

基于标注语料库的情景语义成分分析

刘金凤

(鲁东大学汉语言文学学院, 烟台 264025)

摘要: 语义成分的研究在一定程度上已有显著发展,但仍然存在一些问题。本文基于较大规模中小学语文课本语料库,对情景语义成分进行分析研究,考察语义成分与句法成分的对应关系,以期深化语义成分研究,并构建一个较为适用的语义成分体系。

关键词: 标注语料库; 语义成分; 情景; 分析

The Analysis of Scene according to Tagged Corpus

LIU Jinfeng

(The Linguistics and Literature Faculty of Ludong University, Yantai 264025)

Abstract: The research of the semantic components has received great development in some extent, but there are still many questions. Based on the tagged corpus, this article will analyze and investigate the scene and review the corresponding relation between the semantic components and the syntax components, in order to deepen the research of the semantic components, and found a semantic component system in point.

Keyword: the tagged corpus; semantic component; scene; analysis

汉语是一种缺乏形态变化的语言,意义的研究显得颇为重要。从上个世纪语义学引入我国,国内的学者在意义研究方面做了大量工作,并取得了可喜的成就。但迄今为止,语义成分的研究还存在不少问题,还有深入研究的必要。本文以较大规模标注语料库为基础,在形式和意义相结合、静态研究和动态研究相结合的原则指导下,采用定量和定性相结合的方法,对现代汉语语义成分的一个大类——情景进行研究分析,以期深化语义成分的研究,并试图建构一个较为适用的语义成分体系。

我们选取的语料来自中小学语文课本,以期通过对语料的标注和研究,对中小学语文教学有所裨益,使学生深刻把握句子规律,进而提高汉语教学水平,对对外汉语教学也有所启示;并为大规模文学语料的处理研究提供借鉴,促进计算机对句子的分析、理解和生成。

1 情景概说

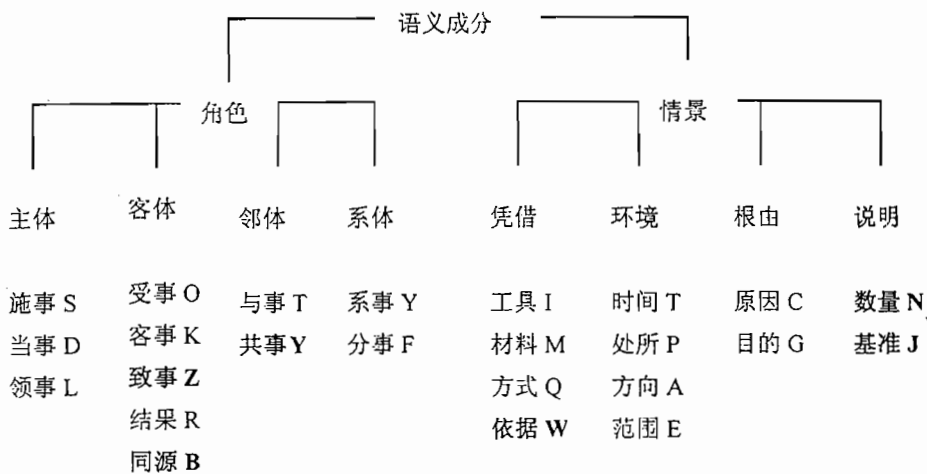
1.1 语义成分系统

对于语义成分种类的确定,笔者主要参考了鲁川、林杏光、范晓和陈昌来几位学者的分类系统,通过比较,选择以林杏光先生的格系统(下面简称林系)为基础,结合所标语料的实际情况,稍作调整,把语义成分定为24小类,其中改同事为共事,增加致事、同源两类。

经过分析和研究,最后,确定的语义成分系统如下:

基金资助: 国家社科规划项目“基于大规模标注语料库的现代汉语句子语义结构系统研究”(05BYY029)

作者简介: 刘金凤(1984-),女,山东日照,2002级本科 lingzhi5202001@126.com



1.2 情景概说

“情景”，《现代汉语词典》解释为“(具体场合的)情形，景象”，在本文指的是语义成分系统的一级分类，最早出现在林系中，“整个格系统有三个层次：第一个层次是角色和情景，”情景下又包含三个要素：凭借、环境和根由共 10 个格。[1]林先生只是提出了“情景”这一说法，但没有进行界定，我们沿用这一术语，并试图对它作出较为恰当的表述。

“情景”的英文单词我们译为“scene”，这与菲尔墨格语法中的“scene”有所不同。菲尔墨的“scene”指场景，即 S=M+P 中的“S”，指“语言之外的真实世界，如物体、事件、状态、行为、变化，以及人们对真实世界的认识、感觉、知觉等”，“句子描述的是场景”。[2]而我们所指的是 P 命题的组成部分，是一部分语义成分的总称，类似于贾彦德先生提出的“一般格”。贾先生指出“有些语义格和谓词或者和其他的项结合，但不与谓词构成句义的框架，而只是说明、描述这一框架。这样的语义成分叫做一般格。”[3]我们的“情景”与此也有差别，最大的不同在于：不包括“其他（属格和描写格）”一项。

