

机器翻译基于从属关系的语义分析

傅爱平

(中国社会科学院语言研究所)

提要:在机器翻译源语分析中,句子的句法结构和语义结构之间存在着一种基于从属关系的对应。根据这种对应,本文指出了分析句子成分之间的语义关系和辨识词义的原则,并分别讨论了基于直接从属关系的分析方法和基于间接从属关系的分析方法。

Dependency – based Semantic Analysis in Machine Translation

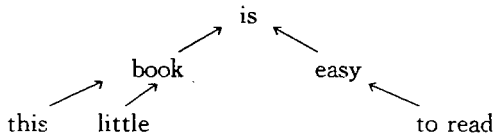
Fu Aiping

(Institute of Linguistics Chinese Academy of Social Sciences)

ABSTRACT: MT source language analysis reveals a close fit between syntactic and semantic structures which is based on dependency relations. On the basis of this fact, the principle of semantic structuring and lexical disambiguation are proposed, together with the processing approach on direct & indirect – dependency conditions.

一. 从属关系语法与机器翻译

从属关系语法(Dependency Grammar)是法国学者 Tesnière 于五十年代创立的一种语法理论。它认为,句子的基本关系是关联性,关联要服从层次原则,有层次的关联性称为从属性。在表示句子关联性的层级结构中,处于较高位置的成分(或范畴)叫做领位,处于较低位置的成分(或范畴)叫做属位。因此,句子中的关联性就表现为领位和属位之间的从属关系。一个句子可以有几个领位和属位,但动词总是扮演领位的角色,其他成分可以作属位,也可以作其下层成分的领位。这样,一个句子的句法结构就可以用从属关系构成的树形结构来表示。例如对于英语句子 This little book is easy to read. 有:



从属关系语法在自然语言处理的研究中得到了应用。以 JFY-5 机器翻译系统为例^①,它根据机器翻译实践的要求,在现有的众多语法理论当中各取所长,制定了语言自动分析的基本原则,从属关系语法是其主要的语言学理论基础。以下是 JFY-5 机器翻译系统在语言学方面的一些基本观点:

1) 在机器翻译的过程中,系统处理的对象在不同的加工阶段可以是词、短语或分句。它们是句子在不同层次上的组成成分。

2) 句子的一些成分与另一些成分之间具有关联性,它体现为支配关系和附加关系,统称为直接联系。支配关系是一种限定性的直接联系,造句时若不满足支配关系,将会造成不合语法的句子。附加关系是一种非限定性的直接联系,多来源于交际上的需要,附加成分不受中心成分的限制,造句时不影响句子在语法上的合法性。

3) 句子结构由成分之间的直接联系组成,表现为句法结构和语义结构两个方面。句法

结构是成分之间句法关系的总和,语义结构是成分之间语义关系的总和。同一句子的句法结构因语言不同而异,同一句子的语义结构对不同的语言则基本相同,因而语义结构是不同语言之间交际或翻译的媒介和基础。

4)直接联系具有方向性。处于支配或中心地位的成分是直接联系的轴心,称为上位成分;处于被支配或附加地位的成分称为下位成分。

5)根据直接联系的有向性,可以划分出成分之间的层级关系:上位成分、下位成分、同位成分和外位成分。一个上位成分可以支配若干个下位成分;而一个下位成分只能从属于一个上位成分。同位成分指属于同一轴心的所有直接联系成分;外位成分是具有准直接联系成分②。

6)谓语动词是句子结构的主轴心,其他形式的谓词是句子的次轴心。

显然,这里的“直接联系”即从属关系语法里的“从属关系”,“上位成分”和“下位成分”则分别对应“领位”和“属位”。

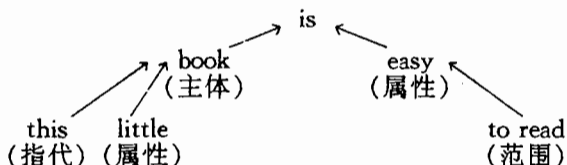
上述观点同样可以以英语句子 This little book is easy to read 为例,用上面的树形图表示。对 book 来说,它的上位成分是 is,下位成分 is this 和 little,同位成分是 easy,外位成分 is to read。

