

文章编号: 1003-0077 (2011) 00-0000-00

基于语义依存图库的兼语句句模研究 *

郑丽娟, 邵艳秋

(北京语言大学信息科学学院, 北京 100083)

摘要: 句子语义分析是语言研究深入发展的客观要求, 也是当前制约语言信息处理技术深度应用的主要因素。在探索深层语义分析方法的基础上, 文章根据汉语的特点提出了语义依存图深层语义分析理论, 并且根据这种理论, 建立了一个包含 30,000 个句子的语义依存图库。以兼语句为重点研究对象, 文章研究了语料库中所有纯粹的兼语句所对应的句模情况, 进而试图构建基于语义依存图的句模系统, 总结句型和句模的映射规则, 从而为更好的建立语义自动分析模型提供相应的知识库。

关键词: 句模; 语义分析; 语义依存图; 兼语句

中图分类号: TP391

文献标识码: A

The Study on the Semantic Sentence Patterns of Chinese Pivotal

Sentence Based on Semantic Dependency Graph

ZHENG Lijuan, SHAO Yanqiu

(College of Information Science, Beijing Language and Culture University, Beijing 100083, China)

Abstract: Semantic analysis of sentences is the requirement of further development of language study, it is also the major bottleneck restricting the large-scale application of language information technology at present. Based on the study of deep semantics analysis methods, according to the characteristics of Chinese, we propose a new semantic analysis method—Semantic Dependency Graph, and construct a corpus consisting of 30,000 sentences. Furthermore, we make a study on the semantic sentence patterns of pure pivotal sentences in the corpus, and try to construct the system of semantic sentence patterns based on semantic dependency graph and summarize corresponding relations between sentence patterns and semantic sentence patterns to offer the automatic semantic parsing system a corresponding knowledge base.

Key words: semantic sentence patterns; semantic analysis; semantic dependency graph; pivotal sentence

1 引言

句子语义分析是语言研究深入发展的客观要求, 也是当前制约语言信息技术大规模深度应用的主要瓶颈。尤其是对汉语这种意合性语言而言, 句子形式灵活多变, 句子形式不同, 表达的语义就会存在细微差别, 同一种句法结构可以表达多种语法意义, 同一种意义也可以由多种形式表达。虽然关于句子的语义分析研究存在很多困难, 但却也不是毫无规律可言, 只要找出这些规律, 就可以进行更深入的研究。

任何一个句子都是句法、语义和语用的共同体, 不存在一个没有语义语用只有句法的句子, 表达句法结构的句型和表达语义结构的句模之间必然存在着某种对应关系, 但这种对应关系必然不是一对一的简单对应关系, 而是一个一对多、多对一或者多对多的映射关系, 并且这种对应关系中必然有一个或几个是很常用的, 而其余的则是低频使用^[1]。目前, 关于句法的计算研究已经比较成熟, 在句法的基础上, 研究句法和语义的映射关系, 可以帮助句子

* 收稿日期: 2015-6-15 定稿日期: 2015-8-10

基金项目: 国家自然科学基金 (No.61170144); 国家语委十二五科研规划重点项目 (ZDH125); 2014 年度北京语言大学中青年学术骨干支持计划; 北京语言大学研究生创新基金资助项目 (中央高校基本科研业务费专项资金) (15YCX102)

语义分析,将语义分析的研究进一步深化。

兼语句是汉语的一种特殊句型。目前为止,关于兼语句的语义研究或者是只注重句中极个别词语的语言关系描写^[2],或者是只重视分析兼语成分前面和后面的V1、V2之间隐藏的深层事件语义关系^[3]。本文以汉语中的一个特殊句型——兼语句为研究起点,研究兼语句句子所对应的句子语义结构,并对每一类兼语句句模进行了详细刻画,试图描写出兼语句句型和句模的对应关系以期为更大规模的句模研究提供新的研究视角,因为好的句模研究不仅有助于机器自动理解句子、生成符合规范的句子,同时也有助于提高机器翻译和机器自动问答的准确率,促进人工智能的发展。

2 句子语义分析及句模相关研究

2.1 句子语义分析研究现状

关于语义的研究很多,建立起来的相应的语义资源也不少。英文中比较著名的语义资源有:加州大学伯克利分校的FrameNet^[4]、宾夕法尼亚大学的PropBank^[5]以及纽约大学的NomBank^[6]。中文比较有代表性的语义资源主要包括:Chinese PropBank(CPB)^[7], Chinese FrameNet^[8]和Chinese NomBank^[9]。

