

句法距离和句法复杂性的测量

司富珍

(北京语言大学 语言科学院,北京 100083)

摘要:主要讨论与句法距离及句法复杂性有关的若干问题,涉及概念主要有句法长度、语法的简约性和复杂性等,基本观点是:影响句子复杂性和句法长度测量的因素有很多,而这些因素又可以大致归为两种大的类型:语法内部的因素(有的可能与普遍语法有关,有的则可能只与特定的具体语言有关)和语法外部的因素。在实际的交际中,则会有更多因素参与其间。

关键词:句法;长度;简约性;复杂性;长距离;结构依存原则

中图分类号:H314.3 **文献标识码:**A **文章编号:**1009—1076(2016)01—0098—05

一、引言

对于多数读者来说,“句法距离”应该算是一个比较新鲜的概念,“句法复杂性”则虽然常常被提及,却也很少被放在一个原则性或规则性的系统中测量和计算过。本文写作的主要目的之一正是要讨论与测量句法距离和句法复杂性相关的几个问题。讨论伊始先来看几个例子:

(1)(a)我不喜欢我喜欢的人喜欢我不喜欢的人,更不喜欢我不喜欢的人喜欢我喜欢的人,也不喜欢我喜欢的人不喜欢我喜欢的人。

(b)在乎你的我只在乎我在乎的你是否像在乎你的我在乎我在乎的你一样在乎在乎你的我。

(c)If if if if it is raining, then A is correct, then B is correct, then C is correct, then D is correct, then E is correct.

(d)The frightened child who the old woman who the rescue worker looked for had comforted greatly survived the crash.

以上中文例子来自微信语料,英文例子摘自国外语言学文献,都是真实的文本语料。一般读者对于这类句子的第一印象大概是它们不仅都“很长”而且还比较“复杂”,甚至还有一点点奇怪。而语言学

研究者们关心的问题则是:应该如何定义、评价或计算一个特定结构的“长度”和句法“复杂性”(或其对立面“简约性”)。哪些因素会影响句法结构的复杂性?

二、影响句法复杂性的若干因素

(一)结构长度因素

一个句法结构的复杂性有时可以通过其长度(即结构中所包含的词语或短语的数量多少)来得到测量。Derenzi & Vignolo (1962)和 Boller & Vignolo (1966)等人曾通过Token Test (TT)来对失语症患者进行观察和诊断,以下是其中一些例子:

(2)Touch +(one or more) adjectives + noun

(a)Touch the red circle.

(b)Touch the large red circle.

(c)Touch the beautiful large red circle.

(d)Touch the beautiful large red circle and the even more beautiful small blue circle.

.....

以上这组例子中动词“touch”的宾语都是受到一个或多个形容词定语修饰的名词性短语。研究显示,随着结构长度的增加,失语症者对结构理解的难度也增加,这一点应该不难理解。不过它的确清楚地证明

收稿日期:2015-10-09

基金项目:本项目受北京市共建项目“动态视角下的北京口语语法研究”资助。

作者简介:司富珍(1966—),女,山西交城人,文学博士,北京语言大学语言科学院外国语言学及应用语言学专业教授、博士生导师,伊犁师范学院兼职教授,奥地利维也纳大学汉学系访问教授,主要研究兴趣包括乔姆斯基的语言学和哲学思想、生成语法、形态学、语言理据、语言哲学、语言接触和文化接触等。

句法结构的长度与认知难度之间具有正相关性。

汉语语言学家们在探讨结构长度和复杂性之间的关系方面也有一些有意义的探索,例如下面的一些研究:

(3)方梅(1995)的语料:

(a)小王买的戒指。

(b)小王昨天买的戒指。

(c)小王昨天在镇上买的戒指。

(d)?小王昨天在镇上用奖金买的戒指。

(e)?小王昨天在镇上用奖金给女朋友买的戒指^①。

以上例子反映的是,至少对于某些特定的结构来说,随着句子长度的增加,句子可接性的程度会随之降低,即使内部结构相同或基本相同。

陆俭明(2014)对短语结构长度和其可接受性之间的关系作了专门的研究,以下是其中的一些代表性例子:

(4)陆俭明(2014)

(a)吃麦当劳

(b)*吃五道口西边的那个麦当劳

(5)(a)姐姐衣服都洗干净了。

(b)*姐姐脏得不像话的衣服都洗干净了。

(c)*姐姐满是油腻、脏得不像话的衣服都洗干净了。

以上例子说明的是,短语或句子结构的长度会影响到人们对于该短语或句子的认知,但长度当然并不是影响短语或句子认知的唯一因素。还有很多其他因素会影响到人们对句子的识解,比如结构的类型。换句话说,只看线性长度对于句子“复杂性”的认识还远远不够,有些结构可能虽然很长但却并不会造成理解上的困难,而有些较短的结构可能理解起来甚至比相应的较长的结构还要困难。

(6)(a) If if if if it is raining, then A is correct, then B is correct, then C is correct, then D is correct, then E is correct.