二.从属关系与机器翻译的语义分析③

机器翻译源语分析的任务是根据源语言句子的表层符号串识别句子的句法结构(Syntactic Structure),确定句子的语义结构(Semantic Structure),并且对多义词进行词义辨识(Lexical Disambiguation)。本文把识别句法结构的工作称为句法分析,把确定语义结构和进行词义辨识的工作称为语义分析。

从属关系语法是关于句法的理论,然而近些年已有学者应用从属关系的思想讨论语义结构(Hudson, 1991, p123 - 163),认为在句法结构和语义结构之间存在着一种平行的对应。即在句法上有从属关系的成分之间在语义上也有从属关系。这种建立在从属关系基础上的结构对应正好符合 JFY-5 机器翻译系统的观点(见第一节的基本观点 3),在 JFY-5 机器翻译系统中,句子里凡是在句法上具有从属关系的那些成分,在语义上也具有从属关系(反之不亦然)。即如果成分 A 在句法上从属于成分 B,那么 A 和 B 之间在语义上也有从属关系,而且该从属关系的方向总是与句法上从属关系的方向相同,即 A 的意义从属于 B 的意义。

根据这种观点,我们在构造句子的语义结构时,可以以句法上的从属关系为线索去分析成分之间的语义关系。也就是说,对于每个得到了句法关系属性的成分,依据同样的关联对象和方向确定它的语义关系属性。实际上在 JFY-5 机器翻译系统中,语义关系(施事、受事、领属、工具、时间等)是作为句法关系(支配、附加等)的补充赋予每个句子成分的。例如:



在上图中,句法上有从属关系的成分之间语义上也有从属关系,我们把这种情况称为直接从属关系(在第一节的基本观点 2 中称为直接联系)。反过来,语义上有从属关系的成分之间在句法上却不一定有从属关系。我们把语义上有从属关系而句法上没有从属关系的情况称为间接从属关系(在第一节的基本观点 5 中称为准直接联系),例如上图中的 book 和 read。间接从属关系反映源语分析的深度,是句子语义结构中不可缺少的部分,对于辨识多

义词也有重要的作用。在分析语义关系时,属于直接从属关系的可以以句法关系为线索,属于间接从属关系的,就要使用转换分析的方法了。

语义分析的另一个任务是多义词辨识。我们认为,象确定语义结构可以以句法上的从属关系为线索一样,多义词辨识也可以以语义上的从属关系为线索。

辨识多义词最常用的办法是找到与其相关联的词,根据它们在词义上的相互制约排除歧义。基于语义从属关系的方法使得“与其相关联的词”具有确切的定义:在语义上与其有直接或间接从属关系的词;同时也使“找到这些词”成为可操作的事情:按照语义结构的上位、下位、同位和外位关系查找。

例如英语名词短语“heavy(重) rain(雨)”,翻译成汉语应该是“大雨”。heavy 直接从属于 rain,在语义上是修饰关系。处于这种从属关系中的 heavy 在其上位成分 rain 的制约下,应选择词义“大”。从句子表层看,从下位成分 heavy 查找上位成分 rain 其实只是在找相邻的词,似乎不必兴师动众地借助语义关系,只用位置关系控制就可以了。然而,下面的例子告诉我们,在很多情况下,仅仅依靠位置关系辨识词义是不够的。英语动词短语“make(制造) mistake(错误)”翻译成汉语应该是“犯错误”,仅就这个短语形式来说,依靠位置关系完全可以确定 make 的词义是“犯”。可是对下面这些句子中同样的 make,位置关系就很难再起作用了:

1. Three mistakes have been made.
2. The mistake he made is not a serious one.
3. I don't know the mistakes made by the students.
4. He ordered no mistakes to be made in the work.
5. The mistake being made is a serious one.
6. The mistakes of the students in their homework, which they made yesterday, have been corrected.
7. This mistake is worth making.
8. This mistake is easy to make.

