

句法结构表示方法的发展及探索*

邢翠鹃¹, 周逸初², 李斌^{1,2}, 贺胜¹

(1. 南京师范大学 江苏 南京 210009; 2. 南京大学软件新技术国家重点实验室 江苏 南京 210023)

摘要: 计算机自然语言处理往往需要从线性的句子中获取词语之间的语义关系, 树形的句法结构大体上可以推导出句子成分的主要语义关系, 在自然语言处理中发挥了重要作用, 但随着近年来语料库建设规模的不断扩大, 研究者发现用投影树无法完整地描写句法结构, 出现了相当数量的非投影树结构和图结构。同时, 语义角色标注的语料加工和计算实验也挑战着树形结构的统治地位。一个名词性成分往往可以充当多个谓词的论元, 这使得句子的论元结构出现了图结构。以 AMR 为代表的研究者提出了基于图的句子语义表示方法。本文则进一步对 CoNLL2009-ST 开放的评测数据进行统计分析, 分析结果表明了根据句法树结构难以推导出语义结构, 明确了需要探索汉语句子的基于图的句法语义一体化标注方案。

关键字: 句法分析; 树结构; 非投影树; 图结构

The development of syntactic structure representation and new exploration

Abstract: The natural language processing need to obtain the semantic relations between words from linear sentence. Tree syntactic structure played a significant role in natural language processing for that we generally can deduced the sentences' semantic relations. But with the continuous expansion of corpus in recent years, researchers have found that using tree structure cannot fully describe syntactic structure, and existing a significant number of non-projective tree structures and graph structures. Meanwhile, the semantic role labeling and computational experiments also challenging the dominance of the tree structures. A nominal composition can often act as the argument of multiple predicates, which makes the argument structure of a sentence appeared the graph structure. Researchers represented by AMR proposed a method of sentence semantic representation based on graph. This paper further analyze CoNLL2009 evaluation data, preliminary proved that it's difficult to derive the semantic structure according to the syntactic tree structure, and the results has been clear about the need to explore syntactic semantic integration annotation scheme based on graph for Chinese.

Keywords: parsing; tree structure; non-projective tree structures; graph structures.

1 引言

句法分析一直是自然语言处理的关键技术之一, 其主要的任务就是确定句子的句法结构。一般地说, 获取句子的句法结构不是自然语言处理的最终目标, 但是句法分析可以将自然语言的句子的句法语义结构清晰地表示出来, 便于各种后续任务的进一步进行, 即它是实现具体目标的重要环节。

对句子进行句法分析一般会关注句子成分之间的句法关系、语义依存关系或论元关系, 但是在实际的句法分析中, 以上三种分析往往都只能揭示句子某一方面的语义关系, 并且每一种分析都是单独进行的, 这和计算机在处理自然语言时需要获取句子的完整的句法语义关系这一目标还有一定的差距。本文将对传统句法结构表示方法及其发展进行一些介绍, 并从理论上和实际数据上论证引进图结构的必要性, 进而探索对汉语句子进行基于图的句法语义一体化标注, 这样才能为进一步的自然语言处理工作做好准备。

*本研究受到国家自然科学基金(61170181、61300158)、国家社科基金(1101065C)的资助。

2 相关术语界定

对本文的重要术语——包括句法依存、语义依存、论元关系等界定如下：

2.1 句法依存和语义依存

根据 Percival^[1]和 Fisher^[2]关于依存关系的分类，本文中的句法依存和语义依存等术语界定如下：

句法依存是依存关系中的一种，也可以称为结构依存，如果两个成分 A 和 B，如果 B 出现，则 A 也出现（当然，A 也可以单独出现）。这时就可以说 A 支配 B，或 B 依存于 A，利用的是比较严格的分布同现。在英语中，例如，及物动词之后必须要有宾语，而不及物动词之后则不能添加宾语；在汉语中，例如，“吃桃子”中，“桃子”出现的时候“吃”也会出现（一般不能用“喝”等其他表示进食的词语），这时“吃”支配“桃子”，也可以说“桃子”依存于“吃”。

语义依存指的是某些词的出现只是为了限定其他词的意义，即如果一个聚合单位 A 的成分在语义上限制了另一个聚合单位 B 的成分，那么 A 就依存于 B，主要是词汇间的联系。在汉语中，例如，定中短语“大桌子”中，“大”属于形容词这一聚合单位，“桌子”属于名词这一聚合单位，并且前者在语义上是依存于后者的。