就研究方法而言,句子语义分析方法主要有两种:浅层语义分析和深层语义分析。面对复杂的语言学现象,深层、全面的语义分析往往需要大量知识的支撑,为了避免深层语义分析研究的困难,许多研究者对语义分析的研究是从浅层语义分析开始的。语义角色标注^[10-11]是目前浅层语义分析的一种主要实现方式,其并不对整个句子做详细、深入的分析,只分析句子中主要谓词与其论元之间的语义关系,并没有对论元内部的各成分之间的语义关系进行分析。此外,语义角色标注通常都是在句法分析的基础上进行的。依据不同的句法分析方法,语义角色标注可以分为基于短语结构分析^[12]、基于依存关系分析^[13]和基于语块的分析^[14-15]。还有部分学者避开转换方法,直接从研究句子语义入手,建立一套语义分析体系,重新标注语料,从而展现句子语义结构的真实状况。如:清华大学李涓子的语义依存分析理论^[16]、武汉大学姬东鸿的基于特征结构理论的语义依存图理论^[17]等。

到目前为止,关于句子的语义分析研究成果很多,也取得了很大的成绩,但也仍然存在一些问题,如:关于语义角色的数量各家都有不同,语义角色数量难以确定;句子的部分语义信息缺失等。

2.2 句模研究现状

句模是句子在语义平面上的结构类型^[18]。如句子“黄曼昨天在教室里告诉我一个秘密。”,其句模为“[施事]+[时间]+[地点]+【告诉】/【传播】+[对象]+[内容]”。句模可分为简单句模和复杂句模两大类^[17],简单句模只包含一个动核结构,上个例句的句模就是一个简单句模;复杂句模则包含一个以上的动核结构,如本文讨论的兼语句。简单句模又根据带不带状元分为基干句模和扩展句模,基干句模只包括谓词和其必有论元。基干句模中各个语义角色位置是相对固定的,位置不同就会形成不同的句模。

随着范晓和胡附先生的三个平面理论的提出,句模研究受到越来越多的重视。对句模研究做过突出贡献的有范晓^[19]、朱晓亚^[20]、徐昌火^[21]和鲁川^[22]等。范晓对句模做出了严格的定义,并认为句模的研究对象是动核结构,动核结构是形成句子语义结构的基础,句模是动核结构生成句子时与句型结合在一起的语义成分的配置模式^[19]。朱晓亚^[20]先将动词分为动作、经验、性状和关系四个大类,然后根据动词在语义平面上所带必有论元数目的不同,从而建立了一个分类层级系统和语义角色系统,进而进行句模分类和描写。徐昌火^[21]认为:“句模研究的是句子在语义平面上的关系语义。并且强调配价研究不应该作为句模研究的起点,而应该是句模研究的归宿。”鲁川^[22]提出了谓词格框架分析系统,他认为句模的语义单位由小到大可以分为概念、事元、事件三个部分,句模的成分包括“中枢事元”和“周边事

元”，它们担任的语义角色分别是“中枢角色”和“周边角色”，而一个“事件”由一个“中枢事元”以及若干个“周边事元”组成。在此基础上，再根据动核的特征把现代汉语句模分为26类并研究每一类别下可以形成的基本句模，共计128个，在研究过程中还讨论了能够进入每一类句模的代表性谓词。

亢世勇、田珍都等在标注了大规模语料的基础上，对句模的研究也做出突出贡献。田珍都^[23]着重讨论了汉语句模体系和句模的使用频度两个问题，把句模分成单句形式和复句形式的句模，并从语料库中提取句模327种，其中单句形式的句模311种，复句形式的16种。在此基础上又将单句形式句模分为简单句模和复杂句模，单句形式的复杂句模又分为一系、二系、三系、四系、多系五种类型，而复句形式的句模又分为相离、相接、融合三种类型。据此建立了现代汉语的句模体系。亢世勇^[24]等人在大规模标注语料库的基础上，建立了汉语的句系系统，从句型角度考察了句模和句型对应关系并统计出句模的频度及分布情况。

现有对句模的研究，有的则过于细化，将语义的细微差别也一起反映到了语义结构的系统分析之中，有的则过于概括，未能完全覆盖汉语复杂多变的句子语义结构；但是过于概括或具体都会失去句模研究在语法研究上的意义。对特殊句式的句模研究、句模与其句法结构对应规律的研究不是太过零散就直接回避。总体来看，对特殊句式的语义分析还不够完善，对其语义成分和句法成分间的对应关系缺乏深入的分析研究。

