结合二语习得研究的相关理论资源与研究成果进一步拓展其应用研究价值？另一方面，今后可否尝试以汉语的语言事实为研究素材，探究汉语“花园幽径句”的特殊性与复杂性，或许这方面的英汉对比研究会有更多有趣且有价值的发现？或许将该研究的被试调换为非母语的汉语学习者会产生新的理论增长点，提出中国学者独到的新发现和新见解？等等。我的这些想法，不揣浅陋，主要是希望不久的将来，能够看到作者在这个领域更多、更新、更好的研究成果。如此，即便我的上述想法不切实际或明显外行，想必作者也不会见怪吧。

杜家利博士师从我国计算语言学大家冯志伟先生，学术训练和积累丰厚，他本人非常谦逊低调，勤奋刻苦且勇于创新，值得我学习的地方很多。很感谢他邀我先睹其团队的新作为快，也正好借此一番“絮言”，表达我对作者们的祝贺和寄望。

是为序。

2018年6月25日于南京大学

目 录

第一章 引言	(1)
第二章 花园幽径现象研究综述	(8)
第一节 国外研究回顾	(10)
一 启动机制研究	(10)
二 诱发条件研究	(17)
三 检测方法研究	(22)
第二节 国内研究回顾	(25)
一 句法研究	(25)
二 语境研究	(27)
三 心理研究	(27)
本章小结	(29)
第三章 花园幽径现象的理论分析	(31)
第一节 花园幽径现象认知的先后性	(36)
第二节 花园幽径现象认知的条件性	(38)
第三节 花园幽径现象认知的螺旋性	(41)
第四节 花园幽径现象理解折返与顿悟的认知基础	(43)
一 认知顿悟的前提——语义触发	(44)
二 认知顿悟的结果——认知折返	(44)

三 认知顿悟的效应——跨越解码	(44)
第五节 基于实例的理论分析	(45)
本章小结	(52)
 第四章 被试成绩的偏度与相关系数研究	(54)
第一节 偏度	(54)
一 被试高考总成绩偏度分析	(58)
二 被试高考英语成绩偏度分析	(59)
三 被试英语专业四级成绩偏度分析	(59)
第二节 相关系数	(60)
一 简单皮尔逊积矩相关系数公式	(61)
二 相关系数的 EXCEL 函数计算	(63)
三 相关系数的临界值	(64)
第三节 被试高考总成绩与英语专业四级成绩的相关 系数研究	(66)
第四节 被试高考英语成绩与英语专业四级成绩的相关 系数研究	(69)
本章小结	(74)
 第五章 花园幽径句阅读时间与效果的非独立 t 值研究	(76)
第一节 测试前准备	(76)
第二节 非独立 t 检验	(78)
一 非独立 t 检验的公式测定法	(79)
二 非独立 t 检验的 EXCEL 函数测定法	(85)
第三节 非独立 t 值研究	(94)
一 非独立 t 检验下显著性差异测试句分析	(98)
二 非独立 t 检验下非显著性差异测试句分析	(105)
本章小结	(109)

第六章 时间和能力因素与花园幽径句阅读效果的

方差分析	(112)
第一节 方差分析公式法	(112)
第二节 时间因素与花园幽径句阅读效果的方差分析	(120)
第三节 英语能力与花园幽径句阅读效果的方差分析	(124)
一 英语专业四级成绩分组的卡方检验	(125)
二 英语能力与阅读效果的方差分析	(129)
本章小结	(196)
结语	(199)
参考文献	(203)
附录一 花园幽径句和对照句测试样例	(254)
附录二 阅读时间与效果的非独立 t 值测试	(260)
附录三 时间因素与测试结果的方差分析	(264)
附录四(一) 高考英语成绩与英语专业四级成绩相关 系数统计	(268)
附录四(二) 高考英语成绩与英语专业四级成绩相关 系数分析	(273)
附录五(一) 高考总成绩与英语专业四级成绩相关系数 统计	(274)
附录五(二) 高考总成绩与英语专业四级成绩相关系数 分析	(279)

附录六(一)	英语专业四级成绩分组的卡方检验	(280)
附录六(二)	英语专业四级不同成绩组测试结果的方差 分析	(281)
附录七	The Penn Treebank 词类标记集	(285)
附录八	Stanford parser 的依存关系代码与解析	(290)
附录九	中英文姓名对照表	(299)
后记		(305)