还有另外一种依存关系的类型，即形式依存，指的是一个聚合单位 A 的成分限制了另一个聚合单位 B 的成分的形式，主要使用的是语法手段或语法关系。在英语中，主要表现为句子中时、数、态等的一致性。由于汉语不严格依赖形态变化，因此我们在句法分析的过程中就主要分析句法依存和语义依存。

2.2 论元关系

根据顾阳^[3]的介绍，句子中由谓词根据其与其相关的名词短语之间的语义关系而分配给这些名词短语的语义角色^[4]，那么这些名词短语和谓词之间的这种关系就是论元关系。在汉语中，例如，“小明用勺子吃米饭。”这句话中，“小明”是谓词“吃”的施事，“用勺子”是谓词“吃”的工具，而“米饭”是谓词“吃”的受事。当然，不同的谓词会有不同数量、不同类型的论元。我在分析句子语义关系的时候，也会分析这些论元关系。

3 句法结构表示方法的发展过程

在语言学的发展过程中，由最初用树形结构来表示句法结构，到接下来发现特殊的树结构——非投影树结构，再到初步尝试和应用图结构，经历了由树到图的发展过程。

3.1 树结构

语言学家们首先是用投影树结构来表示句法结构，到后来在句子分析过程中又进一步发现了非投影树结构。

美国语言学家 Chomsky 1955 年在其博士论文 *The Logical Structure of Linguistic Theory* 中就提出了句法树理论，之后在 1957 年的著作 *Syntactic Structures* 中则详细分析了句法结构的层级性和转换模型。文中就对英语句子进行了树结构的表示^[5]。如下图：

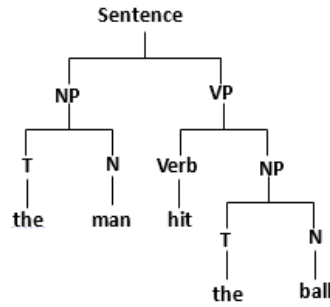


图 1: Chomsky 的句法树结构

德国的 Franz Kern 认为，句子是“一种靠式动词来实现的针对思想的语言表达式”（转引自 Baum 1976: 40）。之后他又发展了这一观点，使句子结构表示为以动词为根的一棵树，之所以会选择用树来表示句法结构主要是基于：二维的树结构有助于不依赖词序来表示句法结构；树结构简单清晰；树结构在数学领域已经有许多形式化的手段和工具来处理，利用起来比较方便；树结构在语言学和计算机科学领域都广为人知^[6]。之后法国的 P. Ihm 和 Y. Lecerf 提出了投影理论（PROJECTIVITY）^[7]。下图就是其投影理论的概况：

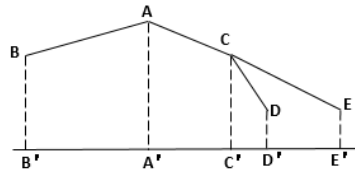


图 2: P. Ihm 和 Y. Lecerf 的投影理论

在这个图中，“BACDE”是一个完整的句子，“B'A'C'D'E'”是这个句子投影后的结果（其实还是“BACDE”，这里只是为了区别）。其中的每个字母代表一个最小的语言单位（词语），实线表示两个语言单位之间有直接的依存关系，虚线则表示投影关系。

在文中他举出一个法语例子“Dieu créa l'homme(上帝创造人。)”(le 是定冠词阳性形式，后面的单词是字母元音，强制缩合成 l'，相当于英语中的 the。)，分析如下：

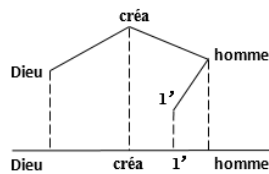


图 3: Dieu créa l'homme 的投影结果

在这句话中，动词“créa（创造）”作为全句的中心，是整棵句法树的根节点。在这个图中，他表示出了单词之间的依存关系，但是没有标注单词之间依存关系的具体类型。

法国的泰尼埃 (Tesnière) 更是在句法分析中普遍使用了依存图式 (Stemma) 这一概念。他还把句子图式分为实图式 (含有实词) 和虚图式 (含有符号，分别用 I,O,A,E 代表动词、名词、形容词和副词)，这也是一种投影树结构^[5]。