2.3 兼语句句模研究

兼语句是指由兼语短语充当谓语或独立成句的句子^[25]。所谓兼语短语是指由动宾短语和主谓短语部分重合在一起，其中动宾短语的宾语兼做主谓短语的主语这样一种短语。如：“请他来”，“他”是请的“宾语”，又是“来”的主语，叫做兼语。在语言学界，兼语句语形可以表示为“N1+V1+N2+V2”。N1指句子主语，V1是句子中的第一个谓词，也即兼语动词，N2指兼语，V2是兼语句中的第二个动词。

从兼语句的定义可以看出，兼语句构成的句模一定是一个复杂句模，兼语句句模是一个由单句形式表示的复杂句模。实际上，已经有学者做过关于兼语句语义结构的研究，如孙道功、许小星、黄晓冬等。孙道功^[2]和许小星^[26]等以语块为基点，按兼语动词的不同将其分为5大类兼语句句模，但却忽略了对兼语句中两个动词之间语义关系的描写。而黄晓冬^[3]从本体研究出发，仅研究了兼语句中两个动词组成的动核结构之间的语义关系，并把这种语义关系按照深层语义的不同，分为四类：表目的或结果、表原因、表领有或存在和表称呼、认定。只是他们的研究都还是停留在兼语句中某个词语所承担的语义角色或者词语之间的语义关系上，都还未对这些兼语句句模进行整体刻画。从严格意义上来，还不能算是句模描写。

3 语义依存图

3.1 语义依存树

汉语语言表达形式灵活，语言现象丰富，实际的句子中经常会出现某个词语和句中其他多个词语发生语义关联的现象，即依存词不唯一；另外，不同的依存词对之间也会出现依存弧交叉的情况。这些情况都是传统的依存树结构无法处理却真实存在的现象，并且这些依存树结构无法表达的语义关系，如果不能正确的表示出来，可能会影响对句子意义的理解。

句子“我有个妹妹很能干。”用依存树分析的结果如下图1所示。从图1中可以看出，依存树结构^[27]必须满足四个形式化的条件：1) 单一父节点；2) 连通；3) 无环；4) 投射，也即弧线之间不交叉。在这个句子中，“有”是句子的核心，它在语义上直接或者间接支配句子中的其他语义成分，如“有”直接支配“妹妹”，并且他们之间的语义关系是领有关系，这种语义关系已在弧线上标注出来；“有”间接支配“个”，它是通过支配“妹妹”而支配量词“个”；“个”受“有”的属事角色“妹妹”直接支配。

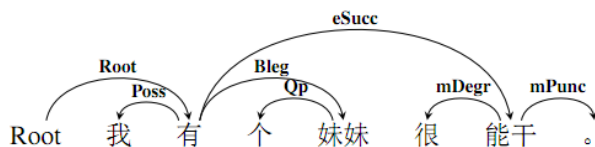


图 1 语义依存树

用语义依存树分析汉语句子清晰，简洁，美观，但有时却会丢失或者错误理解句子的语义信息。如：针对这个句子，机器自动问答想知道“谁”很“能干”，从依存树的标注结果来看，机器得到结果“我”，然而，这个答案是错误的。实际上是“妹妹”很“能干”，机器会得出这个错误的结果，因为用依存树理论分析汉语这种意合性语言还存在一定的缺陷，所以我们对这种理论进行了改造，提出语义依存图理论。

3.2 语义依存图

我们的语义分析以句子为单位，以提取句子中所有的修饰词与核心词对（构建一条从核心词指向修饰词的弧）并标注它们之间的语义关系（在弧上标注词对间的语义关系）为目标，并且通过标注语义依存图的形式体现出来。

实际上，依存图放宽了依存树的四个限制条件，依存图与依存树的不同主要表现在两个方面：1) 允许多父亲节点的出现；2) 允许非投射现象出现，即允许依存弧之间存在交叉 [28]。

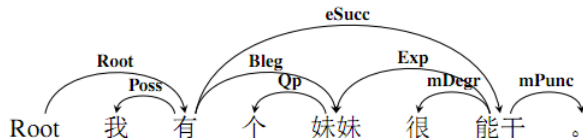


图 2 依存图中的多父节点现象

正如前文提到的依存树结构的不足，为了避免这种缺陷，在依存图结构中，只要词语之间存在真实的语义关系，就允许一个语义成分有多个父节点。这种多父节点的外在表现就是：在依存图中，一个词语可以有多条入弧，比如图 2 中的“妹妹”。“妹妹”有两个父节点“有”和“能干”，相对于两个父节点，“妹妹”承担的语义角色不同：一个是属事角色，另一个是当事角色。这些角色都已在弧线中间表示出来。