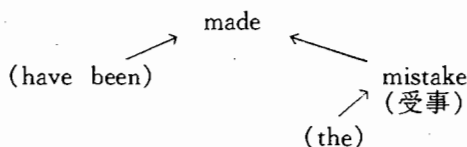
在这些句子里,make 和 mistake 之间虽然位置各不相同,但是在语义关系上,它们与 make mistake 是一致的,即, mistake 直接从属于 make(如例句 1),或间接从属于 make(如例句 2—8);在语义上都是对象关系。处于这种从属关系中的 make 在其下位成分 mistake 的制约下,应选择词义“犯”。我们认为,多义词之所以和与其词义有关联的词相互制约,正是因为它们之间存在语义上的从属关系。所以,以语义从属关系为线索查找在词义上相互制约的关联词,对于准确地辨识多义词是必要的。Grishman 在谈到如何在一个系统中定义并施行语义制约(Semantic Constraints)时,也指出在表层结构施行语义制约很麻烦,如果在句子的功能关系建立起来以后施行,就比较方便了(Grishman, 1986)。这个提法与我们的观点很接近。下面以 JFY-5 机器翻译系统为例,讨论语义分析的方法。

三. 基于直接从属关系的语义分析

1. 确定语义关系

直接从属关系对应句法上的支配关系和附加关系。我们认为,支配关系是由谓词(如英语中的动词、一部分名词和形容词)支配下位成分的能力决定的,因此可以把识别支配关系的规则作为词的用法规则(个性规则)放在词典中谓词的词条下面。支配成分和被支配成分之间的句法关系和语义关系是由支配关系决定的,而且二者相互对应,因此在识别支配关系时可以同时分析其句法结构和语义结构。这叫做句法和语义关系同步分析。例如前一节例句 1,根据词典中存放在 make 下的支配关系识别规则,可以判定 make mistake 的支配关系

是用被动语态的形式表现在句子当中的,同时也可以确定 make 和 mistake 之间的语义关系:



附加关系也可以同步分析,只是有些用个性规则,有些要用共性规则。英语句子中的定语或状语一般可以用共性规则作句法和语义关系的同步分析。例如下句中的 the old, those lovely, 以及 because 引导的状语从句: The old lady always follows those lovely children because she loves them. 但是对一些虚词(如英语的介词、关系代词等)的分析,还必须用个性规则才能保证准确。除此之外,实词的一些特殊用法也常常用个性规则来描述,规则的结论通常也是句法和语义同步的。同步分析句法关系和语义关系的根据就是,在直接从属关系中存在着句法结构和语义结构的对应,而且前者可以作为确定后者的线索。

2. 辨识词义

由于辨识词义需要依据语义上的从属关系,所以原则上应该在语义关系确定之后进行。实际上,对处于支配关系当中的多义词,不论是支配成分还是被支配成分,往往都可以在确定句法、语义关系的同时判断它的词义。例如:

- leave (离开/留下) + NP1 + to + NP2[人] : 把 NP1 留给 NP2
- leave (离开/留下) + NP1[地点] + for + NP2[地点] : 离开 NP1 去 NP2
- feed (喂养/供给) + NP + at + breast(乳房/胸膛) : 给 NP 喂奶

对于其他多义词,就需要以句子的语义结构为准,考察与其有直接从属关系的关联词了。同处于从属关系中的多义词及其关联词可以互为上、下位成分。例如,在第二节的例子中,heavy 的词义靠其上位成分 rain 来辨识,make 的词义靠其下位成分 mistake 来辨识。在英语名词短语 a tough coach 当中,tough 和 coach 的词义要靠它们的相互制约来辨识(有的学者把这种现象叫做词义的“互补制约”)。因此辨识词义的操作应该可以自下而上或自上而下地双向进行。为此,我们定义了两类规则,分别按语义的直接从属关系查找上位成分或下位成分。在多义词的词条下面根据需要使用这两类规则,利用语义制约条件确定词义。