美国的 Hays^[8]详细讲解了投影树结构，他认为一棵树就是一系列已经定义了的元素的集合，基本上每个词语都至少与另一个词语有关系；如果其中有一个元素作为起点，并与其它元素相区别，那么就是这棵句法树的根；如果这棵树上的元素之间的连线标注了性质（不是我们定义的依存关系），那么这棵树就变成了一棵有标记的树。在依存树结构中，根节点就是作为句法中心的语言单位，语言单位之间的关系就是直接依存关系（句法依存），标签只是辅助特征和顺序指示器。图 4 就是其理论的一个形式化：

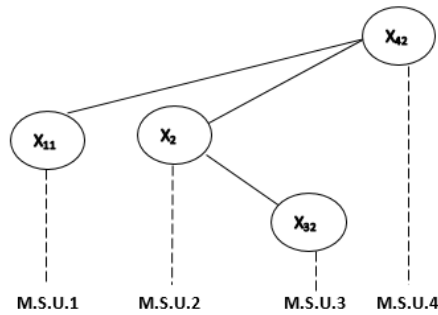


图 4: Hays 的投影理论

在图 4 中 M.S.U.代表最小语法单位 (minimal syntactic unit), 顶点处的 X_{42} 代表整棵句法树的根节点。

这种简洁、明了的句法结构表示方法之后又不断地发展了, 其在汉语中的应用最著名的是哈工大语言技术平台 (Language Technology Platform, LTP), 其对汉语句子进行分析, 并对每一条有向的关系弧进行标注, 形成带标签的投影树结构。图 5 就是其句法分析及标注的可视化呈现:



图 5: LTP 对“上帝创造了人”的分析及标注结果

图 5 中, 关系弧上的标记代表关系弧两端的句子成分之间的关系类型, 关系弧的方向则指向被支配的句子成分。Root 就指向整棵句法树的根节点——谓词“创造”, 谓词“创造”支配主语“上帝”, 宾语“人”, 辅助语“了”和标点符号“。”。

但是在具体的语言分析过程中, 语言学者们相继发现了例外, 就是说, 语言事实中存在一些句子无法进行顺序投影, Holan 等首次提出无限制的依存树 (unrestricted dependency trees)^[9], 后来称为非投影性 (non-projective)^[10]。根据布拉格依存树库 1.0 (The Prague Dependency Treebank, PDT1.0) 的统计 (见下表: 表 1), 在其对捷克语的分析中有大约 23.15% 的句子是非投影性的句子, 而且 Havelka 统计的数据显示, 德语和荷兰语树库中的非投影句的比例更大, 分别是 27.75% 和 36.44%^[11,12]。他统计了 12 种语言中由于错嵌套、非平面和非投影三方面原因造成的非投影结构所占的比例及总句数。具体情况如下表:

表 1: 对几种语言中非投影结构的统计

Language 语言	ill-nested 错嵌套	non-planer 非平面	non-projective 非投影	proportion of all (%) 占有所有句子的比例	All 总句数
Arabic 阿拉伯语	1	150	163	11.16%	1460
Bulgarian 保加利亚语		677	690	5.38%	12823
Czech 捷克语	79	13783	16831	23.15%	72730
Danish 丹麦语	6	787	811	15.63%	5190
Dutch 荷兰语	15	4115	4865	36.44%	13349
German 德语	416	10865	10883	27.75%	39216
Japanese 日语		1	902	5.29%	17044
Portuguese 葡萄牙语	7	1713	1718	18.94%	9071
Slovene 斯洛文尼亚语	3	283	340	22.16%	1534
Spanish 西班牙语		56	57	1.72%	3306
Swedish 瑞典语	71	1076	1079	9.77%	11042
Turkish 土耳其语	14	556	580	11.6%	4997

并且指出，在英语中这类句子尤其存在于关系从句中，下面就是 McDonald 所举出的一个英语中的例子^[10]：



图 6: 英语中的非投影依存树

在“John saw a dog yesterday which was a Yorkshire Terrier. (约翰昨天看到一只约克郡小猎犬)”这句话中，“saw-yesterday”这条弧与多条有向弧交叉，因为这里的“yesterday”位于宾语“a dog”及修饰宾语的从句之间，就导致了句子无法投影成原来的顺序。在这个句子里，每个节点仍然满足入度(In-degree)为1的限制，即其实质还是树结构，只是其不能顺序投影为原始的句子。