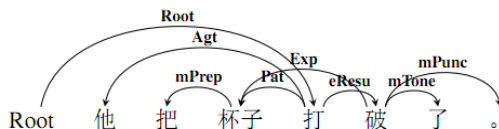


图 3 依存图中的非投射现象

依存图的另一个特点就是允许出现弧线之间的交叉。如图 3 所示：词对（打，他）、（破，杯子）之间的依存弧出现了交叉。而且这样的交叉是必须的，因为如果“破”与“杯子”之间不存在依存弧，就无法明确“破”的当事角色。

3.3 语义依存图库

除了满足形式化的条件外，语义分析还有其自己对语义角色的定义。虽然对汉语语义关系的定义，不同的语言学家给出了不同的划分；董振东在知网中提出事件内部语义关系总计 83 类 [29]，冯志伟提出了 30 种论元关系 [30]，袁毓林提出 40 种语义关系标记 [31]，鲁川提出了共计 26 种语义关系 [22]。各个学者的语义关系体系中所定义的语义关系有交集，有不同，综合考虑以上各种体系的复杂性、可区分性、覆盖性、以及在工程实践标注时的可操作性，以董振东 [29]和鲁川 [22]的体系为基础，我们也定义了一套全面、简单、实用，同时能适应自然语言处理高层应用要求的语义关系体系。该语义关系体系包含五类语义标签：周边角色集、嵌套关系集、反关系集、事件关系集和语义依附标记集。嵌套关系是指由小句或者谓词性短

语充当语义角色的现象，此谓词性短语实际上是降级了的事件。反关系指那些充当论文定语的动词或动词性短语，这种颠倒了的语义关系实际上和正常的动宾搭配语义是一样的，只是动词位置不同，为了区分这种差别，我们又定义了反关系集。

本文旨在总结出一系列语义结构规则，建立相应知识库，从而帮助更加准确地构建语义自动标注模型，其在语义依存研究中所处位置如图 4 所示。

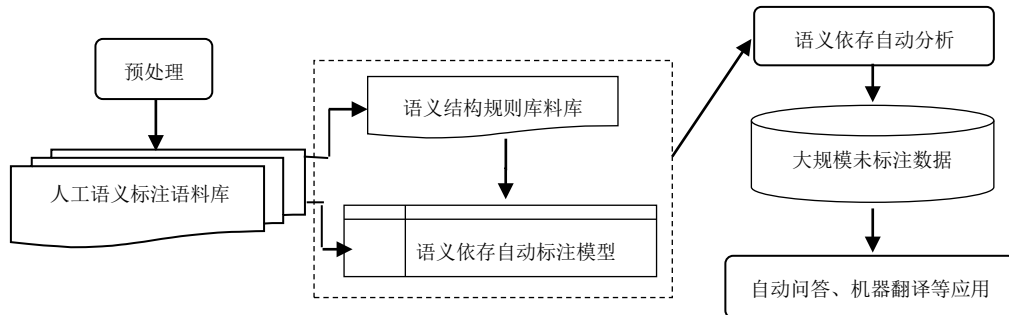


图 4 语义依存图研究内容

依据已经建立的语义关系体系，我们已经建立了一个包含 30,000 个句子的语义依存图语料库，其中已完成对 10038 个句子的校对工作。这些语料都来自于不同的领域，包括新闻语料(10068)、中小学语文课本(10038)、新浪微博语料(5000)和用于机器翻译的语料(4900)。下文对兼语句句模的描写就是以这个已标注的语料库为基础。

4 兼语句句模描写

兼语句作为汉语中一种常用的特殊句式，也常常与其他句式连用，从而使得研究更加困难。为了简化研究，本文以单纯的兼语句（不包括兼语连动连用句）为研究对象。其次，虽然已经标注了 30,000 个句子，但是仍然有 20,000 句等待校对，所以，文章兼语句句模的研究是以 10038 个来自中小学语文课本的校正语料为基础。

