



花园幽径句行进错位的 计算语言学研究

Computational Linguistics Research on Processing Breakdown
in Garden Path Sentences

杜家利 于屏方 著



商務印書館

国家社科基金后期资助项目

花园幽径句行进错位的 计算语言学研究

Computational Linguistics Research on Processing
Breakdown in Garden Path Sentences

杜家利 于屏方 著



商务印书馆
The Commercial Press

2015年·北京

图书在版编目(CIP)数据

花园幽径句行进错位的计算语言学研究/杜家利,
于屏方著.—北京:商务印书馆,2015
ISBN 978-7-100-11440-0

I.①花… II.①杜…②于… III.①计算语言
学—研究 IV.①H087

中国版本图书馆 CIP 数据核字(2015)第 155220 号

所有权利保留。

未经许可,不得以任何方式使用。

花园幽径句行进错位的计算语言学研究

杜家利 于屏方 著

商务印书馆出版

(北京王府井大街 36 号 邮政编码 100710)

商务印书馆发行

北京冠中印刷厂印刷

ISBN 978-7-100-11440-0

2015 年 10 月第 1 版

开本 787×1092 1/16

2015 年 10 月北京第 1 次印刷

印张 26 1/4

定价:62.00 元

国家社科基金后期资助项目 出版说明

后期资助项目是国家社科基金设立的一类重要项目，旨在鼓励广大社科研究者潜心治学，支持基础研究多出优秀成果。它是经过严格评审，从接近完成的科研成果中遴选立项的。为扩大后期资助项目的影响，更好地推动学术发展，促进成果转化，全国哲学社会科学规划办公室按照“统一设计、统一标识、统一版式、形成系列”的总体要求，组织出版国家社科基金后期资助项目成果。

全国哲学社会科学规划办公室

序 言

教育部语言文字应用研究所 冯志伟

杜家利副教授曾是我的博士生，多年来，他一直在研究花园幽径现象（Garden Path Phenomenon）。此前，他曾出版了《非对称性信息补偿假说——花园幽径模式的困惑商研究》一书，提出了“非对称性信息补偿假说”（Asymmetric Information Compensation Hypothesis），尝试利用“困惑商”（Confusion Quotient）这一概念来解释语言理解过程中出现的花园幽径现象。现在，他和于屏方教授合写的新著《花园幽径句行进错位的计算语言学研究》就要在商务印书馆出版了，我对他们表示热烈的祝贺。

本书的重点是研究花园幽径现象中的“行进错位”。那么，什么是花园幽径现象中的行进错位呢？

我这里举一个汉语的例子来说明。

我们知道，“小王研究鲁迅的文章发表了”这个汉语句子是一个花园幽径句。这个句子的前一个部分“小王研究鲁迅的文章”是有歧义的，这一部分可以分析为“小王/研究鲁迅的文章”，其中，“小王”是主语，“研究鲁迅的文章”是谓语，形成一个主谓结构；这个部分又可以分析为“小王研究鲁迅的/文章”，其中，“小王研究鲁迅的”是定语，“文章”是中心语，形成一个定中结构。语料统计发现，把整个句子的前一个部分理解为主谓结构比之于把它理解为定中结构更为优先，也就是说，大多数人倾向于把前一个部分理解为主谓结构，而不倾向于理解为定中结构。可是，当出现了后一个部分“发表了”的时候，就必须把前一个部分“小王研究鲁迅的文章”理解为定中结构，而不能把它理解为主谓结构，原来居于优先地位的理解被抛弃，而原来居于非优先地位的理解却

成为了正确的理解结果，整个句子的理解过程发生了转折，出现了原先预想不到的错位，这就是花园幽径句的行进错位。这种行进错位，在汉语和英语中并不罕见，在理论上和应用上都有研究价值。

关于花园幽径现象中的这种行进错位，国内外学者曾经从心理学和认知科学的角度进行过研究，有丰富的研究成果。杜家利和于屏方另辟蹊径，从计算语言学的角度进行研究，这是他们的创新之处。

在数据结构方面，他们采用良构子串表（Well Formed Substring Table，简称 WFST）来描述行进错位。在良构子串表中，每一个子串在结构上都是合格的，因而也都是良构的，但是这些良构子串形成的整个结构不一定是完全的，这些良构子串甚至不能结合为整个的结构，它们只是形成一个表（table），因此良构子串表可以表示完全结构，也可以表示不完全结构，还可以表示歧义结构。良构子串表能够把剖析过程中那些在局部上良构的中间结构保存下来，不至于因为它们不能形成完全结构而轻易地把它们抛弃，这样就可以有效地描述花园幽径现象中的行进错位。

