

文章编号: 1003-0077(2024)02-0036-10

双重否定结构自动识别研究

王 昱¹, 袁毓林^{2,3}

- (1. 香港理工大学 中文及双语系, 香港 999077;
2. 澳门大学 人文学院 中国语言文学系, 澳门 519000;
3. 北京大学 中文系/中国语言学研究, 北京 100871)

摘 要: 双重否定结构是一种“通过两次否定表示肯定意义”的特殊结构, 直接影响自然语言处理中的语义判断与情感分类。该文以“ $\neg \neg P \Rightarrow P$ ”为标准, 对现代汉语中所有的“否定词+否定词”结构进行了遍历研究, 将双重否定结构按照格式分为了 3 大类, 25 小类, 常用双重否定结构或构式 132 个。结合动词的叙实性、否定焦点、语义否定与语用否定等理论, 该文归纳了双重否定结构的三大成立条件, 并据此设计实现了基于规则的双重否定结构自动识别程序。程序实验的精确率为 98.80%, 召回率为 98.90%, F_1 值为 98.85%。同时, 程序还从 96 281 句语料中获得了 8 640 句精确率约为 99.20% 的含有双重否定结构的句子, 为基于统计的深度学习模型提供了语料支持。

关键词: 双重否定; 自动识别程序; 语义分析

中图分类号: TP391

文献标识码: A

The Research on Automatic Recognition of the Double Negation Structure

WANG Yu¹, YUAN Yulin^{2,3}

- (1. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hongkong 999077, China;
2. Department of Chinese Language and Literature, Faculty of Arts and Humanities, University of Macau, Macau 519000, China;
3. Department of Chinese Language and Literature/ Center for Chinese Linguistics, Peking University, Beijing 100871, China)

Abstract: The double negation structure is a special structure of “expressing positive meaning through two negations”, in which the two negations have an important impact on the semantic analysis and emotional classification in natural language processing. Taking “ $\neg \neg P \Rightarrow P$ ” as the prototype, this paper examines the “negation word + negation word” structures in modern Chinese, and divides them into 3 categories, 25 sub-categories and 132 constructions in total. Then this paper proposes three conditions for the establishment of the double negation structure, and a rule-based method to identify the double negation. The accuracy rate of recognition of the double negation structure is 98.80%, the recall rate is 98.90%, and the F_1 value is 98.85%. The proposed method could identify 8 640 sentences with 99.20% true double negation structure from 96 281 sentences.

Keywords: double negation; automatic recognition program; semantic analysis

0 引言

在否定用法中, 有一种特殊的用法——双重否定。丁声树先生在《现代汉语语法讲话》中将其概括为: “一

句话先后用两个否定词, 如‘不能不去’‘没有人不去’‘非去不可’之类, 都是双重否定的句法。双重否定意思是肯定的, 不过跟单纯肯定不全一样。”例如, “我不得不喜欢他”指“我得喜欢他”, “我不一定不同意这个观点”指“我可能同意这个观点”^[1]。虽然句子使用

收稿日期: 2023-02-14 定稿日期: 2023-07-01

基金项目: 国家科技创新 2030“新一代人工智能”重大项目(2020AAA0106701); 国家社会科学基金(18ZDA295)

的是否定形式,但其表达的却是肯定语义,且句子的语气还与单纯肯定有一定的区别(如“不得不喜欢”的语气比“喜欢”强,而“不一定不同意”的语气比“同意”弱,等等)。可见,双重否定结构的存在会对句子的语义真值与语气强度产生重要影响。

对于自然语言处理来说,双重否定是处理语料时必须考虑的内容。例如,图1展示了ChatGPT对双重否定的判断情况。图中,ChatGPT无法正确判断“不+V1+不……”结构,误将“我不知道他不来”理解为了“我知道他来了”;将“我不认为他不来”误解为了“我认为他不会来”。由此可见,目前自然语言处理应用尚无法有效识别双重否定结构,且双重否定结构的存在会对程序的语义识别与情感分析产生重要影响。综上所述,无论是从汉语本体研究的理论意义考虑,还是从自然语言处理的实用价值出发,我们都有必要对双重否定结构进行更全面、更深入的研究与探索。