4.1 兼语句自动抽取

我们的语料库标注结果采用 CoNLL 数据格式存储，共有十列，如表 1 所示。其中，第一列数字表示词语在句中的位置；第二列是当前词语或标点，第三列为词语原型，这里，中文语料的第二列和第三列相同；第四列和第五列分别为粗粒度和细粒度词性；第六列为句法特征，此处未被使用，以下划线标识；第七列为该词语的父节点即其核心词的序号；第八列为该节点和父节点之间的语义关系；第九和第十列本语料库不予以标记。

表 1 标注结果存储形式

1	我	我	PN	PN	-	3	Agt	-	-
2	无法	无法	AD	AD	-	3	mNeg	-	-
3	使	使	VV	VV	-	10	Root	-	-
4	这	这	DT	DT	-	5	Sco	-	-
5	部	部	M	M	-	6	Quan	-	-
6	机器	机器	NN	NN	-	3	Pat	-	-
6	机器	机器	NN	NN	-	8	Exp	-	-
7	正常	正常	AD	AD	-	8	Mann	-	-
8	运转	运转	VV	VV	-	3	ePurp	-	-
9	。	。	PU	PU	-	3	mPunc	-	-

兼语句的句法结构要求有两个谓词性词语，两个谓词性词语之间有一个名词性短语，并

且这个名词性词语（即兼语）一定有两个父节点。因此，按照结构存储规则和兼语句的句式特点，我们设置了图 5 所示的兼语句自动抽取算法。

依据此抽取算法，我们在 10038 个句子中共抽取了真正兼语句 297 条，伪兼语句 141 条。

```
Sentence_Verbs[]=Sentence 中所有动词; //将 Words 数组中的所有动词找出;
while(Sentence_Verbs 中的动词数目>=2)
    {V1=FirstVerb(Sentence_Verbs); //取 Sentence_Verbs 中第一个动词作为 V1
    SubNouns_of_V1[]=GetSubNouns(V1); //V1 的所有名词性子节点;
    Candidates_of_V2[]=GetSubVerbs(V1); //V1 的所有动词性子节点,动词 V2 的候选词
    for (i=0;i<Candidates_of_V2[ ]中词的数量;i++)
        {V2=Candidates_of_V2[i];
        SubNouns_of_V2[]=GetSubNouns(V2); //V2 的所有名词性子节点
        if(SubNouns_of_V1[]∩SubNouns_of_V2[] <> Null) //某名词有 V1 和 V2 两个父亲节点
            Sentence 是兼语句;
            break;
        }
    if(Sentence 不是兼语句)
        Delete(V1, Sentence_Verbs[]); //将 V1 从 Sentence_Verbs 中删除
    else
        Break;
}
```

图 5 兼语句自动抽取算法

4.2 兼语句句模构建原则

句模描写具有层级性。层级性的原则是由上到下数目逐渐增多。其中数目最少的当属兼语句中两个动词表示的事件之间的逻辑关系了，所以兼语句中两个动词表达的事件之间的语义关系为第一个层级。第二个层级以兼语句中兼语 N2 所承担的语义角色组合的不同为核心，辅以主语的语义角色和兼语动词 V1，将兼语句中可以充当 V1 的动词分为：使令、命令、劝令、委托、提供、推举、协同、协助、跟随、喜恶和有无类^[32]，此外，这里描写的句模直接忽略了具体句子中的出现的时间、空间和方式等表示附加意义的周边语义角色，因为多数动词都可以任意添加这些语义角色表达相同的语义，只是在句中的位置不同。第三个层级是对典型 V1 的举例。

兼语动词 V1 是“有无类”的兼语句是兼语句中比较特殊的一类，这类兼语句在很多情况下，可以通过句式转换，变成一个普通的句式，并且句子表达的语义不会有很大差别。这种可以转换与不可以转换的“有无类”兼语句之间不存在明显差别；另一方面，“有无类”兼语句表达的两个事件之间的语义关系也不是很明确，只能细细体会到事件主体自然而然变了。因此，我们在这里就将所有的“有无类”兼语句表达的事件关系定义为顺承关系。

4.3 基于语义依存图的兼语句句模

根据对兼语句中两个动词之间语义关系的考察，这两个动词的语义关系可以用四种语义关系来概括：结果、顺承、目的和原因。但是这四种事件语义关系却不是同等重要，即其使用频率存在很大差别。顺承和结果相对而言，使用的更为频繁，原因的使用是最少的。在兼语句中，每个兼语都承担着两种语义角色，从理论角度看，兼语承担的语义角色会形成成百上千种不同的组合关系，但事实却不是这样的。根据真实语料，我们发现在四种事件关系中，

