

从依存距离最小化看英语花园路径句暂时性歧义的形成

李紫嫣

四川大学, 四川 成都

收稿日期: 2022年10月10日; 录用日期: 2022年11月21日; 发布日期: 2022年11月29日

摘要

本文基于依存语法的框架, 从依存距离最小化的角度探索英语花园路径句暂时性歧义的形成原因。利用已有的实验材料, 运用计量的研究方法, 将依存距离作为测量指标测算了花园路径句中产生暂时性歧义的部分和完整句子的平均依存距离。研究发现, 暂时性歧义部分的平均依存距离通常小于整个句子的平均依存距离, 证明人们在加工句子的过程中, 总是倾向选择依存距离较短的句法结构, 直到发现解读无法进行下去时, 再重新进行分析, 从而走入花园路径。研究结果再一次验证了依存距离最小化是人类语言的普遍规律, 这种认知机制能有效减轻工作记忆负担, 是“省力原则”的重要体现。

关键词

依存距离, 花园路径句, 暂时性歧义, 依存语法

The Formation of Temporary Ambiguity in English Garden Path Sentences from the Perspective of Dependency Distance Minimization

Ziyan Li

Sichuan University, Chengdu Sichuan

Received: Oct. 10th, 2022; accepted: Nov. 21st, 2022; published: Nov. 29th, 2022

Abstract

Based on the framework of dependency grammar, this paper explores the reasons for the forma-

tion of temporary ambiguity in English garden path sentences from the perspective of dependency distance minimization. Using the available experimental materials, the average dependency distances of the parts of garden path sentences that produce temporary ambiguity and complete sentences were measured using the econometric research method with dependency distance as a measure. The average dependency distance of the temporarily ambiguous part is usually smaller than the average dependency distance of the whole sentence, which proves that people always tend to choose the syntactic structure with shorter dependency distance in processing the sentence, until they find that the interpretation cannot proceed, and then re-analyze it, thus going into the garden path. The results confirm that minimizing the dependency distance minimization (DDM) is a universal rule of human language, and this cognitive mechanism can effectively reduce the burden of working memory, which is an important manifestation of “the principle of least effort”.

Keywords

Dependency Distance, Garden Path Sentence, Temporary Ambiguity, Dependency Grammar

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

花园路径句(Garden Path Sentence, 简称 GP sentence)是一种暂时性歧义现象, 在 20 世纪 70 年代初期由心理学家 T. G. Bever [1]首次提出。“花园路径是故意违反人们已经习惯的语法、语义和逻辑知识, 在实时加工过程中, 句子中某个部分会产生局部歧义, 人们偏向于对歧义句进行最容易接受的解读, 然而那种人们不愿意接受的解释正式句子的正确解读”(晏小琴, 2008) [2]。花园路径句的加工过程和结果对于认识人脑的认知方式、研究人工智能和语言习得以及外语教学实践等方面都具有重要的意义, 一直是语言学和心理语言学的热点话题之一(顾琦一, 程秀莘, 2010) [3]。

花园路径句的研究始于上世纪 70 年代, 彼时的“花园路径现象”讨论多是从心理语言学的角度出发, 这一类研究的基本出发点都是为了发现并描述大脑处理“花园路径”句法的策略。但单从心理语言学的角度开展探讨具有很大的局限性, 因此从 80 年代末起, 越来越多的研究者试图借助对语言结构本身的分析来寻求花园路径现象的答案(蒋祖康, 2000) [4]。近年来, 花园路径句的研究多与实证研究相结合。黄碧蓉(2007) [5]的研究发现“花园路径现象”可以在关联理论框架下得到解释, 幽默话语“花园路径现象”产生过程遵循最佳关联原则。Rah 和 Adone (2010) [6]的研究考察了德语为母语的中高级英语学习者, 发现语言水平与花园路径句的分析机制无关。顾琦一和程秀莘(2010)通过对花园路径句主句和从句理解测试、工作记忆测试和语言能力测试发现中国英语学习者花园路径句理解上存在消歧与曲解并存的现象。卢华萍和吴明军(2021) [7]采用了自定步速阅读任务, 考察了名词短语/内嵌子句(NP/S)和名词短语/零形式句(NP/Z)两种不同句法结构对于中国学习者二语花园路径句重新分析的影响, 结果发现重新分析的难度与句法结构相关。

