

面向深层语义表示的否定定义表达规律探析*

邱立坤, 黄焜, 何保荣, 亢世勇

(鲁东大学文学院, 山东省 烟台市 264025)

摘要: 否定定义是深层语义表示中的一个重要组成部分。本文基于语料库的方法对现代汉语中的否定表达形式及其使用规律进行了深入分析。首先, 系统地收集了否定表达形式, 将之分为显性否定词、隐性否定词、否定结构三类, 并讨论了否定表达形式的非否定用法。其次, 对否定表达形式的使用规律进行了归纳与总结, 涉及单动核结构、情态成分、述补结构、动词性并列结构、连谓结构、兼语结构等, 重点分析多动核结构中否定对命题义的影响, 并总结了在深层语义标注框架下否定定义的标注规则。最后, 基于多领域句法树库考察否定表达形式的领域分布差异。

关键词: 现代汉语; 抽象语义表达; 否定词

中图分类号: TP391

文献标识码: A

Analysis of Negative Expressions of Contemporary Chinese for Deep Semantic Representation

QIU Likun, HUANG Kun, HE Baorong, KANG Shiyong

(School of Chinese Language and Literature, Ludong University, Yantai, Shandong 264025, China)

Abstract: Negative expression plays an important role in deep semantic representation. Using corpus-based methods, this paper focused on analyzing negative expressions and their usages in contemporary Chinese. First, we collected negative expressions and classified them into three types, i.e. explicit negatives, implicit negatives and negative constructions, and then discussed their non-negative usages. Second, we analyzed the rules of negative expressions, especially such negative expressions used in modifying single-predicate structures, modality elements, predicate-complement structures, verbal coordinate structures, serial verb structures and pivotal sentences, and focused on discussing the effect of negative expressions in multi-predicate structures upon the meanings of propositions. The rules of annotating negative meaning under the schema of deep semantic representation are also summarized. Finally, we investigated the distribution of negative expressions in multi-domain treebanks.

Key words: contemporary Chinese; abstract meaning representation; negative words

1 引言

抽象语义表达 (Abstract Meaning Representation, AMR) 的目标是表示语言所反映的“谁对谁做了什么 (who is doing what to whom)”之类的信息^[1], 并且把由于句式变换而产生的不同句法结构所表达出来的相同语义信息等同起来。与浅层语义分析 (即谓词论元结构分析或语义角色标注^[2]) 相比, AMR 所表达的信息更为深入和全面, 因而属于深层语义表示。除了浅层语义分析中所表示的谓词论元关系之外, AMR 还试图描述模态义、否定定义、时体、事件关系等等。本文将针对深层语义表示的需要, 具体分析现代汉语否定定义的表达规律。

汉语学界对否定词的研究主要集中在否定辖域、否定焦点及个体否定词三个方面^[3-13],

* 收稿日期:

定稿日期:

基金项目: 国家社科基金重大项目 (12&ZD227); 国家自然科学基金项目 (61572245, 61272215, 61103089); 鲁东大学人文社会科学研究项目 (WY2013003)

但是较少对否定表达形式进行全面系统的梳理,也很少考虑否定定义与命题义的关系问题。逻辑学界对否定定义的关注度较高,主要在简单命题和模态命题的层面上考虑否定命题^[14],这些命题对应的是语言中的单动核句。对于多动核句则较少涉及,而多动核句却正好是深层语义分析中的难点所在。

在前人研究的基础上,我们针对深层语义表示的需要,对现代汉语的否定形式和否定定义的表达规律进行了全面分析。主要工作包括以下三个方面:(1)全面收集整理现代汉语中的否定表达形式,并分析这些否定表达形式在什么情况下表示否定定义。(2)分析否定定义表达的规律,尤其是述补结构、动词性并列结构、主谓短语做宾语的句子、连谓句和兼语句等多动核结构中的否定表达问题,考察否定与多动核结构对应的多个命题之间的关系,并总结深层语义表示中各种句法结构中否定定义的标注规则。(3)基于多领域树库,分析否定表达形式的领域差异,并且通过分布数据说明绝大多数否定定义的标注可以通过标注规则自动进行,仅有少量需要人工介入。