兼语承担的语义角色总共形成了 23 种不同的组合关系。在表示结果、目的、顺承和原因关系的兼语句中，兼语承担的语义角色组合关系分别有 11 种、9 种、15 种和 3 种。其中一些组合关系是至少两种事件关系当中共有的（10 种），还有一些则是某种事件关系当中所独有的（13 种）。部分兼语句句模的具体情况如下表 2 所示。

表 2 兼语句句模

第一层	第二层			第三层	例句	
事件关系	N1 的语义角色	V1	N2 的双重角色	V2		
结果关系	当事/施事/受事	使令、推举类	与事、当事	所有动词和 可以做谓语的 形容词	使、令、让、叫、 选、使得等	他那有力的论点 使得对手哑口无 言。
			与事、空间			孩子们的欢笑使 花园增添了春 意。
			受事、领事			一条新修的公 路，使我家失去 了四季翠绿的菜 园。
目的关系	施事/领事/当事	使令、命令、 委托、和提 供类	与事、受事		使、让、派、请、 催促、命令、劝、 委托等	一定要保护好森 林资源，使失去 的植被尽快恢 复。
			受事、施事			她开心地迫使那 位售货员把橱窗 里的东西都拿了 出来。
			属事、客事			将来我有什么要 教给我的孩子， 我会告诉他。
顺承关系	空间/时间/当事 /施事	有无、伴随 和称呼类	当事、感事		有、没有、无、伴 随、叫、称等	她那么安静，几 乎没有人注意到 她在这里。
			当事、施事			湖里有许多鸭子 在游泳。
			当事、当事			北宋时候，有位 画家叫张择端。
原因关系	施事/感事	喜恶类	与事、施事		喜欢、讨厌、憎恨、 钦佩、埋怨、责怪、 祝贺、谢谢等	谢谢你告诉我这 个消息。
			与事、当事			妻子埋怨他不务 正业。
			与事、领事			我祝贺你终于有 了一个女儿。

以上表中第一个例句为例，V1 为使令类动词“使得”，V2 为普通动词“哑口无言”，兼语 N2 “他”既是 V1 的“与事”角色同时又是 V2 的“当事”角色，句子主语 N1 “论点”是 V1 的“当事”角色，V1 “使得”和 V2 “哑口无言”之间构成原因—结果关系，但是在这种情况下，原因不是十分突出。在这种表示结果关系的兼语句中，那些以没有实在意义的词“使”“让”做兼语动词的兼语句中，都可以把句子主语当做后一事件的缘故成分去解释，完全不会影响句子语义的表达和理解。如句子“一条新修的公路，使我家失去了四季翠绿的菜园。”，表达这个意义的普通句式为“因为一条新修的公路，我家失去了四季翠绿的菜园”。所以在这里就会出现一定的问题，汉语中是否真的存在兼语句，是否能把这种兼语动词处理为虚词还有待进一步的研究。

“使令类”动词既可以出现在表示结果关系的兼语句中也可以出现在表示目的关系的兼语句中。这样一来，可能会造成结果关系和目的关系难以区分，但是这样的表示目的关系的“使令类”动词确实与表示结果关系的“使令类”动词有很大差别。一方面，出现在两种表示不同的事件关系的兼语句中的“使令类”还是存在一些意义上的差别，这种差别可以在句模中反映出来。在表示目的关系的兼语句中，“使令类”动词是具有实在意义的词语，其意义类似于“命令、要求等”，且兼语句的主语通常是它们的施事；而在表示结果关系的兼语句中，“使令类”词语意义比较虚，且句子主语通常是其当事或者省略；另一方面，表示结果关系的“使令类”动词周围通常会出现一些时体标记“了”、“昨天”等。

在表示顺承关系的兼语句句模中，“有无类”兼语句占相当高的比重。一方面，因为“有无类”兼语句并不存在明显的事件关系，只是单纯的顺其自然的转换话题，如在句子“早上六点钟起，就有群众的队伍入场了。”中，第二个事件的话题自然而然转移到“队伍”；并且这种兼语句通常可以转换成普通的存在句，即把这类兼语句中的 V2 代表的整个事件提前到兼语之前，充当兼语的定语。如“村里有一个离奇的故事一直在流传着。”可以转换成“村里有一个一直在流传着的离奇的故事”。另一方面，在那些不能替换和能替换成普通存在句的“有无类”兼语句之间还未能找到合理的科学的解释。如句子“这一夜，没有什么人睡觉，”和“陆地上几乎没有什么东西覆盖它们。”，兼语都是“什么”，一个可以转换，一个却不能转换。