查找上位成分(用下划线表示)的规则可以用来辨别修饰语的词义,例如下面的 bad:

- bad(坏/...) + NP[食物] : 腐烂的 NP
- bad(坏/...) + NP[疾病] : 严重的 NP
- bad(坏/...) + NP[身体器官] : 疼痛的 NP
- bad(坏/...) + smell/odour : 臭味

查找上位成分(用下划线表示)的规则还可以用来辨别附加语的词义,例如下面的两个副词:

- pretty(漂亮/相当) + AP[性质] : 相当 AP

- It snows/rains heavily(重/大/猛烈/...) : 雨/雪下得大

查找下位成分(用下划线表示)的规则可以用来辨别名词、动词、形容词、介词的词义:

- legend(传说/图例) + of/in + map : 地图的图例
- doctor(医生/博士) + of + NP[学科] : NP 博士
- look(看/显得/...) + big : 显得了不起
- NP[动物] + work(工作/运转/起作用/...) : NP 工作
- NP[机器] + work(工作/运转/起作用/...) : NP 运转
- NP[非动物/非机器] + work(工作/运转/起作用/...) : NP 起作用
- good(好的/有益的/...) + at + NP/动名词短语 : 善于 NP/动名词短语

in(在...中/...) + manner/way/ratio/order : 以...的方式/比率/顺序

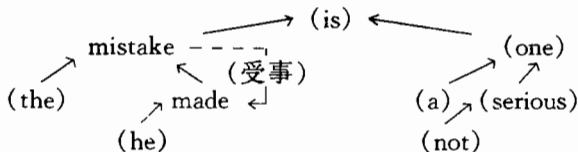
对于某些多义词,单靠上位成分或单靠下位成分还不足以辨别词义,这时可以把两个方向的查找结合起来。例如:in + spite(怨恨/恶意) + of + NP : 不管 NP

四. 基于间接从属关系的语义分析

1. 确定语义关系

句子的成分之间在语义上有从属关系而在句法上没有从属关系,这种情况我们叫做间接从属关系。例如第二节例句 2—8 中的 make 和 mistake。间接从属关系往往是支配关系的典型形式在句法层次发生变化造成的,我们把这种变化叫作句式转换。经过句式转换,支配成分和被支配成分之间在句法上的从属关系被屏蔽了,而语义关系却作为间接从属关系保留下来。在这种情况下,被屏蔽掉的句法关系仍然可以作为分析语义关系的线索。

那么如何操作呢? JFY-5 机器翻译系统认为,在不同的句子里,支配关系典型形式中的各个成分可以有不同的分布,也可以有省略和扩充,这些变化属于句法结构的层次,受句法规律的制约,因此其规律可以由共性的语法规则描述。分析支配关系的共性规则叫做转换规则。转换规则按后置定语、非谓动词等形式分成几类,其条件部分描述被支配成分在转换句式里的分布,以及扩充和省略的规律,结论部分给出与典型形式结论相对应的关系,以及转换句式所表达的语法意义。在词典里以义项为单位建立的个性规则描写的是支配关系的典型形式。源语分析时,系统的控制算法在个性规则和转换规则之间切换,使个性规则和共性规则配合起来,识别句子里以各种形式出现的支配关系。支配关系识别出来以后,语义关系也就同步分析出来了。对于具有间接从属关系的成分,略去其句法关系的结论,只记下语义关系的结论即可。譬如对例句 2:



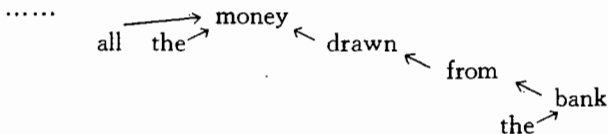
在例句 7 和 8 中,支配关系典型形式的变化又有所不同,在分析 make 支配 mistake 的句式转换时,需要结合描述 worth 和 easy 用法的个性规则。

2. 辨识词义

间接从属关系多数是支配关系的典型形式发生句式转换造成的。在第三节我们指出,对处于支配关系当中的多义词,不论是支配成分还是被支配成分,可以在确定句法、语义关系的同时判断它的词义。因此,我们在用本节第一段介绍的方法为间接从属关系联系的词确定语义关系的同时,就已经可以顺便确定词义了。

这里我们想讨论的是另一种情况,请看英语句子:

He has spent all the money drawn from the bank.