2 相关研究

目前,语言学界进行否定研究的重点集中于以下三个方面:否定辖域研究、否定焦点研究、否定词的研究。

首先是否定辖域和否定焦点方面的研究,这两个方面的研究常常交叉进行。吕叔湘(1985)^[3]指出一个句子的否定辖域是否定词“不”“没”后的全部词语,而这个否定句的焦点主要指的是句子末尾的成分,即句末重音的部分,这是否定辖域最初的一种观点。之后徐杰、李英哲(1993)^[4]提出否定是非线性语法范畴,它没有独立的否定辖域和焦点。袁毓林(2000)^[5]则认为否定在其表层结构上是一种线性语法范畴,否定有其独立的辖域和焦点,并且他还从自然语言的非单调性角度指出了否定句并没有辖域的歧义,也不存在语义模糊。胡建华(2007)^[6]主要探讨了“不”的语法特征,他指出“不”的辖域是其统辖的成分VP,否定词“不”在句法上否定“VP”及“VP”的中心语,而不是对焦点和VP的毗邻成分和否定词的毗邻成分进行否定。

其次是关于否定词的研究,前期研究主要集中在否定词“不”上,邢福义(1982)^[7]、沈开木(1984)^[8]、吕叔湘(1985)^[3]、李瑛(1992)^[9]、史锡尧(1995)^[10]等人针对否定词“不”从各个角度进行研究,例如“不”的语法功能、语义功能及其统辖范围和分布位置等。后期研究中,特殊否定词(如“差点儿”)引起学者们的注意,石毓智(1993)^[11]、周一民(2003)^[12]、袁毓林(2012)^[13]分别对含有“差点儿”“差点儿没”的句子进行了研究。

3 深层语义表示与否定定义

在深层语义分析中,每个句子可以表示为一个有根(rooted)、有向(directed)、非循环(acyclic)图(graph),图中的边和叶子结点上都标注有标签(分别为概念间关系和概念的语义类)。图1和图2是类AMR的深层语义表示示例,原始句子为“王强没有去超市买铅笔”。这个句子包含了两个命题“王强去超市”和“王强买铅笔”,前者一定为否定命题,后者则存在歧义:可能是否定命题,也可能是肯定命题。

图1表示的是第一种情况,即后一个命题是否定命题¹。其中,“买”是谓词,“王强”、“铅笔”和“没有”均是其子结点,其中“王强”是施事角色,“铅笔”是系事角色,“没有”则表示否定。此时,这个句子表示的意思是“王强没有去超市,也没有买铅笔”。

图2表示的是第二种情况,即后一个命题是肯定命题。此时,这个句子表示的意思是“王强没有去超市,但买了铅笔,铅笔不是从超市买的”。

理论上,还可能有第三种情况,即“王强去了超市,没有买铅笔”,但事实上这种意思应该由“王强去超市没买铅笔”这个句子来表示,即否定词直接否定连谓结构的后项。

¹ 在深层语义表示中,否定词的辖域是以否定词的父结点为根结点的子树。比如,在图2中,“没有”的辖域为以“去”为根结点的子树。

在深层语义表示中，句法层面的一个成分可能出现多次，成为多个成分的子结点。比如图 1 中“王强”和“没有”均成为两个谓词（“去”和“买”）的子结点。这种现象在句法层面上通常涉及到多动核结构，正好是深层语义分析的重点和难点所在。下文在分析否定定义的表达和标注时，也将以多动核结构作为重点。



图 1 深层语义表示示例 1



图 2 深层语义表示示例 2

4 否定表达形式

现代汉语中否定表达形式有多种，大致可以分为显性否定词、隐性否定词和否定结构三类。前两者都属于词，两者间的区别在于前者直接表示否定，后者则以一种间接或委婉的形式表示否定。否定结构则大于词，可能是两个以上词组成的短语，也可能是两个以上词组成的框式结构。在前人研究的基础上，我们整理出的现代汉语否定表达形式如表 1 所示。

表 1 现代汉语否定表达形式总表

显性否定词	隐性否定词	否定结构
不、非、匪、弗、未、无、没、靡、莫、罔、无、别、免、毋、勿、休、甬、不是、从不、从未、毫不、绝不、决不、决非、确非、永不、不复、不再、未能、无须、没有、并未、未曾、未尝、未经、不要、切莫、请勿、不用、甬想、休想、不必、不曾、并非、无法	差点儿、险些、几平、防止、避免、拒绝、否认、假装、难以	不得不、不能不、差点儿没、险些没、几乎没、非……不可