(b) If it is raining, then A is correct (X); If X is true, then B is correct (Y); If Y is true, then C is correct (Z); If Z is true, then D is correct (W); If W is true, then E is correct (Q).

(6b)是对(6a)的解释,它包含的词语的数量明显多于(6a),但理解的难度却明显小于(6a)。因此,正如乔姆斯基(1956, 2010 & 2014)所说,句法结构的距离和长度应该基于句法“结构”来进行测量。

有的时候从线性上看本来有先有后的两个词语

与另外一个同向的相关成分之间距离却可能是相等的。Hiroyuki Ura(2000)曾在多重特征核查的框架下讨论过很多日语中的等距现象。英语和汉语中也有类似的例子,以英语为例:

(7)(a) Where_k did you buy what t_k.

(b) What_k did you buy t_k where.

按照生成语法有关疑问词移位的解释,其中C携带有强的疑问特征,吸引距离最近的一个疑问词移动从而核查其强特征,“where”和“what”可以移位到同一C位置的事实说明二者与C位置之间是句法等距的,尽管在线性顺序上看,“what”居“where”之前。这再次说明,句法距离的测量本质上是基于结构的,而不是基于线性顺序的。英语中更多例子还有:

(8)(a) E^② is [sc John_k my best friend_j].

(b) John_k is [t_k my best friend_j].

(c) My best friend_j is [John t_j].

Hiroyuki Ura(2000)认为上面例子中的“John”和“my best friend”与T之间句法等距,其中T带有强的EPP特征,它吸引与它等距的两个候选成分之一移位至其指示语位置,核查其强特征,移位的结果就是(8b)或(8c)。

有时候,一个从线性顺序上看距离更远的成分从结构上看却更为接近其相关成分,下面是Chomsky(2014)举到过的例子:

(9)(a) The boys expect to meet each other.

(b) I wonder who the boys expect to meet each other.

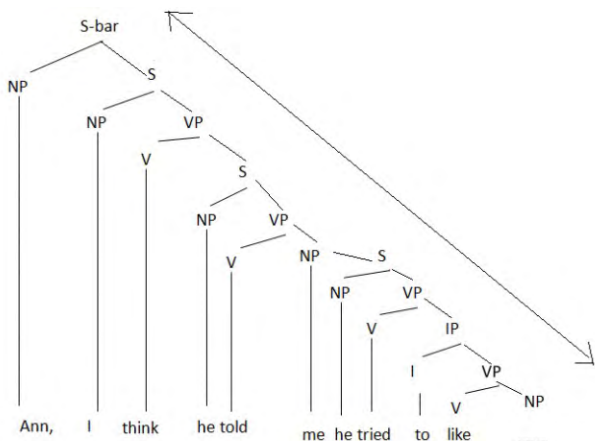
很显然,以上(9a)例中相互式反身代词“each other”回指的是“the boys”,然而这样的语义解读在(9b)中却不再成立:在(9b)中,“each other”回指的是“who”,而不是“the boys”,尽管从线性距离看“each other”与“the boys”的距离更近,然而从句法结构上看,“who”与“each other”却比“the boys”更近,当然这里面存在一个句法移位的问题,这是另外一个话题。总的来说,句法距离的测量是依存于结构类型的,相关原则被称作是“结构依存原则(Structure Dependency Principle)”(简称SDP)。所有的句法操作在语言的计算系统中都是依存于结构的(All syntactic operations are structurally dependent in computational system of language)。

(二)句法依存关系的跨度

决定句法复杂性的因素很多,存在依存关系的句法成分之间的距离跨度是其中之一。从树形图上

看,这种依存关系的跨度可以通过对诸如CP,IP,DP,NP等句法节点的计算来得到测量,下图箭头所指显示的是先行词“Ann”与其语迹“t_{NP}”之间的句法距离,其距离可以通过中间所跨越的句法投射节点的数量来进行计算:

(10) Ann, I think he told me he tried to like t_{NP}.