在算法方面，他们采用递归转移网络（Recursive Transition Network，简称 RTN）来表达行进错位的动态过程。由于语言符号具有递归性，类别相同的语法结构会多次在语言中出现，在语言研究时可以把语法结构加以抽象化，用有限的语法结构和规则来描述无限的、千变万化的句子。递归转移网络正是根据语言符号的这种递归特性研制出来的。在递归转移网络中，主网络可以分解为若干个附属于它的子网络，句子的剖析要经过主网络和子网络之间多次的下推（PUSH）和上托（POP）操作，往往下推了还要再下推，上托了还要再上托，这样就可以把花园幽径现象的行进错位过程生动地、具体地表示出来。

他们还使用 Stanford Parser^① 进行自动句法剖析。Stanford Parser 是美国 Stanford 大学开发的一个自动剖析器，可以用来做短语结构分析和依存关系分析，打开网址之后，点击 try out our parser online，就可以进行自动句法剖析了。Stanford Parser 是用 Java 实现的，它把优化的概率上下文无关语法（Probabilistic Context Free Grammar，简称 PCFG）剖析器、词汇化的依存语法剖析器（lexicalized dependency parser）和词汇化的概率上下文无关语法剖析器（lexicalized PCFG parser）结合在一起。Stanford Parser

^① <http://nlp.stanford.edu/software/lex-parser.shtml>.

还开发了图形界面 (GUI)，可以让用户直接看到剖析输出的短语结构树 (phrase structure tree)。Stanford Parser 大大减轻了两位作者的程序设计工作量，使得他们可以集中精力来研究花园幽径现象行进错位中的理论问题。对于语言学背景的计算语言学研究来说，这不失为一个非常明智的做法。

本书是杜家利博士出版的第三部关于花园幽径现象的专著。现在，于屏方博士也加入了花园幽径现象的研究。他们的研究力量越来越强大了，我希望他们再接再厉，在已经取得的成果的基础之上更上一层楼。

2015 年 4 月 1 日

杭州钱塘江畔

目 录

图目录	I
表目录	IV
例目录	VI
引 言	1
第一章 花园幽径现象研究综述	9
第一节 花园幽径现象国外研究综述	9
第二节 花园幽径现象国内研究综述	52
小 结	62
第二章 花园幽径效应图式研究	63
第一节 花园幽径效应图式模型	63
第二节 语音及词汇变异引发的花园幽径效应分析	68
第三节 句法变异引发的花园幽径效应分析	75
第四节 认知错位的花园幽径效应分析	78
第五节 语义错位的花园幽径效应分析	80
小 结	86
第三章 花园幽径现象理解折返性和顿悟性研究	89
第一节 花园幽径现象理解折返的结构分析	90
第二节 花园幽径现象错位性顿悟研究	102
第三节 花园幽径模式行进错位的认知基础	136
小 结	164
第四章 花园幽径现象层级存在性及解码策略研究	165
第一节 花园幽径现象的层级存在性	165
第二节 花园幽径现象的解码策略研究	186
小 结	206

第五章 花园幽径行进错位的计算语言学研究	207
第一节 递归转移网络的应用效应研究	207
第二节 良构子串表的应用效应研究	225
第三节 花园幽径模式解码系统的理论构建与应用	232
小 结	235
第六章 花园幽径句测试及结果分析	237
第一节 花园幽径句测试	237
第二节 限定反应时花园幽径句测试研究	237
第三节 花园幽径句行进错位效应分析	252
小 结	264
结 语	267
参考文献	271
附录一: The Penn Treebank 词类标记集	297
附录二: Stanford Parser 的依存关系代码与解析	301
附录三: “She told me a little white lie will come back to haunt me” 算法程序	309
附录四: 花园幽径句和对照句测试样例	337
附录五: 严式花园幽径句错位效应系统测试和卡方值表	341
附录六: 宽式花园幽径句及对照句的卡方值表	371
附录七: 外国人名中英文对照表	389
后 记	401

图目录

图 1	例 1 普通句依存关系图	2
图 2	例 2 行进错位的依存关系图	4
图 3	例 2 花园幽径效应依存关系图	4
图 4	例 3 行进错位依存关系图	16
图 5	例 3 花园幽径效应依存关系图	16
图 6	例 7 成功解码的树形图	21
图 7	例 7 解码的递归转移网络	21
图 8	例 7 行进错位的非良构子串表	25
图 9	例 7 花园幽径效应的良构子串表	26
图 10	例 14 树形图	30
图 11	例 14 的递归转移网络	31
图 12	例 14 行进错位的非良构子串表	35
图 13	例 14 花园幽径效应良构子串表	36
图 14	例 18 的树形图	41
图 15	例 18 的递归转移网络	41
图 16	例 18 行进错位的非良构子串表	44
图 17	例 18 花园幽径效应良构子串表	45
图 18	花园幽径效应图式	63
图 19	例 32 的依存关系图	66
图 20	例 33 的依存关系图	67
图 21	例 37 依存关系图	69
图 22	例 38 依存关系图	69
图 23	例 39 依存关系图	70
图 24	例 40 依存关系图	71
图 25	例 41 依存关系图	72