4.1 显性否定词

显性否定词可以进一步分为单纯否定词、复合否定词、古汉语遗留下来的否定词。在描述否定词的词类时，我们均采用《北京大学现代汉语语法信息词典》^[15]中的词类，词类标记“a、d、v、p、t、b、r、h”分别表示形容词、副词、动词、介词、时间词、区别词、代词和前接成分。

单纯否定词根据否定定义的不同可以分为五类：

- (1) 否定主观意愿类：不(d)、莫(d)、勿(d)；
- (2) 否定性质类：非(d, h, v)；
- (3) 否定变化类：未(d)、没(d)、没有(d)；
- (4) 劝阻禁止类：别(d)、甬(d)；
- (5) 否定存在或具有类：没(v)、没有(v)、无(v)。

复合否定词根据其构成可以分为七类：

- (1) 时间副词+否定词：从不(d)、从未(d)、永不(d)
- (2) 否定词+时间副词：未曾(d)、未尝(d)、未经(d)、不曾(d)、何尝(d)、何曾(d)

- (3) 否定词+频率副词：不复(d)、不再(d)
- (4) 数量词+否定词：毫不(d)
- (5) 连词+否定词：并未(d)、并非(v)
- (6) 否定词+助动词：未能(v)、无须(d)、不要(d)、不用(d)、不必(d)、甬想(v)、妄想(v)、休想(v)
- (7) 其它：绝不(d)、决不(d)、决非(v)、确非(v)、切莫(d)、请勿(v)、无法(v)

古汉语否定词包括“匪、弗、靡、罔、免、毋、休”。这些否定词在现代汉语中都较少单独使用，其否定用法通常只保留在某些成语或固定短语之中，比如“受益匪浅、无远弗届、靡不有初”等。

4.2 隐性否定词

多数学者均将“差点儿”视为否定副词。为了区分，在这里我们采用了袁毓林先生的观点^[13]，称之为隐性否定词。和“差点儿”相似的还有“险些”“几乎”等，《现代汉语词典（第六版）》分别释义如下^[16]：

差点儿(d)：表示事物接近实现或勉强实现。

险些(d)：差一点儿（发生不如意的事情）。

几乎(d)：表示十分接近，差不多。

从中可以看出三者的释义是相似的，都表示“事情接近发生，但是没有发生”。这三个词只是间接否定，而不是像“不”“没有”那样直接否定，所以称之为隐性否定词是合适的。比如“小红差点儿就放弃了比赛”，表示小红没有放弃比赛；“小明险些就撞到了柱子上”，表示小明没有撞到柱子上；“他几乎就要完成任务了”，表示他没有完成任务。

除此之外，袁毓林（2012）^[13]提出，现代汉语中本身含有否定意义的动词也应归为隐性否定词，包括“防止”“避免”“拒绝”“否认”“假装”“难以”等等。例如“老板假装不了解他”就是“老板了解他”，“假装”否定了“老板了解他”这个命题；“小红难以相信这件事儿”就是“小红不相信这件事儿”，“难以”否定了“小红相信这件事儿”这个命题。

4.3 否定结构

否定结构主要包括双重否定结构和表达肯定义的特殊否定结构。

双重否定结构可以分为“否定词+反问语气”和“连用两个否定词”两种类型，双重否定结构可以视为一种否定结构对另一种否定结构的管辖。例如“我们不是完成这项工作了吗？”和“我们不得不离开上海”分别属于这两种类型。

表达肯定义的特殊否定结构，又称为“异形同义格式”^[17]，这种结构中，否定词可以出现也可以不出现，但整体上都表示肯定义。例如“好热闹=好不热闹（热闹）”“好容易=好不容易（容易）”“当心摔跤=当心别摔跤”“难免出错=难免不出错”。

上述否定结构本质上都表示肯定，但是其中包含了否定词，在标注时需要予以特殊处理。

4.4 否定表达形式的非否定用法

需要说明的是上述词语和结构在现代汉语中并不一定表示否定，它们的非否定用法可以分为三种情况。

第一种情况，有些词兼属多个词类，只有部分词类是表达否定义。比如，“别”兼属动词和副词，只有充当副词的“别”才表示否定义。

第二种情况，有些词在同一个词类下有多个义项，只有部分义项表示否定义。比如“没”在充当动词时，有两组义项，第一组义项与“没有”相同，表示否定；表示第二组义项时读作 mò，不表示否定。