相对于其他三种句模，表示原因的兼语句句模的种类和数量是最少的，但也是最具有区别度的。因为在这种句模中，兼语动词常常是喜恶类动词，并且喜恶类动词不会进入其他三种兼语句句模结构。表示原因的兼语句可以通过转换变成一个表示原因—结果关系的复句。如句子“妻子埋怨他不务正业”，可以转换成“妻子埋怨他，因为他不务正业。”，这个单句和复句在意义上没有任何差别，只是语言形式不同。

除了上述列举的 12 种兼语句句模，本文共总结了 38 种兼语句句模，其中结果关系句模 11 种，目的关系 9 种，顺承关系 15 种和原因关系 3 种。但是由于文章长度限制，在这里，不能展示所有的具体句模，只能在有限的空间内展示尽可能多的具体句模。同时，由于语料规模较小，在大规模语料中可能会出现这 38 种兼语句句模无法覆盖的情况。但即使是这样，这些句模的总结也将更好地帮助构建语义依存关系自动分析模型。

5 结语

鉴于依存分析的优点和汉语语言的特点，本文提出了一种新的句子语义分析理论——语义依存图理论。语义依存图结构有两个不同于语义依存树的特点：1) 多父节点；2) 非投射。此外，两者对语义角色的定义也不同，语义依存图理论定义了一套有系统、分层级的语义关系体系。根据这套理论，我们建立了一个包含 30,000 个句子的语义依存图库。在 10,000 个来自该语料库的句子的基础上，对其中的兼语句（不包括兼语连动连用句）的语义结构进行了总结。这种语义结构的描写是分层次、有结构的描写。第一层是兼语句中两个谓词代表的

事件之间的语义关系：1) 结果；2) 目的；3) 顺承；4) 原因，第二层以兼语承担的语义角色组合关系为核心，辅以主语的语义角色和兼语动词 V1 的动词类，共总结 23 种不同的组合关系、11 种 V1 所属语义类和 38 种具体的兼语句句模，第三层是对典型 V1 的举例。这样对兼语句句模进行描写可以帮助建立句型与句模的映射关系，从而促进句法-语义界面研究和语义研究的发展，同时，也能帮助机器自动生成符合规范句子，促进人工智能的发展。

由于语料规模有限，可能还存在以上兼语句句模不能覆盖的兼语句现象，这一点，还有待于语料规模的扩大。此外，兼语句中第二个谓词表示的事件框架往往与第二个谓词的谓词意义和特点有很大的关系，因此，对第二个谓词的分类研究和进行更加完善全面的兼语句句模描写和结构设计将是下一步工作。再者，研究其他特殊句式的句模和建立汉语句子的句模系统也将是我们未来的研究工作。