Frazier (1987) [8]为分析花园路径模型, 提出了两个重要原则: 迟关闭原则和最小附加原则。迟关闭原则是指在合乎语法的情况下, 人们总是习惯将新的语言材料加到刚刚已经假定的子句或者短语结构上, 再去理解新的信息。这个原则与 Kimball [9]提出的早关闭原则相对, 早关闭原则意为除非下一个节点是该短语的直接成分, 否则句法结构会尽早关闭(Kimball, 1973)。最小附加原则即句子加工过程中并未假定

任何潜在节点,因此分析花园路径句时,句法分析器总是倾向于最简单最可能的结构。这两个模型针对性地解决了花园路径句句法分析常常需要回溯的难题,目的是实现花园路径句的高效分析(赵怪怡,刘海涛,2014) [10]。

那么关于花园路径句需要重新分析的特点,是否存在另一种解释呢?本文以依存句法为理论基础,以依存距离为衡量标准,提出假设“花园路径句暂时性歧义的形成与人类语言依存距离最小化倾向有关”,下文将通过实验材料对这一假设进行论证。

2. 方法

依存语法是语言学家泰尼埃(Tesnière)提出的描述词间关系的语法。依存语法中存在两个重要概念,其一是依存关系,即支配词和被支配词之间存在一种有向的、非对称的关系,这种关系通常用带箭头的弧线来标记,弧的起点是支配词,终点是被支配词;其二是依存距离,根据依存语法,两个有依存关系的词在句中的线性距离称为依存距离(刘海涛,林燕妮,2018) [11]。也就是一个句子中存在依存关系的两个词之间的位置差。依存距离的正负可以显示支配词和被支配词的位置,当距离为正时,支配词在被支配词之前,距离为负时,支配词在被支配词之后(李辉,刘海涛,2017) [12]。

依存距离这一概念(Dependency Distance,简称DD)由Heringer等人在1980年首次提出,被刘海涛(2009) [13]定义为“支配词与被支配词之间的线性距离”。它可以有效地测量句法复杂度以及语言加工的难度(王琳,2014) [14],被作为语言复杂性的量度标准(Liu,2008) [15]。

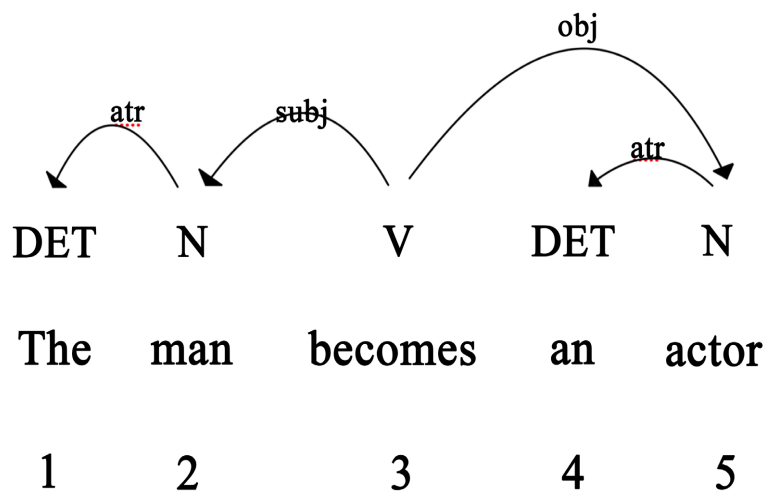


Figure 1. The dependency relation in the sentence “The man becomes an actor”
图 1. 句子 “The man becomes an actor” 的依存关系