图 目 录

图 3 - 1	Kempen(1996)的句法分析器框架模型	(34)
图 3 - 2	Vtr + NP 节点模式	(35)
图 3 - 3	Vtr + NP + PP 节点模式	(35)
图 3 - 4	普通句认知解码的顺序流程图	(36)
图 3 - 5	花园幽径句认知解码的顺序流程图	(37)
图 3 - 6	花园幽径句认知解码的选择流程图	(38)
图 3 - 7	普通句解码流程图	(39)
图 3 - 8	花园幽径句解码流程图	(40)
图 3 - 9	花园幽径句系统解码完整程序	(41)
图 3 - 10	WHILE 型循环结构	(42)
图 3 - 11	UNTIL 型循环结构	(42)
图 5 - 1	正态分布曲线中的数值分布图	(86)
图 6 - 1	句补关系错位的花园幽径句依存图	(138)

表 目 录

表 1-1	普通句—复杂句—对照句的剖析对照表	(5)
表 4-1	均值、中位数和偏度关系列表	(55)
表 4-2	平均数的 EXCEL 函数计算列表	(55)
表 4-3	奇数中位数的 EXCEL 函数计算列表	(56)
表 4-4	偶数中位数的 EXCEL 函数计算列表	(56)
表 4-5	奇偶数组的偏度对比	(57)
表 4-6	奇偶数组标准差的 EXCEL 函数计算列表	(57)
表 4-7	高考总成绩偏度表	(58)
表 4-8	高考英语成绩偏度表	(59)
表 4-9	英语专业四级成绩偏度表	(60)
表 4-10	高考成绩 X 与英语专业四级成绩 Y 的取样相关 系数计算表	(62)
表 4-11	高考成绩 X 与英语专业四级成绩 Y 的取样相关 系数函数计算表	(63)
表 4-12	自由度($df=8$)的相关系数临界值量表	(65)
表 4-13	高考成绩和英语专业四级成绩数据列表	(66)
表 4-14	高考成绩和英语专业四级成绩的相关系数 计算与分析	(68)
表 4-15	高考成绩和换算成绩与英语专业四级成绩 数据列表	(69)

表 4-16	换算后高考成绩与英语专业四级低分成绩 对比表	(70)
表 4-17	高考成绩和英语专业四级成绩的相关系数 计算与分析	(73)
表 5-1	测试句 S1 的 5 秒和 10 秒反应时的评分及 差异值	(79)
表 5-2	测试句 S1 的 5 秒和 10 秒反应时的非独立 t 检验值	(84)
表 5-3	取样测算值的标准差计算表	(87)
表 5-4	测试句 S1 在 5 秒和 10 秒反应时的 TTEST 函数 概率	(91)
表 5-5	测试句 5 秒和 10 秒反应时的 t 值和 sig 值列表	(94)
表 5-6	反应时延长差异性影响阅读效果的测试句列表	(99)
表 5-7	反应时延长对阅读效果不明显的测试句列表	(105)
表 6-1	反应时 5 秒—7 秒—10 秒的 S1 阅读效果评分及 平方表	(115)
表 6-2	反应时分组与阅读效果的 ANOVA 分析列表	(121)
表 6-3	反应时分组与阅读效果 $p < .05$ 的 ANOVA 列表	(123)
表 6-4	被试的英语专业四级成绩列表	(126)
表 6-5	英语专业四级成绩分组的卡方检验	(128)
表 6-6	英语能力组与阅读效果评分及平方列表	(129)
表 6-7	能力分组和阅读效果的方差分析列表	(131)
表 6-8	能力分组与效果评分的 100 个测试句 F 值列表	(132)
表 6-9	能力分组与效果评分的显著性差异项 F 值列表	(134)
表 6-10	频数观察中的 plan 和 plan to 差异性列表	(147)
表 6-11	在 BNC 语料库中 plan 和 plan to 的频数卡方检验 ..	(147)
表 6-12	四个对照句的结构对比与显著性差异对比	(166)
表 6-13	嵌套结构引导词 that 对花园幽径句 S3 的系统 剖析对照	(176)

第一章

引 言

花园幽径句是局部歧义的复杂结构。其认知解码是“单车道单向通行”的独立排他模式。这种复杂结构存在语义流中途折返、另路通过的特殊语义短路现象。

例 1 - 1 The boy walked towards the door. 男孩朝门走过去。
该句是简单句，读者没有任何阅读障碍。利用系统进行自动剖析其结果也很正确。请见斯坦福解析器（Stanford parser）对该句的剖析结果（具体代码请见附录）：

Your query
The boy walked towards the door.
Tagging
The/DT
boy/NN
walked/VBD
towards/IN
the/DT
door/NN
. /.
Parse
(ROOT
 (S