图 26	例 42 行进错位的依存关系图	72
图 27	例 43 行进错位的依存关系图	73
图 28	例 42 花园幽径效应依存关系图	74
图 29	例 43 花园幽径效应依存关系图	74
图 30	例 44 行进错位的依存关系图	76
图 31	例 45 的依存关系图	77
图 32	例 46 行进错位的依存关系图	79
图 33	例 46 花园幽径效应的依存关系图	80
图 34	词集合结构	91
图 35	语法线性结构	93
图 36	歧义图状结构	95
图 37	“Failing student looked hard” 的数据结构	98
图 38	理解折返的树形结构	98
图 39	“The new record the song” 的数据结构	100
图 40	例 63 行进错位的依存关系图	104
图 41	例 63 raced 替换后的依存关系图	105
图 42	例 63 花园幽径效应的依存关系图	107
图 43	例 66 的树形图	113
图 44	例 66 的递归转移网络	114
图 45	例 66 行进错位的非良构子串表	116
图 46	例 66 花园幽径效应的良构子串表	124
图 47	循环算法样例图	142
图 48	错位纠偏的认知折返图	146
图 49	基于 ERP 波形的错位效应对照图	148
图 50	多模式 ERP 波形对照图	160
图 51	ERP 测试电极分布图	161
图 52	多模式脑电地形图	163
图 53	引导词补足后的依存关系图	199
图 54	词类更迭导致的行进错位效应的依存关系对照图	204
图 55	递归转移网络样例流程图	208
图 56	汉语示例的递归转移网络图	211
图 57	汉语花园幽径句的下推和上托图	214

图 58	行进错位模式的树形图	220
图 59	汉语样例的非良构子串表图示	230
图 60	汉语样例的良构子串表图示	232
图 61	花园幽径模式的解码系统模型	233
图 62	严式花园幽径句卡方值对比图	245
图 63	非严式花园幽径句卡方值对比图	250

表目录

表 1	例 7 行进错位的算法矩阵	24
表 2	例 7 花园幽径句的算法矩阵	25
表 3	例 14 行进错位的算法矩阵	34
表 4	例 14 花园幽径句的算法矩阵	35
表 5	例 18 行进错位的算法矩阵	43
表 6	例 18 花园幽径句的算法矩阵	44
表 7	Failing(adj) 与 Hard 的歧义搭配	96
表 8	Failing(Grd) 与 Hard 的歧义搭配	97
表 9	Failing 与 hard 的歧义搭配	97
表 10	例 63 行进错位和花园幽径效应的依存关系对比表	107
表 11	例 64 行进错位和花园幽径效应的依存关系对比表	109
表 12	例 66 行进错位的算法矩阵	115
表 13	例 66 花园幽径效应的算法矩阵	124
表 14	感知认知顿误的 NS 流程简图	140
表 15	循环算法示例	141
表 16	UNTIL 型 NS 流程解码简图	188
表 17	从句插入前后的词类标注和依存关系对照表	197
表 18	词类更迭错位的 UNTIL 循环算法	205
表 19	良构子串表的数字符号排列矩阵表	226
表 20	底层终极符号标记展示	228
表 21	自底向上二级符号标记展示	228
表 22	自底向上三级符号标记展示	229
表 23	自底向上四级符号标记展示	229
表 24	汉语样例的 UNTIL 型循环算法	230
表 25	行进错位效应中三级符号标记展示	231

表 26	行进错位效应中四级符号标记展示	231
表 27	行进错位效应中五级符号标记展示	232
表 28	严式花园幽径句错位效应卡方值表	242
表 29	严式花园幽径句卡方值排序表	244
表 30	非严式花园幽径句错位效应卡方值表	246
表 31	非严式花园幽径句卡方值排序表	249
表 32	人机解码行进错位效应的交叉对照表	251
表 33	错位强度分类及频数分布表	252
表 34	错位类别卡方值表	254
表 35	兼类词在 BNCweb 语料库中的频数对照表	255
表 36	兼类词不同义项的卡方值表	255
表 37	floated 与 curled 的同形频数分布	259
表 38	floated 与 curled 的同形词卡方值表	259
表 39	S44 人工解码的卡方检验	261
表 40	S59 人工解码的卡方检验	262
表 41	S45 人工解码的卡方检验	262
表 42	S4 人工解码的卡方检验	262
表 43	动词 eat 后续成分 (+1) 的及物与不及物频数统计表	263
表 44	动词 eat 的及物与不及物分类的卡方值表	263
表 45	动词 hunt 的及物与不及物分类的卡方值表	263