根据比较和分析，我们这样界定“情景”：语言学术语，是语义成分系统的一级分类，与“角色”相对。指与谓词结合，一般只起说明、描述作用的语义成分。在我们的语义成分系统中，情景包含凭借、环境、根由和说明四个要素，共有工具、材料、方式、依据、时间、处所、方向、范围、原因、目的、数量和基准 12 个语义成分。限于篇幅和能力，笔者在十二种情景语义成分中选取了记录数在 100-200 之间的四个进行分析，分别是：基准、方式、方向和工具。

2 情景分析

2.1 基准

基准是事件中进行比较或测量所参照的对象。据语料统计，基准可以处于补语、宾语、状语三种句法位置上。

基准作补语数量较少，仅得两例，并且都是文言句式。语料中只有一例是基准作宾语的情况：饮食起居亦不如[O 前]J，这是由语料选取的特殊性决定的。实际生活中，基准宾语还是很多的，如：学习上，我不如他。|我吵不过他。基准更多的是充当状语，这在语料中占有极大比例。基准状语都有介词引导，“比”使用最多，另外还有“与、跟、和、同”等。“比”有明显的比较意味，而“与、跟、和、同”等介词引入的成分算作基准，很大程度上取决于后面的谓语，这些谓语本身就含有比较的意思，如：陕西[D 同山西]J，不是差不多吗。

有人将“园子荒芜冷落得如同一片野地”、“我被调养得像千金小姐”等句中“如同、像”引导的补语看作是基准，我们认为这样是不合适的。首先，这些补语都表程度或结果，无比较义；更重要的是，此处的“如同、像”都是动词，宜看成是动宾结构作补语，应在其内部进行进一步切分标注，而不是作为基准处理。基准状语也有类似的问题，要注意区分。

2.2 方式

方式，是事件中所采用的方法或形式，多是抽象的事物。

在语料中，方式只处于状语一种句法位置上。按照词语的特点，可以分为以下几类：

(1) 介词类 作方式形式标记的介词有“以”和“用”两个，如：也[D 以同样的耐心和方式]Q2 想过我为什么要出生 [D 用这样的方法]Q，寄托我们的哀思。从例句中可以看出，充当方式成分的词语有这样的特点：都是以方式、方法等名词为中心的偏正结构，“这些成分的特点正是方式的‘抽象性’决定的”。[4]

(2) 名词类 语料中“轻声、低声、小声、大声、高声、尖声”以及“大步、大口、真心”等词语，不带形式标志而直接充当状语，表示方式。除了上述几个外，还有“口头、书面、正面、手工”等。这类词的特点是表示动作行为的某种方式，因为比较简短，可以不带形式标志；结构上，都是单音节形容词名词或形容词性名词性语素相结合。典型例句如：她[D 轻声]Q 与丈夫谈话。

(3) 代词类 方式状语以代词为最多，约占 67%。可以分为两类：1、指示代词 这样、这么、那样；2、疑问代词 怎样、怎么、如何。其中“这样、怎么”出现频率最高，分别为 13 处、11 处。

(4) 其他 [D 无意中]Q 回头一看 |一对令人羡慕的中年情侣[D 不觉中]Q 成了两个老人 [D 潜意识下]Q 并且还觉得有点儿落寞。这些句子中的“无意中”“不觉中”“潜意识下”等也都是方式状语，数量较少，规律性也不大，所以归为一类。

另外，方式还可以作宾语，只是语料中没有例句，如：唱高音、画水彩、抄仿宋体等，据统计能带方式宾语的动词约有 52 个[5]。

2.3 方向

方向是事件中的时空趋向，可能出现的形式标记有“向、朝、往”三个，它们引进动作行为的方向，具有很强的时空趋向性。方向可以处于补语、宾语、状语三种句法位置上。

语料中的方向补语都由介词“向”引导，而且只能由它来引出，不能换成“往”或者“朝”。表示方向的成分有方位词（如“西”）、处所词（如“窗外”），还有普通名词“大海、笼子”等，限制不大。句子的谓语都是光杆动词，无论是自动词“奔、飞、飘、退”还是他动词“卷、转、摔”都如此，在他动词作谓语的句子里，主语是受事。这些句子的补语可以提前转换为状语，但是谓语动词后必须加趋向动词“去”。

方向宾语都是方位词，带方向宾语的动词最典型的是“向”，除此之外，还有“朝”等，数量很少。如：天空出现一匹马，马头向[O 南]A2，马尾向[O 西]A3。

方向状语可以分为两类：有形式标记的和无形式标记的。方式状语的形式标记有“往、朝、向”三个，通过分析，可以看出，方向状语更多的由方位词（短语）来充当，“朝”后更是只能跟方位词。通过方位词和前面的形式标记，我们可以很容易地判断出方向状语。