参考文献

- [1]孙道功, 亢世勇, 孙茂松.基于标注语料库的现代汉语单句句型句模的对应关系研究[C].全国第八届计算语言学联合学术会议论文集. 2005:234-240.
- [2]孙道功, 亢世勇, 孙茂松.面向语言处理的单句句型句模对应关系研究——基于标注语料库的定量考察[J]. 计算机工程与应用. 2006,33: 170-173.
- [3]黄晓冬.试论兼语短语的语义结构.四川师范大学学报. 2009,36(6):38-43.
- [4]Fillmore,Charles J. and Collin F.Baker. FrameNet:Frame Semantics Meets the Corpus[D].Poster presentation,74th Annual Meeting of the Linguistics Society of America, 2000.
- [5]Palmer,M.,Gildea,D.,Kingsbury,P.The Proposition Bank:A Corpus Annotated with Semantic Roles.Computational Linguistics,2005,31(1):71-105.
- [6] Chang Liu, HweeTou Ng. Learning Predictive Structures for Semantic Role Labeling of NomBank. In Proceedings of ACL 2007: 208-215.
- [7] N. Xue, M. Palmer. Annotating the Propositions in the Penn Chinese Treebank. Q. Ma, F. Xia, (Editors) Proceedings of the Second SIGHAN Workshop on Chinese Language Processing, 2003, 47-54.
- [8]Liping You, Kaiying Liu.Building Chinese FrameNet Database1.Natural Language Processing and Knowledge Engineering, 2005.
- [9]NianwenXue.Annotating the predicate-argument structure of Chinese nominalizations.In Proceedings of the fifth international conference on Language Resources and Evaluation, Genoa, Italy, 2006.
- [10] Min Zhang, WanxiangChe, Guodong Zhou, Ai Ti Aw, Chew Lim Tan, Ting Liu,Sheng Li. Semantic Role Labeling Using a Grammar-Driven Convolution Tree Kernel. IEEE Transactions on Audio Speech and Language Processing 2008, 16(7):1315-1329.
- [11]WanxiangChe, Min Zhang, Ai Ti Aw, Chew Lim Tan,Ting Liu, and Sheng Li. Using a Hybrid Convolution Tree Kernel for Semantic Role Labeling. ACMTransactions on Asian Language Information Processing 2008, 7(4):1-23.
- [12]Xue NW. Labeling Chinese Predicates with Semantic Roles[J]. Computational Linguistics, 34(2):225-255.
- [13]Hacioglu K. Semantic Role Labeling using Dependency trees[C]//COLING04 Proceedings of the 20th international conference on Computational Linguistics. 2004: 1273-1281.
- [14]丁伟伟, 常宝宝. 基于语义组块分析的汉语语义角色标注[J]. 中文信息学报, 2009, 23(5):53-61,74.
- [15]Carreras X, Mrques L. Introduction to the conll2005 shared task: semantic role labeling[C]// Daniel Gildea. CONLL 05 Proceedings of the 9th Conference on Computational Natural Language Learning.Stroudsburg: Association for Computational Linguistics. 2005: 152-164.
- [16]尤昉, 李涓子, 王作英. 基于语义依存关系的汉语语料库的构建[J]. 中文信息学报, 2003,17(1): 46-53.
- [17]B.Chen, D.Ji. Chinese Semantic parsing Based on dependency Graph. In: Proceedings of the First

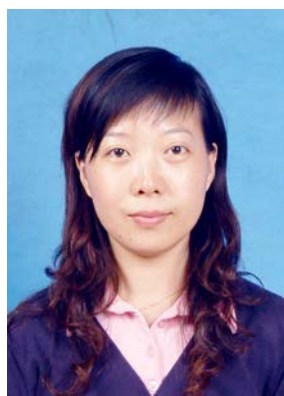
International Conference on Electronic & Mechanical Engineering and Information Technology (EMEIT), 2011:1730-1734.

- [18]范晓.三个平面的语法观[M].北京:北京语言学院出版社.1996.
- [19]范晓,朱晓亚.论句模研究的方法[J].徐州师范大学学报.1999,25(4):18-23.
- [20]朱晓亚.现代汉语句模研究[M].北京:北京大学出版社.2001.
- [21]徐昌火.试论句模研究的对象、起点和基本原则[J].南京师范大学学报.1999,(4):101-108.
- [22]鲁川.汉语语法的意合网络[M].北京:商务印书馆.2001.
- [23]田珍都.基于标注语料库的现代汉语句模研究[D].烟台师范学院.2002.
- [24]亢世勇,许小星.现代汉语句系系统的构建和研究[J].中文信息学报.2010,(1):103-109.
- [25]黄伯荣,廖旭东.现代汉语(第四版)[M].北京:高等教育出版社.2007.
- [26]许小星.基于标注语料库的现代汉语特殊句式语义分析[D].烟台:鲁东大学.2007.
- [27]Robinson,J.J.Dependency Structures and Transformation Rules[J].*Language*.1970,46(2):259-285.
- [28]郑丽娟,邵艳秋,杨尔弘.中文非投射语义依存现象分析研究[J].中文信息学报.2014,28(6):41-47.
- [29]Qiang Dong, Zhendong Dong.*Hownet and Computation of Meaning*.World Scientific Publishing Company.2006.
- [30]冯志伟.中文信息处理与汉语研究[M].北京:商务出版社.1992.
- [31]袁毓林.基于认知的汉语计算语言学研究[M].北京大学出版社,2008.
- [32]游汝杰.现代汉语兼语句的句法和语义特征[J].汉语学习.2002,(6):1-6.

作者简介: 郑丽娟(1991—),女,硕士研究生,主要研究领域为计算语言学。Email: lijuanzhengzai@126.com; 邵艳秋(1970—),通讯作者,女,博士,北京语言大学教授,硕士生导师,主要研究领域为计算语言学。Email: yqshao@blcu.edu.cn;



郑丽娟



邵艳秋