Gibson (1998 & 2000)曾提出一种依存局域性理论(Dependency Locality Theory,简称DLT),该理论认为,一个句子的复杂性会随句法依存跨度的增大而增加。按照Gibson(2000)的观点,一个句法结构的复杂性可以通过两个要素来预测:“存储代价”和“整合代价”。“存储代价”是将前面已经说过的词语存储在记忆中所需要花费的代价,而“整合代价”则是将这些词语按照某种关系整合、联系后存储于记忆中时需要付出的代价,较在跨度的依存关系花费的存储代价更大,也因此会增加说话人的认知难度。

(三)范域的宽度

句法距离有时不是通过句法节点来度量,而是以范域为单位进行测量。这一方面Hawkins有独到的见解,在他关于“最小化的范域(Minimized Domain)”的研究中有这样的论述:“要处理的任务所涉及的范域越大,其复杂程度就越高”。从类型学上来看,VO型语言偏好前置或后置式的移位手段,原因是通过将结构比较“重”^③的成分前置或后置处理,原有的结构就可以得到简化。这方面的例子在英语里比较多,人们倾向性地认为英语在句法结构上看是一种偏好后重结构的语言。英语的形式主语在很多情况下是为了填补后置的“重”型结构所留下的空位而存在的。如:

(11)(a) To express it properly in a public speech is a demanding task.

(b) It is a demanding task to express it properly in

a public speech.

对于说英语的人来说,(11b)比(11a)理解起来所担负的认知负担更小,因为它通过把比较“重”的不定式主语后置从而将主句的范域宽度有效地得到了缩小,从而降低了认知难度。

(四)范域的深度

语言中大量存在内嵌式结构(embedded structures),在多数情况下,内嵌结构的深度直接影响到句子解读的难易程度。比如下面这个Chomsky和Miller(1963)曾经提到的一个句子就属于这种内嵌深度造成的句法复杂性影响人们对它的理解的典型例子:

(12) The rate the cat the dog chased killed ate the malt. (Chomsky & Miller, 1963)

将这个句子拆分,把内嵌的结层剥离,每个层次单独放置后结果如下:

(13) The dog chased the cat, the cat killed the rate, and the rate ate the malt.

可以看到这个剥离分解后的新句子从线性上看其长度增加了,但理解和记忆起来负担却小多了。它再次证明句法复杂性并不是简单地由线性的长度所决定的。

从语用角度看,为了交际方便起见,人们通常总是尽量采用适当的话语策略来降低句法结构的复杂性。上面提到的英语将“重”型成分后置的策略,以及上例中将套嵌较深的结构浅层化处理的策略都是这方面的例子。各种语言由于各种语言结构的差异而可能会采用不同的策略来化简结构复杂性。根据陆俭明(2014)的研究,在汉语里,人们有时会通过停顿、添加语气词或插入“把”的方式来化简如下结构:

(14)(a) 弟弟杯子打了。

(b)? 弟弟那个很漂亮的杯子打了。

(c) 弟弟,那两个很漂亮的喝咖啡用的杯子打了。(停顿)

(d) 弟弟呀,那个很漂亮的杯子打了。(停顿加插入语气词)

(e) 弟弟把那两个很漂亮的喝咖啡用的杯子打了。(插入“把”)

(五)句法结构的类型

生成语法中最早讨论到句法复杂性的文献是Chomsky(1956 & 1965),Chomsky(1956)对形式语言进行了等级划分,这通常被称作“乔姆斯基等级(the Chomsky Hierarchy)”,而在Chomsky(1965)中,他则对具体的句法结构类型的复杂性程度进行了比较分

析,他特别提到了以下五种类型的结构,以及它们对于可接受性程度的影响:

(15)Chomsky(1965)提到的五种结构

(i)嵌套式结构(Nested constructions)

(ii)自嵌式结构(Self-embedded constructions)

(iii)多分支结构(Multiple branching constructions)

(iv)左分支结构(Left branching constructions)

(v)右分支结构(Right branching constructions)

这五种结构的复杂程度及其与可接受性之间的关系如下:

(16)五种结构的可接受程度(Chomsky, 1965)

(i)重复叠加使用嵌套式结构会导致句子不可接受;

(ii)自嵌式结构也更容易导致句子的不可接受;

(iii)多分支结构在可接受性上是可选结构;

(iv)将既长又复杂的成分嵌套会降低句子的可接受性;

(v)没有明显证据表明使用左分支结构或右分支结构会导致不可接受的例子。

在这五种类型的结构当中,对可接受性影响程度最小的是多分支结构(并列结构),本文前面提到例(13)和例(14)从嵌套式结构到多分支结构的变化所带来的可接受程度的提高就是这方面的典型例子,这方面例子有很多,由于篇幅所限这里不再多举例。Chomsky 和 Miller(1963)曾评论说:“嵌套式结构即使在没有自嵌式的情况下也会很快造成理解困难甚至使得结构完全不能理解……那些包含了嵌套式的依存成分或自嵌式结构的自然语言句子在超过一定程度后就可能变得连本族人也不可能理解了。(...nested [i.e. center-embedded] structures even without self-embedding quickly become difficult or impossible to understand. ...sentences of natural languages containing nested dependencies or self-embedding beyond a certain point should be impossible for (unaided) native speakers to understand.)