图 1 中,带箭头的弧线代表两个词的依存关系,弧线的起点是支配词,终点是被支配词。单词下方的数字是每一个词语在整个句子中的线性位置,用于计算依存距离。刘海涛(2009)提出了一种依存距离的计算方法,假设 $W_1 \dots W_i \dots W_n$ 是一个词串,在具有依存关系的两个词之间, W_a 是支配词, W_b 是被支配词,它们之间的依存关系可以用 $a-b$ 来计算,则相邻两词的依存距离为 1,当 a 大于 b 时,DD 为正数,表示支配词在被支配词之后;当 a 小于 b 时,DD 为负数,表示支配词在被支配词之前,但通常取绝对值。

平均依存距离(Mean Dependency Distance,简称MDD)是句中涉及的所有依存距离绝对值的平均值。它可以有效测量语言理解难度以及表现人类认知负荷,被应用于多种语言的测量(蒋跃,范璐,王余蓝,2021) [16],可被表示为图 2。

$$\text{MDD}(\text{the sentence}) = \frac{1}{n-1} \sum_{i=1}^{n-1} |DD_i|$$

Figure 2. The formula of MDD

图 2. 平均依存距离公式

此处 n 表示句子中词语的总数, $n-1$ 表示句子中依存关系总量, $|DD_i|$ 表示句子中第 i 个依存距离的绝对值。一个句子中, 只有一个单词即根节点没有支配词, 它的依存距离为 0。因此图 1 的例句 “The man becomes an actor” 中, 从左至右在单词下方标示的序号为 1, 2, 3, 4, 5。根据依存语法, “The” 和 “man” 之间, 前者为被支配词, 后者为支配词, 两者的依存距离为 “2 - 1 = 1”, 得出两词的依存距离是 1; “becomes” 和 “actor” 之间的依存距离为 “3 - 5 = -2”, 说明支配词 “becomes” 在被支配词 “actor” 之前, 取其绝对值, 得出依存距离为 2。整个句子的平均依存距离为: $(|2-1|+|3-2|+|5-4|+|3-5|)/4=1.25$ 。

现在已有大规模的实证研究表明人类在语言加工过程中的依存距离最小化倾向, 所谓最小化即 “自然语句中, 有依存关系的词在一个句子中靠的比较近” (梁君英, 刘海涛, 2016) [17]。麻省理工学院(MIT)大脑与认知科学系的弗特勒尔等人(2015) [18]对 37 种自然语言做了基于语料库的大数据分析后发现, 依存距离最小化时人类语言的普遍特征 Liu (2008) [15]。基于语料库考察了 20 种语言, 发现人类语言存在依存距离最小化趋向, 平均距离的阈值一般不超过 3 个单词。依存距离的最小化一方面与人类认知机制密不可分, 另一方面也与人类行为所遵循的 “省力原则” 息息相关。依存距离越大意味着词语被储存在工作记忆的时间越长, 相应给工作记忆造成的负担也越重, 出于 “省力” 的考虑, 人类会尽量避免在语言中使用可能导致的长距离依存关系(陆前, 刘海涛, 2016) [19]。

这些研究充分体现了语言研究是一个具有交叉学科特点的研究领域, 是多语言、大数据技术、语言的普遍规律及认知科学充分融合的学科(徐春山, 梁君英, 2015) [20]。

3. 结果与讨论

英语中比较典型的花园路径句从句法层面可以分为三类: 第一类是主动词/缩略关系型花园路径句(缩写成 MV/RR), Bever 于 1970 年提出的经典句子 “The horse raced past the barn fell” 就属于此类, 在这个句子中, raced 可以看作是 the horse 的谓语动词, 但加工到 fell 时会发现句子解读错误, 此时需要将 raced past the barn 看作 the horse 的修饰部分, 重新解读句子; 第二类是直宾/主语型花园路径句(缩写成 DO/S), 如: while the man hunted the deer ran into the woods, 这个句子的两种解读分别是将 the deer 作为 hunted 的直接宾语, 但会走入花园路径, 第二种将 the deer 作为句子主语才是正确的解读; 第三类是直宾/补语型花园路径句(缩写成 DO/SC), 如: Mary convinced her children were noisy (王嘉颖, 2012) [21]。在 were 之前, 这个句子可能有两种解读: 一种是将 her children 作为 convinced 的直接宾语, 一种是将 her children were noisy 整个从句作为补足语, 这时 her children 是从句里的主语。但对于第一种情况, 当句子加工到 were 时就无法继续进行, 因此需要重新解读, 形成了花园路径现象。