第三种情况，本文所述的否定结构本质上均表示肯定义，因此在处理时可将之作为一个整体予以理解，从而可视为否定表达形式的非否定用法。比如“不得不”整体上表示肯定，“非……不可”这样的框式结构，整体上也表示肯定义。

第一种情况比较容易区分,只需要在识别否定词时根据词类予以限制即可。第二种情况则比较复杂,需要通过词义消歧方能解决。不过,第二种情况涉及的词比较少(主要是动词“没”),而且动词“没”的两组义项中,后一组义项也较为少用,对整体影响不大。

为表述方便,下文统一使用否定词来指代各种否定表达形式。

5 否定词使用规律

现代汉语中,否定词可以出现的句法环境包括两大类:(1)在单动核结构、述补结构、动词性并列结构、主谓短语做宾语的句子、连谓句、兼语句、名词短语等结构中充当修饰成分,以及充当情态成分的修饰成分;(2)直接后接宾语充当谓语中心语。

单动核结构指的是谓语部分只含有一个动词或形容词的结构,通常只形成一个命题。与之相对的是多动核结构,有两个或两个以上的动词或形容词,通常可以形成两个或两个以上的命题,现代汉语中述补结构、动词性并列结构(包括形容词性并列结构)、主谓短语做宾语的句子、连谓句、兼语句等均属于多动核结构。

需要说明的是,一个小句可能同时包含多种句法环境。比如,单动核结构与多动核结构中均可能出现情态成分,如果否定词直接修饰情态成分的话,则它与动核没有直接关系,我们只须将之处理为修饰情态成分即可。又如,多动核结构中,否定词可能仅修饰其中的一个动核,与其它动核没有直接关系,此时只须将之处理为修饰单动核结构即可。

5.1 否定词充当修饰成分

5.1.1 单动核结构

一般而言,由充当修饰成分的否定词形成的否定可以分为三种情况:否定单动核结构,否定多动核结构中的某一个动核,否定整个多动核结构。前两种情况是基本一致的。

否定词在修饰单动核结构时可以占据两类句法位置:一类是出现在动核之前、主语之后,例如“花没有开”,否定所形成的命题是“花开了”这一命题的否定命题;另一类是出现在主语之前,例如“并非人人都是经济动物”,所形成的命题是“人人都是经济动物”这一命题的负命题。

单动核结构可以进一步分为动词性动核和形容词性动核。少数否定词(比如“非”)还可以直接修饰名词,用法上与修饰单动核结构类似。区别在于,“非+名词”整体上相当于区别词,通常充当定语。

5.1.2 模态成分

所谓模态成分指的是表达认识情态、能力情态等的助动词(能愿动词),可以分为可能类、必要类和意愿类三类(表2)。没有情态成分的单动核结构和多动核结构一般形成简单命题,加上模态成分之后则形成模态命题。对模态命题的否定通过在模态成分前添加否定词的形式予以实现。例如“我能去上海”的否定形式是“我不能去上海”,其中否定词“不”直接修饰助动词“能”。

表 2 否定词修饰情态成分示例

表示可能	表示必要	表示意愿
我不能去上海。	我们不应违背家人对我们的期望。	他不愿参加这场宴会。
我不可能去上海。	他不得擅自改变自己的学号。	他不想参加这场宴会。
不可直接食用。	我们不该这么做。	他求她不要离开上海。
他不会打我的。	我们不应该这么做。	他不肯遵循学校的规章制度。
他不能够完成这项任务。	我们本不应当错过这场表演的。	他不敢违背学校的规章制度。
这里不可以吸烟。		

5.1.3 述补结构

在述补结构中,否定词可以占据两种句法位置。第一种是修饰述语,形成“主语+否定副词+谓语+补语”的格式,例如“我没有吃完饭”。第二种是修饰补语,形成“主语+谓语+

否定副词+补语”的格式，例如“我吃不吃饭”。这两种格式中都存在两个简单命题，前一种格式中的两个命题是“我吃饭了”和“我吃完饭了”，后一种格式中的两个命题是“我吃饭”和“我能把饭吃完”。