语言学家 Hudson 也介绍了这类现象，他称之为“例外(exception)”^[13]，对这类句子，他通过添加词语等方法避免句子成分之间弧的交叉。下图就是 Hudson 《词语法》原文中的一个例子：

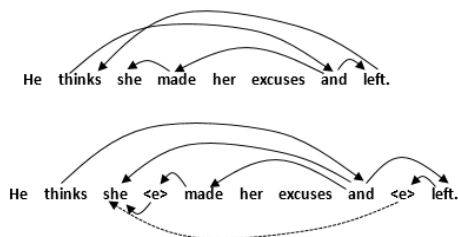


图 7: Hudson 的 WG 最终的例外及解决办法

他认为，这个句子中的关系弧之所以会交叉，是因为说话者省略了两个“she”，在图 7 中下面的句子中他用<e>补充出来，这样就使整个句子避免了关系弧的交叉。

但是并不是所有的交叉都能通过这种方法解决的，就像图 6 中的句子就无法补出省略的内容，因为造成关系弧交叉的原因不是只有省略这一种。也就是说，这并不是解决问题的根本办法。

造成非投影树的原因就是 Tesnière 说的句子中存在一个潜在的关联，它是语义上的关联而不是结构上的关联。潜在的关联在图式中用虚线表示。例如：

Alfred aime son père. (阿尔弗列德爱他的父亲)

在这个句子中，son (他的)这个词不仅与其依存的词 père (父亲) 有结构上的关联，而且，它和 Alfred 还有语义上的关联。图示如下：

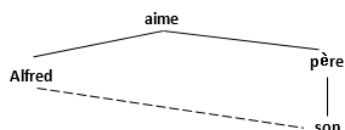


图 8: 潜在的关联

3.2 图结构

语言中往往存在某一句子成分被多个句子成分支配的现象，这就突破了每个词语只能和另外一个句子成分存在依存关系的限制，体现在句法结构的表示上就是某些节点的入度 (In-degree) 可能会大于 1。这就挑战着树结构的句法结构表示方法在自然语言处理中的统治性地位。因此，我们要找到一种合适的方法来描述句子成分之间的这种“潜在的关联”。Drewes 等提出了超边替代图语法 (Hyperedge Replacement Graph Grammar) [14]。另外 Sagae & Tsujii 也认为，尽管树结构无论在计算还是在语言学上都有很多令人满意的优点，但是用树结构难以完全获得那些超出了浅层语法的语言学现象的结构。他们继而提出一种称为有向无环图 (directed acyclic graphs, DAG) [15]的方法，这种方法对英语的谓词-论元分析有明显优势，同时通过用长距离依存和共指、引用关系等表示方法对丹麦语的句法语义分析也有帮助。

基于 Drewes 等提出的超边替代语法 (Hyperedge Replacement Grammar, HRG) [16]和一些上下文无关文法，Kevin Knight, Philipp Koehn, Martha Palmer 等提出了用抽象的意义表示 (Abstract Meaning Representation, AMR) [17]来标注英语句子。AMRs 是有根节点、方向、边界标记和叶子节点标记的图；对有相同含义的句子分配同样的 AMR，而不考虑句子的外在形式；充分利用了命题库框架集 (PropBank framesets) [18,19]中的框架结构。对句子进行语义上的分析和抽象。下图就是一个例子：

```
(d / describe-01
  :arg0 (m / man)
  :arg1 (m2 / mission)
  :arg2 (d / disaster))
```

图9: 框架集“describe-01”的 AMR 表示

在这个表示中, d 是 describe 的首字母, 在框架集“describe-01”有三个预定义的空位(也就是“论元”, 论元 0 是描述者, 论元 1 是被描述的事物, 论元 2 是此事物被描述为什么)。这个 AMR 可以表示以下句子:

The man described the mission as a disaster. (这个人把这次任务描述为一次灾难。)
 The man’s description of the mission: disaster. (这个男人对这次人物的描述是: 灾难。)
 As the man described it, the mission was a disaster. (据这个男人描述, 这次任务是一次灾难。)

其实还不止这些句子, 只要句子的主干是“describe-01”这个框架, 就可以从这个句子抽象出图 10 所示的这个 AMR。

AMR 还有另一种表示方法——图格式 (GRAPH format), 更一目了然:

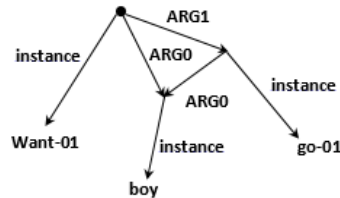


图 10: “The boy wants to go. (这个男孩儿想要走。)”的 AMR 的图格式

句中的“boy”, 在语义上既是“want”的论元, 也是“go”的论元, 导致“boy”这个节点的入度大于 1, 这就形成了图结构。图 10 中的实线也表示依存关系, 只是这里不再只是句法依存关系, 还有语义依存关系。计算机需要从线性的句子中获取词语之间的语义关系, 这样才能进一步对自然语言进行自动处理。因此接下来的发展也更加关注语义依存关系。语义分析需要更大、覆盖率更广的数据库, 以下这些项目就是为满足这样的需求而建立的, 例如, 格罗宁根意义库^[20], UCCA^[21], 语义树库 (the Semantic Treebank, ST), 布拉格依存树库 (the Prague Dependency Treebank) 和通用网络语言 (Universal Networking Language, UNL, 即在自然语言处理中用知识表示语言^[22,23]等。

3.3 树结构和图结构的比较

树结构有其自身的优点: 相对于直接成分分析法和层次分析法, 树结构的句法结构表示在可视化方面更加清晰明了; 在深度和广度上可以无限扩展; 在算法上也容易实现等。

但是, 树结构也有一些无法克服的不足: 树结构硬性规定每个语言单位只能与另外一个语言单位有依存关系, 这就不符合语言事实, 因为在实际的交际中, 某个语言单位很可能和多个语言单位都有依存关系。尤其是在分析句子成分之间的语义依存关系时, 这种情况就会更多。用树结构不能完全揭示这些复杂的句法语义依存关系, 也就无法使计算机通过树结构来获取句子的语义关系, 这就不利于自然语言处理工作的进一步展开。

相对来看, 图结构就有明显的优势: 图结构中, 每个节点不限制其父节点和子节点的个数, 这就对应于语言事实中某一语言单位可能和多个语言单位都有依存关系这一现象, 因此能够更好地展现这些关系。

如果对句子进行单独的句法依存、语义依存或论元关系的分析就会呈现出非投影树结构, 然而, 正如上文所述, 这样单独的分析结果却不能完全反映出句子中所包含的语义关系, 为

了能够得到句子中包含的完整的语义关系,我们需要将句法依存关系和论元关系合并到一起,或将语义依存关系和论元关系合并到一起。

4 汉语中图结构问题探究

我们在进行汉语句法语义分析的过程中,发现无论是投影树结构还是非投影树结构,都无法完整地揭示句子中的语义关系,而图结构则能够做到。但是之前的基于图的句法分析都是针对英语等其他西方语言的,我们则尝试对汉语句子进行基于图的语义依存标注,发现图结构对汉语句子的语义依存关系的解释力也很强大。另一方面我们还对 CoNLL2009-ST 开放的评测数据进行了统计分析,从数据上验证了我们的设想即需要探究对汉语的基于图的句法语义一体化标注方案。

4.1 汉语中的图结构

图结构对汉语句子的语义关系的解释力也更好。下面就是一个例子:

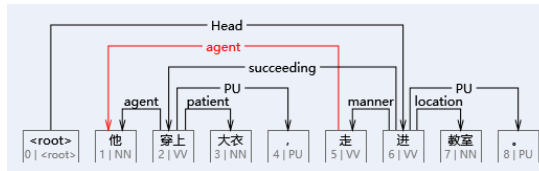


图 11: “他穿上大衣,走进教室”的依存图标注结果

这是一个典型的连动句,第一个动词“穿”的主语是“他”,对于第二个动词“走”,一般的处理办法是作为“穿”的并列动词,但是这没有体现出其与主语“他”的语义关系。图 11 中,句子中红色的那条有向弧就使这个树结构变成了图结构(“他”的入度为 2),找到这条有向弧,才可能完全揭示句子成分之间蕴含的语义关系。