例目录

例 1: The horse raced past the barn.	1
例 2: The horse raced past the barn fell.	2
例 3: The poor record the music.	12
例 4: I confirm that, as agreed, I have arranged for your Mr R. Simpson to deal <i>with</i> any matters arising in connection <i>with</i> the above during my absence on leave from 7-21 August 1993.	17
例 5: I have arranged for Mr R. Simpson to deal with any complaints about National Health Service and the matters arising in connection with the problem of homelessness in the city during my absence on leave.	17
例 6: I have arranged for Mr R. Simpson to deal with any complaints about National Health Service and the matters arising in connection with the problem of homelessness in the city must be discussed by the board meeting during my absence on leave.	17
例 7: The tycoon sold the offshore oil tracts for a lot of money wanted to kill JR.	18
例 8: The data was the result of active processes of project.	27
例 9: The data the scholars investigated was the result of active processes of project.	27
例 10: The data the scholars the police chased investigated was the result of active processes of project.	27
例 11: The data the scholars the police the journalist interviewed chased investigated was the result of active processes of project.	27
例 12: The data the scholars the police the journalist the children	

respected interviewed chased investigated was the result of active processes of project.	27
例 13: The raft floated down the river sank.	28
例 14: She told me a little white lie will come back to haunt me.	28
例 15: The logs are trimmed and then floated down the river.	37
例 16: The raft that was floated down the river sank.	37
例 17: She told me that a little white lie will come back to haunt me. ...	37
例 18: Until the police arrest the drug dealers control the street.	37
例 19: Until the police make the arrest, the drug dealers control the street.	37
例 20: The professors instructed about the exam were confused.	38
例 21: The waiter served in a new restaurant was happy.	38
例 22: Until the police arrest the baby control the street.	38
例 23: The girl told the story cried.	46
例 24: I know the words to that song about the queen don't rhyme.	46
例 25: Who could the little child have forced us to sing those stupid French songs for, last Christmas?	48
例 26: Who could have forced us to sing those stupid French songs last Christmas?	48
例 27: Who could the little child have forced to sing those stupid French songs last Christmas?	48
例 28: For whom could the little child have forced us to sing those stupid French songs last Christmas?	48
例 29: Someone shot the brother of the actress who was on the balcony.	49
例 30: John read the letter to Mary.	49
例 31: John read the note, the memo and the letter to Mary.	49
例 32: We knew that when that snake bit you, you would die.	64
例 33: We knew that when you bit that snake, it would die.	66
例 34: While the boy scratched the dog yawned loudly.	67
例 35: While the boy scratched, the dog yawned loudly.	68
例 36: The sky is blue and the /sʌn/ is bright for he really scores a	

notable success in an international tournament game.	68
例 37: The deer train the dogs.	68
例 38: The dogs train the deer.	69
例 39: The train dogs the deer.	70
例 40: That deer dogs the train.	70
例 41: The old dog dogs the deer.	71
例 42: The old dog the deer.	72
例 43: The strong train the weak.	73
例 44: The experienced soldiers warned about the dangers conducted the midnight raid.	75
例 45: The experienced soldiers warned about the dangers before the mid- night raid.	76
例 46: The woman brought the flowers smiled broadly.	78
例 47: The dog walked along the bank was barking.	80
例 48: The waiter served a steak enjoyed it immensely.	82
例 49: Time flies like an arrow.	84
例 50: Time flies like an arrow; fruit flies like a banana.	84
例 51: The new singers the song.	91
例 52: The old women the boat.	92
例 53: The building window the sun.	92
例 54: The new singers sing the song.	93
例 55: The old women sail the boat.	94
例 56: The building window reflects the sun.	94
例 57: Failing student looked hard.	95
例 58: Failing(adj) student looked hard(adj).	95
例 59: Failing(adj) student looked hard(adv).	95
例 60: Failing(Grd) student looked hard(adj).	95
例 61: Failing(Grd) student looked hard(adv).	95
例 62: The new record the song.	99
例 63: The boy raced to hospital was the son of chairman.	102
例 64: The boat floated down the river sank.	107
例 65: Fat boys eat accumulates.	110