在前一种格式中，补语为结果补语，否定词否定的是动作的结果，说明动作“吃”的结果“完”没有发生。否定词只对两个简单命题中的后一个命题进行了否定，前一个命题仍然是肯定命题。在后一种格式中，否定词否定的是动作的可能性，因此形成一个否定模态命题，相当于“我不能把饭吃完”。否定词同样只对两个简单命题中的后一个命题进行了否定。

两种格式中形成的否定命题的区别在于，前者否定的是结果，形成否定关系命题；后者否定的是模态，形成否定模态命题。

5.1.4 动词性并列结构

动词性并列结构通常存在两个或两个以上的命题，并列结构中的每个并列成分分别与主语和宾语形成命题。在动词性并列结构中，否定词可以占据两种句法位置。前一种是修饰整个动词性并列结构，例如“他从来不请客送礼”。后一种是仅修饰并列结构中的一个并列成分，例如“他不喜欢北京却喜欢上海”。

在前一种句法位置上，否定词对两个命题分别进行否定。在后一种句法位置上，否定词则只对它直接修饰的那个并列成分进行否定。

5.1.5 主谓短语做宾语的句子

在主谓短语做宾语的句子中，否定词占据的句法位置可以分为两类：直接修饰句子谓语的的中心语，直接修饰充当宾语的主谓短语。此类句子通常包含两个命题，这两个命题的谓词分别由句子谓词和充当宾语的主谓短语中的谓词充当。例如“我们不知道他会来”中，第一个命题的谓词是“知道”，第二个命题的谓词是“来”。

对于有些句子而言，这两个句法位置上出现否定词时的意义不一样。比如“我们不知道他会来”跟“我们知道他不会来”意思是不一样的。对于另外一些句子而言，这两个句法位置上出现否定词时的意义基本一致，比如“我们不认为他会来”跟“我们认为他不会来”意思是一样的。

就目前的观察而言，两种情况之间的区别主要是由句子的谓语动词导致的。前一类情况的动词包括“知道、理解、羡慕、害怕、庆幸”等等，后一类情况的动词包括“相信、认为、希望、感觉、打算”等等。

5.1.6 连谓句

连谓句通常也会包括两个或以上的命题。比如“王强去超市买铅笔”这个句子包含两个命题：“王强去超市”，“王强买铅笔”。

在连谓句中，否定词出现在整个连谓结构前时，所表示的意义可能是两种。

前一种情况中，否定词修饰整个连谓结构（比如“王强没有去超市买铅笔”），其意思是“王强没去超市，也没有买铅笔”，对连谓结构的前后项均进行了否定。

后一种情况下，否定词仅修饰第一个谓词性成分。比如，“王强没有去超市买铅笔”这个句子可能表示“王强没有去超市，但买了铅笔”，仅否定前项。又如“不采用污染较轻的改良技术炼焦”中“不”仅对“采用污染较轻的改良技术”予以否定。

但通常情况下，区分这两种意义有一定的难度，有时需要综合考虑上下文予以判断。比如“不利用职权谋取私利”属于否定词修饰整个连谓结构，“不采用污染较轻的改良技术炼焦”则属于否定词修饰连谓结构的前项。

5.1.7 兼语句

兼语句通常也包含两个命题。在否定词修饰兼语结构时，一般情况下仅否定前项，对后项不置可否。例如“老板没有派他去上海”、“我们没有授权农业部制定实施办法”分别否定了“老板派他”、“我们授权农业部”这两个命题，至于“他有没有去上海”和“农业部有没

有制定实施办法”则难以判断。

此外，在使用“不”否定兼语句时，直接修饰兼语结构与直接修饰兼语结构的后项在意义上是一致的。例如“爸爸不让儿子打别的小朋友”和“爸爸让儿子不要打别的小朋友”都是命题“爸爸让儿子打别的小朋友”的否定。这一点与连谓句和并列结构的否定不同。

5.2 否定词充当谓语中心语

有些否定词可以直接充当谓语中心语，比如“无、没有”。否定词直接充当谓语中心语可以根据宾语的不同分为两种类型：带名词性宾语、带谓词性宾语。只能带名词性宾语的否定词主要有“无、没、没有”，只可带谓词性宾语的否定词主要有“假装”，既可带名词性宾语也可带谓词性宾语的否定词则包括“拒绝、否认、防止”。