为了验证英语花园路径句暂时性歧义的形成可以用依存距离最小化来解释, 本文根据典型的英语花园路径句三大类别即包含主动词/缩略关系型花园路径句(缩写成 MV/RR)、直宾/主语型花园路径句(缩写成 DO/S)以及直宾/补语型花园路径句(缩写成 DO/SC), 每一类别收集了 10 个句子, 一共 30 个句子。这些句子主要是根据 Christianson 等人(2001)、顾琦一等人(2010)和杜家利等人(2016) [22]的实验材料挑选或改编的花园路径句。在依存语法的理论框架下, 运用计量的研究方法, 将依存距离作为测量指标测算了花园路径句中产生暂时性歧义的部分和完整句子的平均依存距离, 验证了花园路径句的理解体现了依存距离最小化倾向。

对这些花园路径句进行了依存句法分析，并计算了这些句子以及暂时性歧义部分各自的平均依存距离。以第一类主动词/缩略关系型的第一个句子为例，在“The boat floated down the river fell”这个句子中，可能存在两种解读：第一种是将 floated 作为主语 The boat 的谓语动词，直到加工到 fell 这个单词时才发现解读错误，因此 fell 前面的部分形成了花园路径式的暂时性歧义。对其暂时性歧义部分(图 3(a))和整个句子(图 3(b))进行依存句法标注和计算得出前者的平均依存距离为 1.2，后者为 1.83，也就是说暂时性歧义部分的平均依存距离小于整个句子的平均依存距离。

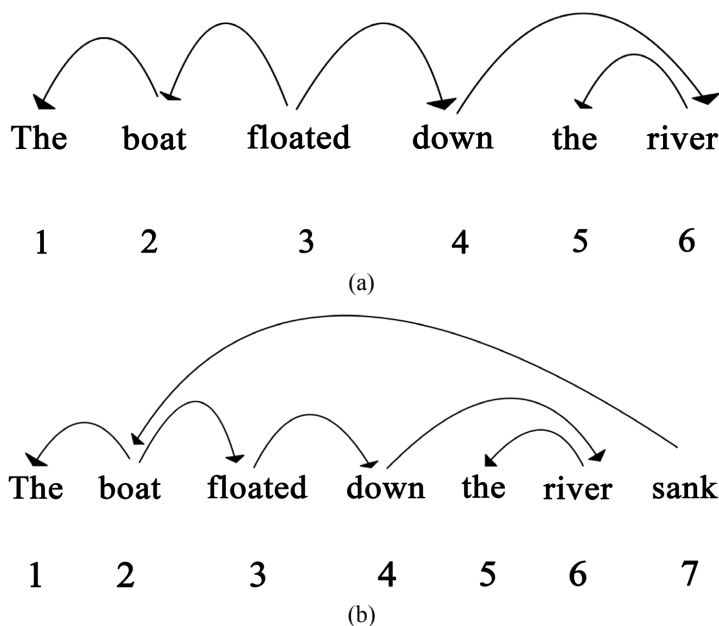


Figure 3. (a) The part of temporary ambiguity (MDD = 1.2); (b) the complete sentence (MDD = 1.83)
图 3. (a) 暂时性歧义部分(MDD = 1.2); (b) 完整句子(MDD = 1.83)

