

链语法述评

冯志伟

提要 本文介绍链语法的基本原理和方法,并指出这种语法的词汇主义倾向。

Review on Link Grammar

Feng Zhiv ei

Abstract The basic theory and approach of link grammar are introduced in this paper. The author points out the lexicalism tendency of link grammar.

链语法(link grammar)是斯里特(D. Sleator)和汤佩雷(D. Temperley)于1991年在《用链语法剖析英语》(Parsing English with a Link Grammar, Technical Report of Carnegie Mellon University, CMU - CS- 91- 196)中提出的。这种语法的构思方式与范畴语法十分接近,对于单词的特性十分重视,带有强烈的词汇主义倾向。熟悉范畴语法的读者,不难领会链语法的原理。链语法便于语言工程的实现,是计算语言学中引人注目的一种新的语法理论。

一部链语法由一组词组成,语法中的每一个词都有一些特定的链接要求,这些链接要求被一一登录在链语法词典的相应词条里,根据这些链接要求对单词的链特性进行链接运算,便可以得出句子的结构。可以说,链语法是一种立足于单词的链接特性的语法。

单词的链接要求通过链来描述。链有两种:一种是链头,一种是链座。可以把链同电路相比拟,链头相当于电路中的插头,链座相当于电路中的插座。在链语法中,如果两个词要合法地链接,它们必须带有同一类的链,并且,一个词带链头,一个词带链座,链头应该恰如其分地插在链座中。例如,在句子“代表团昨天参观博物馆”中的单词“代表团”“昨天”“参观”“博物馆”可以通过连接子(connector)分别描述如下:

“代表团”!它的连接子为($()$, (s)),其中,左边的($()$)是链座,现在为空,右边的(s)是链头,是 s (主语)类的链,表示这个词要向右找一个链座为 s 的词相链接。

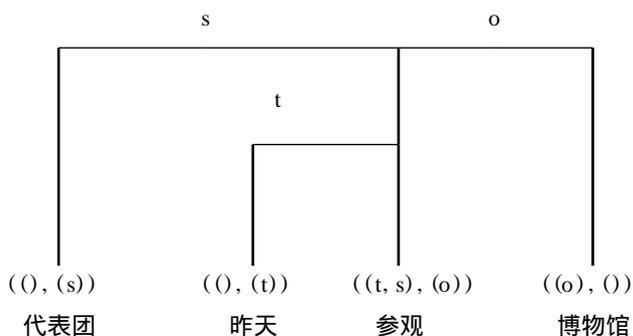
“昨天”!它的连接子为($()$, (t)),其中,左边的($()$)是链座,现在为空,右边的(t)是链头,是 t (时间词)类的链,表示这个词要向右找一个链座为 t 的词相链接。

“参观”!它的连接子为((t, s) , (o)),其中,左边的(t, s)是链座,按从后向前顺序分别为 t 和 s ,表示这个词要向左首先找一个链头为 t 的词相链接,接着再向左找一个链头为 s 的词相链接,右边的(o)是链头,是 o (宾语)类的链,表示这个词要向右找一个链座为 o 的词相链接。

“博物馆”!它的连接子为((o) , ($()$)),其中,左边的(o)是链座,表示这个词要向左找一个链

头为 o 的词相链接, 右边的 () 是链座, 现在为空。

如果一个连接子的链头能够插入类别和它相同的链座之中, 那么, 就说这个连接子的链接要求得到满足, 如果一个句子中的各个词的连接子的链接要求都得到满足, 那么, 链接这些词的一组链就叫做这个句子的一个链系统(linkage)。句子“代表团昨天参观博物馆”的链系统如下:



“代表团”的连接子要向右找一个链座为 s 的词相链接, “昨天”的连接子要向右找一个链座为 t 的词相链接, 而“参观”的连接子首先要向左找一个链头为 t 的词相链接, “昨天”的连接子特性正好满足这个条件, “参观”的连接子然后还要再向左找一个链头为 s 的词相链接, “代表团”的连接子正好满足这个条件,

因此, 可以把“参观”先同“昨天”链接起来, 然后再同“代表团”链接起来。“参观”的连接子还要求向右找一个链座为 o 的词相链接, 而“博物馆”的连接子正好满足这个条件, 于是, 最后把“参观”同“博物馆”链接起来, 造出句子的连锁。从上面的图中可以看出, 这个链系统包括 s, t 和 o 三条链, 每一条链的链头都正好插入链座之中, 完全满足链接的条件。

我们再以英语为例。在英语句子“The cat chased a snake”(那只猫追赶一条蛇)中的单词“the”“cat”“chased”“a”“snake”可以用连接子描述如下:

“the”: 它的连接子为 $((), (d))$, 其中, 左边的 $()$ 是链座, 现在为空, 右边的 (d) 是链头, 它表示这个词要向右找一个链座为 d 的词相链接。