5.3 深层语义表示中的否定定义标注

由上述分析可知，在深层语义表示中，单动核结构、情态成分的否定比较容易表示，分别为单动核结构的动核和情态成分添加一个表示否定定义的子结点即可。多动核结构的否定定义表示则比较复杂，一个多动核结构可能对应多个命题，对整个多动核结构进行否定时各个命题是否标注否定定义需要区别对待。具体标注方式如表 3 所示。

表 3 多动核结构中否定的使用规律及其标注策略

	修 饰 范 围	例 句	否 定 前 项 还 是 后 项	与 直 接 否 定 后 项 是 否 相 同	标 注 方 式
并列结构	整体	他从来不请客送礼	同时	/	分别标注前后项
述补结构	整体	他没打碎杯子	后项	不同	仅标注后项
	后项	他打不碎那个杯子	后项	直接否定后项，后项形成否定模态命题	仅标注后项
主谓短语做宾语的句子	整体	我不认为他应该完成这项工作	后项	相同	仅标注后项
	整体	我们不知道他会来	前项	不同	仅标注前项
连谓句	整体	妈妈没有上街买菜	同时	/	分别标注前后项
	前项	不采用污染较轻的改良技术炼焦	前项	/	仅标注前项
兼语句	整体	老板没有让小王去上海	前项	不同（没有相应的直接否定后项的用法）	仅标注前项
	整体	老板不让小王去上海	前项	相同	仅标注前项

其中，并列结构和连谓结构的第一类同时对前项命题和后项命题标注否定定义，述补结构仅对后项命题标注否定定义，其它结构则仅对前项命题标注否定定义。需要特别说明的是：（1）述补结构中否定词直接修饰补语时，仅标注后项，但是形成的命题是否定模态命题；（2）兼语结构中，“没有”与“不”的否定定义不一样，使用前者否定兼语结构时与直接否定兼语结构的后项意义不同，使用后者否定兼语结构时则与直接否定兼语结构的后项意义相同。

因此，假定句法分析已进行完毕，在进行深层语义标注时，对单动核结构、模态成分、并列结构、述补结构和兼语句的否定定义标注可遵循如下规则全自动地进行：

- (1) 对单动核结构、模态成分,分别在动核和模态成分下标注表示否定义的子结点即可;
- (2) 并列结构(假定有两个并列项),对前项命题和后项命题分别标注否定义;
- (3) 述补结构,如果补语中心语也对应一个命题,则仅对该命题标注否定义;如果述补结构仅对应一个命题,则对该命题标注否定义;
- (4) 述补结构,如果否定词修饰补语,则需要对补语对应的命题添加表示可能的模态义,之后在模态义的下面再添加否定义;
- (5) 兼语句,则仅对句子谓语对应的命题标注否定义。

对主谓短语做宾语的句子否定义标注需要人工判断,但一定程度上可以通过对谓语句词进行分类的方式予以自动标注:

对主谓短语做宾语的句子,根据谓语句词予以区分,如果属于表 3 中的第一种情况,则仅对充当宾语的主谓短语对应的命题标注否定义;如果属于表 3 中的第二种情况,则仅对句子谓语对应的命题标注否定义。

对于连谓句,则完全需要人工判断。其更为细致的表达规律,还需要进一步予以研究。

否定词充当谓语中心语时,一般不需要单独在语料库中为之标注否定义,只需要在知识库中添加“没有”等于“否定+有”之类的信息就可以了。

此外,复合否定词“不要、不曾”等等不仅表示否定义,还包含模态和时间等信息。因此,在深层语义标注时,需要在词典中对这些复合否定词所表达的语义进行分解,以便在深层语义标注时将其所表达的否定义、模态义、时间等信息完整地揭示出来。

6 否定词的语料库分布

句法树库²。本文所使用的句法树库为邱立坤等(2015)所构建的多领域依存树库^[18],该树库分为新闻、医药、口语、专利、微博五个领域,分别含有 64000、32227、40620、17035、29840 个句子,平均句长分别为 24、19.7、8.9、32.2、13.7 个词。该树库中,医药领域语料来自皮肤病领域教材和论文摘要,口语领域语料来自对外汉语口语教材,专利领域语料来自中文专利文献,微博领域语料为随机抽选的微博,新闻领域语料来自 1998 年和 2000 年《人民日报》。