以同样的方法标注和计算 30 个句子，得到了其暂时性歧义部分(a)和完整句子(b)的平均依存距离，如下表所示。

MDD (a)	MDD (b)	第二类	MDD (a)	MDD (b)	第三类	MDD (a)	MDD (b)
1.2	1.83	1	1.6	2.1	1	1.33	1.4
1.2	1.83	2	1.6	2	2	1.66	1.66
1.25	1.8	3	1.6	2	3	1.8	1.75
1.2	1.71	4	1.33	1.8	4	1.86	1.62
1.71	2.1	5	1.6	2.17	5	1.33	1.38
1.2	1.71	6	1.5	1.75	6	1.33	1.5
1.2	1.83	7	1.8	2.27	7	1.33	1.4
1.25	1.6	8	1.6	2	8	1.33	1.5
1	1.6	9	1.6	1.89	9	1.33	1.5
1.25	1.71	10	1.83	2	10	1	1.25

在这 30 个句子中,有 90%的暂时性歧义部分的平均依存距离都小于整个句子的平均依存距离,说明暂时性歧义部分的加工难度低,更好理解。例外的 3 个句子为第三类直宾/补语型花园路径句(缩写成 DO/SC)的第 2、3、4 句。第 2 个句子和第 4 个句子中的暂时性歧义部分形成了主谓宾结构的简单句,这个简单句的主语和宾语成分前面存在较长的修饰词,加大了工作记忆的负担。第三个句子的暂时性歧义部分形成了一个“主语 + 谓语 + 直接宾语 + 间接宾语”的结构,间接宾语 dog 与谓语 gave 的线性距离较远,是产生较大依存距离的原因。

暂时性歧义部分的平均依存距离小于整句的平均依存距离原因可能如下:首先,暂时性歧义部分的字数长度一般少于整句,也就意味着需要加工的成分更少。其次,暂时性歧义部分的结构简单,一般是主谓宾结构,相对于整句复杂多样的结构来说,加工的复杂性更低。因此,暂时性歧义部分认知加工的难度小于整句。从人类语言的依存距离最小化倾向来看,人们倾向于选择容易加工的结构,于是走入了花园路径。

4. 结论

本文从语言学的角度,结合已有的花园路径句研究材料,对于花园路径句暂时性歧义的形成原因进行了探讨,证明了人们在进行花园路径句加工时,总是倾向于选择依存距离更小的暂时性歧义部分。本研究为花园路径句暂时性歧义的形成提供了一种解释,印证了依存距离最小化是人类语言的普遍规律,与短期工作记忆密切相关。

鉴于本文的研究语料数量较小,在选择花园路径句的结构类型上也有待进一步完善,仍需要跟专业的心理学实验加以证明,期待未来会产生更权威更详实的研究。

参考文献

- [1] Bever, T.G. (1970) The Cognitive Basis for Linguistic Structures. In: Hayes, J.R., Ed., *Cognition and the Development of Language*, John Wiley and Sons, New York, 279-352.
- [2] 晏小琴. 英语花园路径句加工的定性研究[J]. 外国语言文学, 2008(1): 39-45.
- [3] 顾琦一, 程秀苹. 中国英语学习者的花园路径句理解——与工作记忆容量和语言水平的相关研究[J]. 现代外语, 2010, 33(3): 297-304+330.
- [4] 蒋祖康. “花园路径现象”研究综述[J]. 外语教学与研究, 2000(4): 246-252+320.
- [5] 黄碧蓉. 幽默话语“花园路径现象”的关联论阐释[J]. 外语研究, 2007(6): 25-29.
- [6] Rah, A. and Adone, D. (2010) Processing of the Reduced Relative Clause versus Main Verb Ambiguity in L2 Learners at Different Proficiency Levels. *Studies in Second Language Acquisition*, 32, 79-109. <https://doi.org/10.1017/S027226310999026X>
- [7] 卢华萍, 吴明军. 不同句法结构对二语花园路径句重新分析的影响研究[J]. 现代外语, 2021(2): 233-245.
- [8] Frazier, L. (1987) Sentence Processing: A Tutorial Review. In: Coltheart, M., Ed., *Attention and Performance XII: The Psychology of Reading*, Erlbaum, Hillsdale, 559-586.
- [9] Kimball, J. (1973) Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, 2, 15-47. [https://doi.org/10.1016/0010-0277\(72\)90028-5](https://doi.org/10.1016/0010-0277(72)90028-5)
- [10] 赵怿怡, 刘海涛. 歧义结构理解中的依存距离最小化倾向[J]. 计算机工程与应用, 2014, 50(6): 7-10.
- [11] 刘海涛, 林燕妮. 大数据时代语言研究的方法和趋向[J]. 新疆师范大学学报(哲学社会科学版), 2018(1): 72-83.
- [12] 李辉, 刘海涛. 基于句法标注语料库的汉语儿童三词句习得研究[J]. 语言文字应用, 2017(1): 107-116.
- [13] 刘海涛. 依存语法的理论与实践[M]. 北京: 科学出版社, 2009.
- [14] 王琳. 汉英语码转换的句法变异问题探索——基于树库的动词句法配价分析[J]. 外语与外语教学, 2014(5): 47-53.
- [15] Liu, H.T. (2008) Dependency Distance as a Metric of Language Comprehension Difficulty. *Journal of Cognitive Science*, 9, 159-191. <https://doi.org/10.17791/jcs.2008.9.2.159>