事实上本文开头所提到的几个例子(为方便读者这里将这几个例子重述如下)一般人都会觉得它们“很长”“很复杂”,甚至很费解,它们是合语法的句子,但从理解和认知的角度来看却又不是“正常”的句子,原因就在这里:

(17)(a)我不喜欢我喜欢的人喜欢我不喜欢的人,更不喜欢我不喜欢的人喜欢我喜欢的人,也不喜

欢我喜欢的人不喜欢我喜欢的人。

(b)在乎你的我只在乎我在乎的你是否像在乎你的我在乎我在乎的你一样在乎在乎你的我。

(c)If if if if it is raining, then A is correct, then B is correct, then C is correct, then D is correct, then E is correct.

(d)The frightened child who the old woman who the rescue worker looked for had comforted greatly survived the crash.

(六)其他因素

影响句法复杂性的因素还有很多,比如音韵方面的因素、短语或句子的语义类型、其他可能对记忆和解读造成负担的因素等。试看下面这个句子:

(18)全国政协邀请已故知名人士和党外全国政协委员夫人茶话迎春。

这是一则发表于官方网站(转引自《人民网》,原稿为中新社所发:<http://politics.people.com.cn/n/2015/0210/c70731-26542564.html>)的新闻稿中截取出来的例子,该标题从语法上来讲是没有任何问题和漏洞的,却成了语言学专业的学生开怀大笑的经典例子,原因就是标题所用短语中“夫人”一词的定语是一个并列结构,而这个并列结构中两个并列项长度有差别,前项为“已故知名人士”,后项为“党外全国政协委员”,再加上没有采用一般修饰结构都会使用的联结词“的”,就使得人们很容易把该句看作是“邀请”“已故知名人士”和“党外全国政协委员夫人”。这也是显示句法复杂性的一个很好的例子。

三、结语

总结起来,影响句子复杂性和句法长度测量的因素有很多,而这些因素又可以大致归为两种大的类型:语法内部的因素(有的可能与普遍语法有关,有的则可能只与特定的具体语言有关)和语法外部的因素。在实际的交际中,则会有更多因素参与其间。即以两个简单的对话人来说,如George von der Gabelentz所说,说话人希望表达时是“舒服的”(比如可能希望省力、经济等),而听话人则希望听到的是“清晰的”(即易于理解)。如乔姆斯基所说:“语言系统对于计算的有效性要求和思想表达来说可以说是经过了完好设计的,但对于语言使用特别是交际来说却可能造成种种问题……”就是说,所谓“简约性”与“复杂性”还存在一个视角问题,这方面可以探讨的地方还有很多。

注释:

①注意(d)(e)是具有歧解的结构,这里的分析

针对的是“的”居于动词“买”和宾语“戒指”之间的结构。

②E 此处意思指这是一个暂时由空语类(empty category)占据的位置。

③在语言学上人们常常把结构比较复杂或比较长的结构称作是“重”的结构。

参考文献:

[1]CHOMSKY N. The Logical Structure of Linguistic Theory[M]. Cambridge, MA: The MIT Press, 1955.

[2]CHOMSKY N. On the Poverty of Stimulus [C]. In the Proceedings of Glow-in-Asian-VIII. BLCU. Beijing: Beijing Language and Culture University Press. 2010.

[3]CHOMSKY N. Language Architecture and Its Import for Evolution[M]. Amsterdam: the Amsterdam Symposium. 2014.

[4]CHOMSKY N, MILLER G. Introduction to the formal analysis of natural languages [J]. in R. D. Luce, R.R. Bush, and E. Galanter (eds.), Handbook of Mathematical Psychology, vol 2. New York: Wiley, 1963 (2): 269-321.

[5]GIBSON E. Linguistic complexity: locality of syntactic dependencies [J]. Cogni-

tion, 1998, 68: 1-76.

[6]GIBSON E. The dependency locality theory: a distance-based theory of linguistic complexity[C]. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), Image, language, brain. Cambridge, MA: MIT Press. Oya, 2000: 95-126.

[7]HUDSON R. The Difficulty of (So-called) Self-embedded Structures[C]. UCL Working Papers in Linguistics 8, 1996.

[8]NEWMAYER F J, Preston L B. Measuring Grammatical Complexity [M]. Oxford University Press, 2014.

[9]方梅. 汉语对比焦点的句法表现手段[J]. 中国语文, 1995(4).

[10]陆俭明. 句法成分长度对句法结构的制约刍议[R]. 北京:北京语言大学语言科学院“走近语言科学前沿”第十四讲, 2014.

[11]司富珍. Syntactic Distance and Syntactic Complexity[C]//第三届“语言中的显著性和局部性”会议论文, 2014.

[12]萧国政. 句子的信息结构和汉语句法实体存活[J]. 世界汉语教学, 2001(4).

【责任编辑:何苗苗】