否定表达形式的领域分布。通过统计可知,上述五个领域中的否定词数量分别为 10563、4407、4062、2376 和 11616 个。就否定词数量占句子数量的比例来看,微博领域否定词占比最高,达到 0.389,其它四个领域占比则在 0.1 到 0.16 之间,这说明微博领域中否定词的使用比例要远远高于其它领域,尤其是“不、没、没有、别”的大量使用。

五个领域的 TOP15 高频否定词及其频次如表 4 所示。观察该表可以发现:(1)有些词语在各个领域均出现,而且占比比较稳定,比如“不、没有、不、无、几乎”等。(2)还有一些词语表现出比较强的书面语色彩,比如“勿、未经”主要出现在医药类文本中,“未、并非”则主要出现在医药类和专利类文本中。(3)另外一些词语表现出比较强的口语色彩,比如“别、不用、从未、不再、永不”等主要出现在微博中;在医药和专利类文本中,“没”很少使用,主要使用“没有”;在口语、微博、新闻类文本中,“没”的使用比例则较高,从微博、口语到新闻,“没”的排名和占比逐渐降低。上述分析说明,微博类文本较为自由,口语性最强;口语类文本由于均来自教材,所以内容较为规范,口语性次之;《人民日报》新闻则是口语和书面语的混合体;医药和专利类文本则属于典型的书面文本。

否定词的句法位置分布。除了相对频次存在领域差异外,各否定词在不同领域中也存在一定的差异。受篇幅所限,下面仅详细分析否定词“不”的句法位置分布,如表 5 所示。该表显示了否定词“不”在微博、新闻、医药(这三个领域分别作为口语、书面语和口语混合体、书面语的代表)中常用的句法结构类型。“不”不能带宾语,其主要功能是第 5.1 节

² 该树库为句法树库,尚未进行深层语义表示标注,因而也未进行否定义标注。深层语义表示的标注工作尚在进行之中。

所示的七种功能。在进行统计时,我们进一步根据动核的类型将单动核结构区分为不同类型,比如“d-v-ADV”表示否定副词“不”修饰动词(v)并且充当状语成分(ADV);述补结构进一步区分为可能补语、结果补语、趋向补语三种类型。

表 4 五个领域语料库的 TOP15 高频否定词

新闻		医药		口语		专利		微博	
频次	词语	频次	词语	频次	词语	频次	词语	频次	词语
6523	不	2220	不	2402	不	1262	不	8018	不
1257	没有	1077	无	900	没有	228	未	1274	没
497	无	200	未	296	没	208	没有	1033	没有
342	未	169	避免	64	未	200	非	351	别
300	没	160	没有	64	别	151	无	233	无
293	非	157	非	53	无	137	防止	97	非
171	难以	130	防止	41	不用	65	避免	87	否认
143	几乎	82	几乎	32	几乎	30	难以	76	未
106	防止	52	难以	23	非	26	几乎	68	不用
99	不要	32	并非	23	拒绝	16	不再	37	从未
91	避免	31	不再	20	不要	16	并非	36	不再
90	否认	28	未能	19	难以	13	不必	32	拒绝
77	拒绝	27	勿	19	不必	8	无须	31	几乎
68	别	21	未经	15	避免	5	拒绝	27	不必
68	并非	12	不必	14	从未	3	免	26	永不

表 5 否定词“不”句法位置的领域分布差异

微博		新闻		医药	
频次	句法位置类型	频次	句法位置类型	频次	句法位置类型
4513	d-v-ADV	3095	d-v-ADV	1009	d-v-ADV
1177	模态成分	1109	模态成分	698	d-a-ADV
737	d-a-ADV	1070	d-a-ADV	329	模态成分
707	可能补语	748	可能补语	69	可能补语
345	主谓短语做宾语	172	d-d-ADV	28	主谓短语做宾语
106	兼语句	146	主谓短语做宾语	27	d-d-ADV
103	独用	34	兼语句	15	d-p-ADV
67	d-d-ADV	25	独用	9	独用
44	连谓句	23	连谓句	4	并列结构
39	d-p-ADV	22	d-p-ADV	4	结果补语
37	结果补语	22	并列结构	4	连谓句
12	d-r-ADV	18	趋向补语	1	d-b-ADV
11	并列结构	16	结果补语	1	并列结构-并列结构