-
- [16] 蒋跃, 范璐, 王余蓝. 基于依存树库的翻译语言句法特征研究[J]. 外语教学, 2021(3): 41-46.
- [17] 梁君英, 刘海涛. 语言学的交叉学科研究:语言普遍性、人类认知、大数据[J]. 浙江大学学报(人文社会科学版), 2016, 2(1): 108-118.
- [18] Futrell, R., Mahowald, K. and Gibson, E. (2015) Large-Scale Evidence of Dependency Length Minimization in 37 Languages. *Proceedings of the National Academy of Sciences*, **112**, 10336-10341.
<https://doi.org/10.1073/pnas.1502134112>
- [19] 陆前, 刘海涛. 依存距离分布有规律吗?[J]. 浙江大学学报(人文社会科学版), 2016, 2(4): 63-76.
- [20] 徐春山, 梁君英. 依存距离最小化是人类语言的普遍规律吗?[J]. 浙江大学学报(人文社会科学版), 2015, 1(6): 197-199.
- [21] 王嘉颖. 新世纪国外“花园幽径句”研究综述[J]. 绍兴文理学院学报(哲学社会科学), 2012(6): 71-75.
- [22] 杜家利, 于屏方. 中国英语学习者花园幽径句错位效应强度研究: 计算语言学视角[J]. 中文信息学报, 2016(6): 100-116.

附录

第一类:

- 1) The boat floated down the river sank.
- 2) The horse raced past the barn fell.
- 3) The girl told the story cried.
- 4) The statue stands in the park are rusty.
- 5) I told the girl the cat scratched Bill would help her.
- 6) The man returned to his house was happy.
- 7) The man pushed through the door fell.
- 8) The author wrote the novel was likely to be a best-seller.
- 9) The senator criticized on radio resigned.
- 10) The man caught the thief was a policeman.

第二类

- 1) While the man hunted the deer ran into the woods.
- 2) While the chef stirred the soup boiled over.
- 3) While the reporter photographed the rocket sat on the launch pad.
- 4) When Fred eats food gets thrown.
- 5) As the detective investigated the robbery occurred.
- 6) While Anna dressed the baby played in the crib.
- 7) Because he always jogs a mile seems a short distance to him.
- 8) As the professor lectured the students take notes.
- 9) While the clown juggled the balls fell on the ground.
- 10) After the secretary typed the incriminating report would be distributed to many people.

第三类

- 1) Mary convinced her children were noisy.
- 2) The old secretary doubted the incriminating report would be distributed to many people.
- 3) Mary gave the child the dog bit a cake.
- 4) The modern fashion writer read the famous magazine was having trouble with its finances.
- 5) John convinced his parents are interested in their children.
- 6) Susan knew her father hated her boyfriend.
- 7) Anna convinced her friends were unreliable.
- 8) Susan cheated her boyfriend was a policeman.
- 9) Jane concluded the paper made no sense.
- 10) Mary knew Jack loved her.