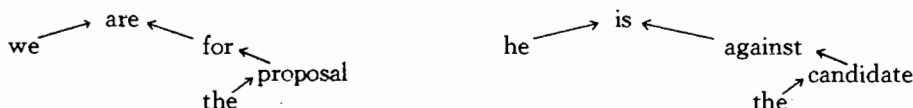
要确定 bank 的词义(银行/岸),需要查到 money。两个词之间不直接互为上下位词,但在语义结构中仍有间接的上下位关联,可以通过从属关系找到。我们把这种关联也叫做间接从属关系。在辨识多义词时,对其上位成分或下位成分作延长查找,就可以找到对排除歧义有作用的关联词。

五. 余论

本文认为在机器翻译系统中,确定句子的语义结构应以句法结构为基础,辨识词义可以语义关系为线索。在词义辨识时,利用从属关系(直接的或间接的)可以排除大部分歧义,但不是全部。有些词义要用别的方法辨识,其中比较可行和可靠的方法仍然与句子的语义结构有关。

在句子的语义结构中除了从属关系以外,还存在着非从属关系,比如第一节中基本观点5)提到的同位成分,在语义上就是一种非从属关系。非从属关系也表示一种关联性,也可以用作词义辨识的线索。请看英语句子:

We are for the proposal. 和 He is against the candidate.



要确定介词 for 和 against 的词义(分别为“赞成”和“反对”),除了需要知道它们的句法成分和介词宾语的义类以外,还要找到句子的主语,看它是不是具有主观意愿。在这两句中作主语的 we 和 he 与介词 for 和 against 在语义上是非从属关系,但仍然可以在语义结构中找到它们之间的联系。具体作法是:从 for(或 against)向上,通过上位成分 are(或 is)转接,找到 we(或 he),作两个词之间的临时联系,待词义辨别完成后,即放弃这种临时联系。

这种辗转查询可以多次转接,从理论上说能在整个句子的范围内查找非从属关系的关联词。但实际上,加大查找距离就趋向于语境分析了,那时单凭语义关系可能会感到难于控制。显然,这个问题已经超出了本文讨论的范围。

除此以外,机器翻译系统还需要其他排除歧义的方法,例如兼类词判别,专业主题限制,语境分析,等等。这些就与句子的语义结构没有关系了。

附注:

- ①JFY-5 机器翻译系统由中国社会科学院语言研究所刘倬、傅爱平和李维研制,项目主持人是刘倬研究员(刘倬等,1989)。
- ②句子中有的成分之间在句法关系上没有直接联系,但在语义关系上相关联,这种关联性称为准直接联系。
- ③在机器翻译中,对“语义分析”通常有两种理解:一是语言分析的手段,即借助语义信息作语言分析;二是语言分析的目的,即为了确定语义结构或辨识词义而进行的分析,本文讨论的“语义分析”是第二种。

参考文献

- [1]冯志伟(1987)《现代语言学流派》,陕西人民出版社,p139-151
- [2]Hudson, R. (1991)《English Word Grammar》,Oxford: Basil Blackwell
- [3]Grishman, R. (1986)《Computational Linguistics: An Introduction》, Cambridge University Press
- [4]李洁(1986) Kalevi Tarveinen 的从属关系语法导论,《国外语言学》,No. 4
- [5]李维 刘倬(1990) 机器翻译词义辨识对策,《中文信息学报》,No. 1
- [6]刘倬 傅爱平 李维(1989) JFY-IV 机器翻译系统概要,《中文信息学报》,No. 4
- [7]朱靖波 王宝库 姚天顺(1995) 机器翻译中的消歧策略,《计算语言学进展与应用》,清华大学出版社,p289-295