“往、朝、向”作为方向语义成分的形式标记在使用的过程中，并不是可以随意替换的，它们的用法存在一些细微的差别：1、“往”的义项为“引进动作行为的方向”，它后面的成分全部是方向；2、“朝”和“向”的义项是“引进动作的方向、对象”，它们后面的成分不完全是方向，要注意区别；3、由“朝”引导的介宾结构只能作状语，“向”引导的可以做补语。像“向一匹猿尽力刺去”、“朝他看”这两个句子中“向、朝”引导的成分就不是方向，而是动作的对象，是受事。

2.4 工具

工具是事件中所用的器具，通常是无生命的，即具有无生性。典型的标记是“用”，另外还有“以、将、拿”等。工具在句子中可以处在状语、宾语和主语的位置上，以作状语为主，约占 86%。

语料中，引导工具作状语的形式标记有“以、用、将、被”四个，其中“用”使用最多，占 83%左右，“将”有两个例句，“被”只有一个，如：要是[D 以这园子里的声响]I 来对应四季呢 |黑熊只好[D 用舌头]I 舔着自己又肥又厚的脚掌 |我[D 将他给我做的紫毛大衣]I 铺好坐位 |[D 将两个指头的长指甲]I2 敲着柜台|当时的我只是[D 被一件破蓝布袄]I 草草地包裹着。“将、被”一般不引导工具，而将上面例句的状语算作工具是根据句义判断的，这样的用法很少。

工具宾语在语料中只有三个句子：美国的橘子包着[O 纸]I |补上[O 一块新橡皮]I |里外包着[O 皮]I3。有人统计，能带工具宾语的动词约有 97 个[5]。像“乘三小时汽车”“拄着拐杖”“用这名字”“说话用嘴”等句中的宾语，虽然在日常生活中，它们都是某种工具，但在句子里它们并不是工具语义成分，而是动作支配的对象，是受事，典型的的就是“用”“使”和“V 用”类动词。

有人否认工具主语的存在[4]，认为下面的都是施事主语句：一只脚踏在石凳上 |眼睛都望着碟子 |言语是不够表现我的心情的 |棒打狍子瓢舀鱼。实际上，上述句子中的主语都不能自主地发出动作，施事是事件中白发动作行

为的主体，包括人和自然力两类，很显然，它们与施事的定义不符。这些主语前面都可以加上“用”来作状语，当然有的句子需要稍加调整，如“眼睛都望着碟子”可以变成“都用眼睛望着碟子”。

3 语义成分和句法成分的对应关系

语义成分是语义结构的组成成分，句法成分是句法结构的组成成分，底层的语义结构都会映射为句法结构，这时，每个语义成分都会出现在一定的句法成分的位置上。

通过上面对四种情景语义成分的分析，我们可以得出下表：

句法位置 语义成分	状语	比例 (%)	补语	宾语	主语
基准	+	98	+	+	
方式	+	100			
方向	+	90	+	+	
工具	+	86	+		+

从表中可以看出：

- 1、除方式外，基准、方向和工具均可以处在多种句法位置上；
- 2、所有四种情景语义成分都以状语为主要句法位置。

在主语、宾语、状语和补语四种句法位置上，最为活跃的是状语，这从一个侧面说明状语位置上对情景语义成分的限制条件比较少。

语义成分和句法成分属于两个层面，它们并不是一一对应的关系，一个语义成分可以对应多个句法位置，一个句法位置同样可以对应多个语义成分。那么，如何判断一个语义成分在什么条件下对应为某一句法成分？我们的依据主要有三个：形式标记、动词和句义，它们不是并列的，优先选择的顺序为：形式标记>动词>句义。

四种语义成分的典型形式标记依次为：基准——“比”；方式——“以、用”方向——“往、向、朝”；工具——“用”。根据这些形式标记，基本可以做出判断。但是存在的问题是有的形式标记具有多义性，这就需要引入动词标准，依据动词的语法特点和词汇意义来判断，如果在这样的情况下还不行，就求助于语义的分析。经过这样的分析过程，一般不会再有判断困难。

参考文献：

- [1] 林杏光. 词汇语义和计算语言学[M]. 北京：语文出版社，1999. 185-186.
- [2] 俞如珍，金顺德. 当代西方语法理论[M]. 上海：上海外语教育出版社，2000. 258-276.
- [3] 贾彦德. 对现代汉语语义格的认识和划分[J]. 语文研究，1997（3）： 23-29.
- [4] 陈昌米. 现代汉语语义平面问题研究[M]. 上海：学林出版社，2003. 233-234, 200-209.
- [5] 黄伯荣. 动词分类和研究文献目录总览[M]. 北京：高等教育出版社，1998. 237, 232.