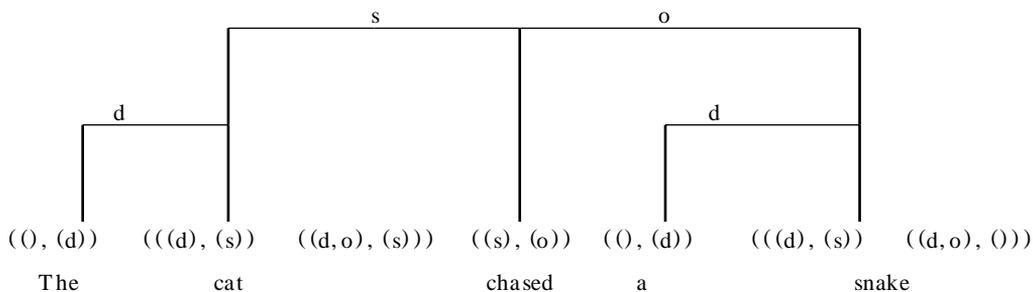
“cat”: 它的连接子为 $((d), (s)) \ ((d, o), (s))$, 这个连接子中的 $((d), (s))$ 以及 $((d, o), (s))$ 叫做“选言肢”, 这两个选言肢之间的关系是“逻辑或”(用 \vee 表示)的关系。选言肢 $((d), (s))$ 中, 左边的 (d) 是链座, 它表示这个词要向左找一个链头为 d 的词相链接, 右边的 (s) 是链头, 它表示这个词要向右找一个链座为 s 的词相链接; 选言肢 $((d, o), (s))$ 中, 左边的 (d, o) 是链座, 它表示这个词首先要向左找一个链头为 d 的词相链接, 然后再向左找一个链头为 o 的词相链接, 右边的 (s) 是链头, 它表示这个词要向右找一个链座为 s 的词相链接。

“chased”: 它的连接子为 $((s), (o))$, 其中, 左边的 (s) 是链座, 它表示这个词要向左找一个链头为 s 的词相链接, 右边的 (o) 是链头, 它表示这个词要向右找一个链座为 o 的词相链接。

“a”: 它的连接子为 $((), (d))$, 其中, 左边的 $()$ 是链座, 现在为空, 右边的 (d) 是链头, 它表示这个词要向右找一个链座为 d 的词相链接。

“snake”: 它的连接子由两个选言肢组成: $((d), (s)) \ ((d, o), (o))$ 。选言肢 $((d), (s))$ 中, 左边的 (d) 是链座, 它表示这个词要向左找一个链头为 d 的词相链接, 右边的 (s) 是链头, 它表示这个词要向右找一个链座为 s 的词相链接; 选言肢 $((d, o), (o))$ 中, 左边的 (d, o) 是链座, 它表示这个词首先要向左找一个链头为 d 的词相链接, 然后再向左找一个链头为 o 的词相链接, 右边的 (o) 是链头, 现在为空。根据单词的连接子的性质, 对它们进行链接, 我们可以得到句子“The cat chased a snake”的链接系统如下: (见下页)

在构成句子的链系统时, 对于具有两个或两个以上的选言肢的连接子, 由于连接子中选言肢之间是“逻辑或”关系, 能而且只能有其中的一个选言肢满足链接的条件。例如, 在 cat 的连



接子中, 只有((d), (s))这个选言肢能够满足条件从而同它左边的 the 和右边的 chased 链接起来; 在 snake 的连接子中, 只有((d, o), (0))这个选言肢能够满足条件从而先同它左边的 a 链接起来, 再同它距离更左的 chased 链接起来。

严格地说, 链语法是由一组词以及这些词相应的连接子组成的。而词的连接子是由一系列逻辑选言肢(即“逻辑或”)组成的。即有

$$w = (d_1 \ d_2 \ \dots \ d_k) \quad d = ((l_1, l_2, \dots, l_m), (r_n, r_{n-1}, \dots, r_1))$$

其中, $d = \{d_1, d_2, \dots, d_k\}$, (l_1, l_2, \dots, l_m) 称为左链, $(r_n, r_{n-1}, \dots, r_1)$ 称谓右链。带选言肢 d 的词 w 可以同处于该词两边的其他词相链接, 但是, 从左边相链接的词必须与 (l_1, l_2, \dots, l_m) 中的链相匹配, 不能有重复, 也不能有遗漏, 从右边相链接的词必须与 $(r_n, r_{n-1}, \dots, r_1)$ 中的链相匹配, 不能有重复, 也不能有遗漏。对于被采用的选言肢 $d = ((l_1, l_2, \dots, l_m), (r_n, r_{n-1}, \dots, r_1))$, 与 l_i 相链接的词和 w 的距离, 随着 i 的增加而增加, 与 r_j 相链接的词和 w 的距离, 随着 j 的减少而增加。句子的链系统应该满足如下四个条件:

1. 平面性(planarity): 在一个句子上面画出的词与词之间的链互不交叉。
2. 连接性(connectivity): 画出的链可无遗漏地把这个词序列中的所有词都链接起来。
3. 顺序性(ordering): 在一个选言肢的左链中从左到右排列的成分必须同它们分别要链接的在其左边的词从近到远的顺序相一致, 在一个选言肢的右链中从左到右排列的成分必须同它们分别要链接的在其右边的词从近到远的顺序相一致。
4. 排他性(exclusion): 一对词之间最多只能有一条链相链接, 也就是说, 不允许在同一对词之间出现一条以上的链。

像链语法这样的基于词的语法系统, 比起上下文无关的短语结构语法来, 更容易得到词间关系的统计数据, 如果通过大规模真实文本的语料库来获取每种链所链接的单词对的出现频度, 那么, 在句法系统遇到链接的歧义时, 便可以凭借这样的统计数据作出判断, 从而为句子的分析选出概率意义上的最佳结果。

链语法的语法信息主要记录在词典中的单词上, 依靠单词的语法信息来进行句法语义分析, 单词中记录的信息是整个链语法运算的基本资源, 这种语法特别强调词汇的作用。所以, 链语法是当代计算语言学中词汇主义倾向的具有代表性的一种语法理论, 应该引起我们的注意。

参考文献

[1] D. Sleator, D. Temperley, 1991, Parsing English with a Link Grammar, Technical Report of Carnegie Mellon University, CMU - CS- 91- 196

[1] J. Lafferty, 1998, The Link parser Application program Interface (API), <http://www.link.cs.cmu.edu/link> (冯志伟 国家语委语言文字应用研究所, 邮编: 100010)