4.2 图结构的可推导性

我们使用的数据是来自于 CoNLL2009 评测的中文数据,其中的每一个句子都被标注成三种不同的结构,分别是句法依存关系、语义依存关系以及论元(SRL)关系。我们为了得到句子中更多的语义关系,需要将句法依存关系和论元关系合并到一起,或将语义依存关系和论元关系合并到一起。在合并的过程中,我们发现其中的有些论元关系是可以直接从句法依存关系或语义依存关系推导得出的。如果所有的论元关系都能推导出来,那么说明我们不需要图结构就能完整的表达出句子中包含的信息。针对这一问题,我们展开了一系列的统计分析工作。

如果我们将句法依存关系和论元关系画到同一个树结构中,将会得到如下的图结构:

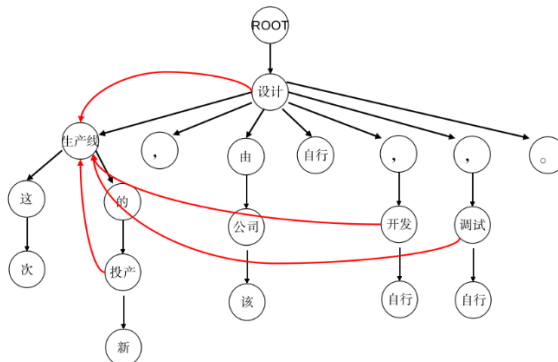


图 12: 汉语中图结构的实例

图 12 中，黑色的有向弧表示句法关系，红色的有向弧表示论元关系，可以很明显的看到，“开发”和“调试”处于树结构的同一深度，可以通过树结构本身含有的信息直接推导出来，而“设计”和“投产”就不那么明显了，不能通过一些简单的规则进行推导。对于这些不同的情况，我们对这些不同节点关系进行了分类，得到了三种的不同的节点关系，分别如下：

1. 兄弟关系。这里的兄弟关系是广义上的，如果两个节点在树上的深度是相同的，那么，我们就认为这两个节点是兄弟关系。比如图中的“开发”和“调试”。

2. 祖先关系。一个节点是另外一个节点的祖先。比如图中的“设计”和“投产”。

3. UNK。除去上述两种关系以外的其他情况都属于 UNK。这些关系是很难利用某种规则进行推导的，我们标记为 UNK。比如图中的“开发”和“投产”。

在算法上，这三种关系中，兄弟关系是最容易从树结构中推导出来的，祖先关系推导起来比较困难，而 UNK 关系则基本无法进行推导，为了检验在真实的汉语语境中，不可推导的情况有多少，我们利用 CoNLL2009 的中文数据进行了一系列的统计分析，统计结果如下表所示：

表 2：汉语树结构论元关系节点的统计

依存关系	节点关系	关系个数	所占比例
句法依存关系	兄弟关系	499	13.06%
	祖先关系	2963	77.52%
	UNK	360	9.42%
语义依存关系	兄弟关系	556	14.55%
	祖先关系	2613	68.37%
	UNK	653	17.09%

从表中可以看到，祖先关系占有比较大的比重，而可以直接推导出的兄弟关系只占有不到 20% 的比例。为了验证祖先关系是否可推导，我们对祖先关系进行更进一步的统计分析，详细统计了这些祖先关系中具体的深度差。

表 3: 祖先关系不同深度层次的统计

依存关系	深度差	计数	所占比例
句法依存关系	1	840	28.35%
	2	1338	45.16%
	3	645	21.77%
	4	102	3.44%
	5	34	1.15%
	6	4	0.13%
语义依存关系	1	843	32.26%
	2	960	36.74%
	3	562	21.51%
	4	148	5.67%
	5	69	2.64%
	6	20	0.765%
	7	5	0.191%
	8	2	0.07%
	9	1	0.038%
	10	2	0.07%
	11	1	0.038%

从此表中可以看到, 祖先关系中, 其分布主要集中于深度差 1-3 之间, 但是也没有出现特别明显的集中现象, 这就证明了我们的设想——不能通过规则来抽取论元关系, 这也就是说单纯的树结构无法包含句子所有的信息, 因此引入图结构模型是一种比较合理的选择。