从表 5 可以看出,三个领域中的句法位置基本一致,仅有的差异是从微博、新闻到医药领域兼语句比例逐渐降低,这同样体现了书面语和口语角度的领域差异。此外,各种句法位置中,简单句法位置(否定词充当状语、修饰模态成分、修饰可能补语等)较多,因此绝大多数否定定义的标注可以使用第 5.3 节的规则全自动地予以标注。需要人工进行判断的连谓句数量较少。

7 结论

本文在深层语义表示的框架下讨论了现代汉语否定定义的表达规律,具体包括全面收集否定表达形式,对否定表达形式尤其是多动核结构否定定义的表达规律进行分析,以及提出若干条深层语义标注框架下否定定义的标注规则。之后,基于大规模句法树库考察否定表达形式的分布,统计结果显示否定表达形式的分布存在明显的领域差异,绝大多数结构中否定定义的标注可以通过少数几条规则全自动地进行。本文工作可用于指导深层语义标注资源构建过程中否定定义的人工标注,同时也可否定定义自动标注提供可以借鉴的知识与规则。

参考文献

- [1] Banarescu L, Bonial C, Cai S, et al. Abstract Meaning Representation for Sembanking [C]//In Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, 2013: 178-186.
- [2] Palmer M, Gildea D, Xue N. Semantic role labeling [J]. Synthesis Lectures on Human Language Technologies, 2010, 3(1): 1-103.
- [3] 吕叔湘.疑问 否定 肯定[J].中国语文, 1985(4): 241-250.
- [4] 徐杰,李英哲.焦点和两个非线性语法范畴:“否定”“疑问”[J].中国语文, 1993(2): 81-102.
- [5] 袁毓林.论否定句的焦点、预设和辖域歧义[J].中国语文, 2000(2): 99-108.
- [6] 胡建华.否定、焦点与辖域[J].中国语文, 2007(2): 99-112.
- [7] 邢福义.论“不”字独说[J].华中师范学院学报, 1982(3): 124-133.
- [8] 沈开木.“不”字的否定范围和否定中心的探讨[J].中国语文, 1984(4): 404-412.
- [9] 李瑛.“不”的否定意义[J].语言教学与研究, 1992(2): 61-70.
- [10] 史锡尧.“不”否定的对象和“不”的位置[J].汉语学习, 1995(1): 7-10.
- [11] 石毓智.对“差点儿”类羡余否定句式的分化[J].汉语学习, 1993(4): 12-16.
- [12] 周一民.北京话里的“差点儿没 VP”句式[J].语言教学与研究, 2003(6): 24-30.
- [13] 袁毓林.动词内隐性否定的语义层次和溢出条件[J].中国语文, 2012(2): 99-104.
- [14] 张君,申长军.逻辑学教程[M].中南大学出版社,长沙,2009: 40、43—44、67、112.
- [15] 俞士汶,朱学锋等.现代汉语语法信息词典详解[M].北京:清华大学出版社,2003 第二版.
- [16] 中国社会科学院语言研究所词典编纂室,现代汉语词典(第六版)[M].商务印书馆,北京,2012: 396、138、1414、595.
- [17] 邵敬敏.现代汉语通论[M].上海教育出版社,上海,2001.6: 217-218.
- [18] 邱立坤,史林林,王厚峰.多领域中文依存树库构建与影响统计句法分析因素之分析[J].中文信息学报, 2015(5): 69-75.

作者简介：邱立坤（1979——），男，博士，副教授，主要研究方向为计算语言学、语料库语言学，Email: qiulikun@gmail.com; 黄焜(1993——)，女，本科生，主要研究方向为现代汉语语法，Email: 505744026@qq.com; 何保荣（1990——），女，硕士研究生，主要研究方向为中文信息处理，Email: 1532484154@qq.com; 亢世勇（1964——），男，硕士，教授，主要研究方向为中文信息处理，Email: kangsy64@163.com。



邱立坤

黄焜

何保荣

亢世勇

作者联系方式: 邱立坤, 地址: 山东省烟台市芝罘区红旗中路 186 号鲁东大学文学院, 264025, 13583566312, qiulikun@pku.edu.cn