5 总结及未来的工作

本文通过对句法结构表示方法的一系列总结和分析, 验证了树结构在完整揭示句子语义关系方面不具优势, 而图结构则能更好地做到, 因为图结构在广度和深度上同样是可以无限扩展的, 这就有助于表现更加复杂的句子; 更重要的是图结构不限制节点的入度, 这就能够更全面地表现句子中句法成分之间复杂的语义关系; 只有图结构才能全面地表现句子成分之间的句法依存关系、语义依存关系和论元关系。然后通过对 CoNLL2009-ST 开放数据的统计分析, 也从数据上明确了在分析句子的时候我们需要通过句法关系、语义依存关系和论元关系的多重分析, 才能完整地获得句子的语义关系, 才能为进一步的自然语言处理做好准备。当然要对句子进行这样的多重分析, 就会形成图结构, 验证了引入图结构的合理性。对汉语句子进行基于图的句法语义一体化标注则是我们下一步的工作重点。

参考文献

- [1] Percival, W.K.1990. Refelections on the History of Dependency Notions in Linguistics. *Historiographia Linguistica* 17:29-47.
- [2] Fischer, K.1997. *German-English Verb Valency*. Tübingen: Gunter Narr Verlag.
- [3] 顾阳. 论元结构理论介绍[J]. 国外语言学, 1994(1):1-11.
- [4] 袁毓林. 论元角色的层级关系和语义特征[J]. 世界汉语教学, 2002(3):13-22.
- [5] 刘海涛. 依存语法的理论与实践[M]. 北京: 科学出版社, 2009. 6
- [6] Schubert, Klaus(1986):Syntactic Tree Structures in DLT. *Utrecht: BSO/Research*.
- [7] P. Ihm and Y. Lecercf, Eléments pour une grammaire générale des langues projectives (Bruxelles,

- 1963).
- [8] Hays D. Dependency Theory: A Formalism and Some Observations. *Languages*, vol. 40, no. 4, 1964:511–525.
- [9] Tomáš Holan, Vladislav Kuboň, Karel Oliva, and Martin Plátek. 1998. Two Useful Measures of Word Order Complexity. In *Alain Polguère and Sylvain Kahane, editors, Proceedings of Dependency-Based Grammars Workshop, COLING/ACL*, pages 21 – 28.
- [10] McDonald R., F. Pereira. Non-projective Dependency Parsing using Spanning Tree Algorithms. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*. 2005:523–530.
- [11] Jiří Havelka. 2007. Beyond projectivity: Multilingual evaluation of constraints and measures on nonprojective structures. In *45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 608 – 615, Prague, Czech Republic.
- [12] Nivre, Joakim. 2007. Incremental non-projective dependency parsing. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT' 07)*.
- [13] Hudson, Richard. 2005. Word Grammar. In *K. Brown (Ed.), Encyclopedia of Language and Linguistics (second ed., p633–642)*. Elsevier.
- [14] Frank Drewes, Annegret Habel, and Hans-Jorg Kreowski. 1997. Hyperedge replacement graph grammars. *Handbook of Graph Grammars and Computing by Graph Transformation*, pages 95 – 1626.
- [15] Maier, Wolfgang and Laura Kallmeyer. 2010. Discontinuity and non-projectivity: Using mildly context-sensitive formalisms for data-driven parsing. In *Proceedings of the Tenth International Conference on Tree Adjoining Grammars and Related Formalisms (TAG+), New Haven, CT*
- [16] D. Chiang, J. Andreas, D. Bauer, K. M. Hermann, B. Jones, and K. Knight. Parsing Graphs with Hyperedge Replacement Grammars. *Proc. ACL*, 2013.
- [17] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. Abstract Meaning Representation for Sembanking. *Proc. Linguistic Annotation Workshop*, 2013.
- [18] R. Martins. 2012. Le Petit Prince in UNL. In *Proc. LREC*. P. Kingsbury and M. Palmer. 2002. From TreeBank to PropBank. In *Proc. LREC*.
- [19] M. Palmer, D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1).
- [20] V. Basile, J. Bos, K. Evang, and N. Venhuizen. 2012a. Developing a large semantically annotated corpus. In *Proc. LREC*.
- [21] Abend and A. Rappoport. 2013. UCCA: A semantics based grammatical annotation scheme. In *Proc. IWCS*.
- [22] H. Uchida, M. Zhu, and T. Della Senta. 1999. A gift for a millennium. Technical report, *IAS/UNU Tokyo*.
- [23] H. Uchida, M. Zhu, and T. Della Senta. 1996. UNL: Universal Networking Language—an electronic language for communication, understanding and collaboration. *Technical report, IAS/UNU Tokyo*.