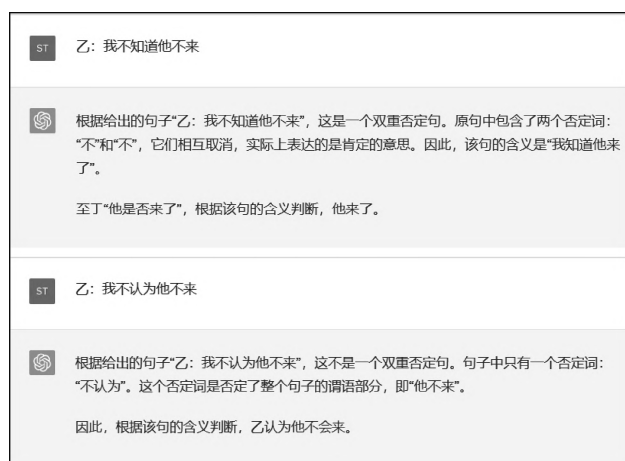


图1 双重否定结构处理情况示例图

0.1 文献综述

目前国内关于双重否定结构的理论研究主要集中在双重否定结构的定义、范围、格式和语义等领域。关于双重否定结构的定义与范围,学界存在不少争论,学者们争议较多的问题有:含有否定词的反诘句是否为双重否定,紧缩复句是否为双重否定,双重否定是否包含复句,等等。关于双重否定结构的格式,已有研究都是从分类与举例的角度进行讨论,尚未有研究从形式的角度对双重否定格式进行具体、详细的遍历分析。目前对双重否定格式概括最为全面的是茆崧^[2],他将双重否定划分的八大类型,25个格式。然而,仍有许多常在自然语料中出现的双重否定结构未被茆崧收入其中,如“不可能

不”“不应该不”“不是……不……”,等等。关于双重否定结构的语义,叶文曦^[3]、方绪军^[4]、何爱晶^[5]等引入了Ladusaw^[6]的形式语义学、Horn^[7]的元语否定等理论,对一些典型的结构进行了分析,得出了具有解释力的成果。然而,由于双重否定的范围、格式还未确定,目前学者只集中分析了几个典型的结构,覆盖面十分有限,缺乏系统性的梳理与总结。

目前关于双重否定结构的应用研究主要集中在情感分析领域,具体根据研究方法可以分为以下两种。

(1) 通过搜集典型的双重否定结构(图2),构建双重否定词典,以服务相关的情感分析,如王勇等^[8],吴杰胜、陆奎^[9],等等。这种方法的准确率很高,但是覆盖面不足。

双重否定词典	绝非不、并非不、不是不、不能不、不会不、不可不、不要不、不得不、没有不、无不、不无
--------	---

图2 王勇等^[8]双重否定词典

(2) 在否定词的基础上,通过统计修饰每一个情绪词的否定词个数来判断双重否定,并以系数的形式将双重否定的语气功能加入到情感分析的结果当中,如封洋^[10],等等。这种方法涵盖范围广,但是错误率很高,因为任何含有两次否定的结构都会被判断为表示肯定的双重否定结构。

0.2 本文选题及目标

综上所述,目前学界对汉语双重否定结构的研究成果颇丰,但仍存在一些不足之处。例如,双重否定结构的格式与范围尚不完整;双重否定语料资源匮乏;系统化、全面化的双重否定结构自动识别尚未实现,等等。鉴于此,本研究将“双重否定结构”作为研究对象,试图通过遍历分析与语料考察相结合的方法,对双重否定结构进行以下探索:

- (1) 梳理双重否定结构格式,使其能够全面覆盖CCL语料库;
- (2) 总结双重否定结构成立条件,并据此提出相应的计算机识别策略;
- (3) 建立高 F_1 值的双重否定结构自动识别程序;
- (4) 进一步验证语言学知识在双重否定结构自动识别过程中的贡献,通过程序测试上述成立条件在双重否定结构识别过程中的作用;
- (5) 搜集双重否定语料资源,为基于统计的双重否定识别深度学习模型提供支持。

1 双重否定结构的定义标准与考察范围

鉴于语义真值识别和情感极值判断是计算机对否定结构进行语义识别时所面临的主要问题,本文借鉴形式语义学,为双重否定拟定了一个工作定义:只要两次否定与肯定在语义真值上相同,即“ $\neg \neg P = \Rightarrow P$ ”,便属于双重否定。

目前我们的考察范围为所有“否定词(+……)+否定词”中双重否定表肯定的结构。暂不考虑下列情况:

- (1) 否定词为隐性否定词(即本身语义内含有否定意思的动词,如“讨厌”“拒绝”等);
- (2) 否定类型为语用否定的结构(如“我不是不喜欢你,而是恨你”中的“不是不喜欢”);

(3) “反问句+否定词”结构(如“难道……不……”等,具体参见刘彬、袁毓林^[11])。

我们结合吕叔湘^[12]、王力^[13]、朱德熙^[14]等前贤研究,梳理出了十三个否定词,前十个为否定副词,后三个为否定动词,具体如下:

“非、不、别、甬、未、莫、勿、没、没有、休”
“无、没、没有”

结合语料,我们对“否定词+……+否定词”组合中符合要求的双重否定结构的格式进行了遍历梳理,整理出了 25 种“否定词+否定词”可表肯定的结构,具体如表 1 所示^①。

下面,我们将对上述 25 种“否定词+……+否定词”格式展开具体分析,梳理每一种格式中双重否定结构的成立条件,并设计与之相应的计算机识别策略,在此基础上总结所有现代汉语中常用的双重否定结构。

表 1 双重否定结构格式调查情况表

	不	没	没有	无	非	莫	别
不	不……不	不……没……	不……没有……	不无	▲	▲	▲
没	没不 没……不……	没……没……	没……没有……	▲	▲	▲	▲
没有	没有不 没有……不……	没有……没……	没有……没有……	▲	▲	▲	▲
无	无不 无……不……	▲	▲	▲	▲	▲	▲
非	非不 非……不……	非没 非……没……	非没有 非……没有……	非无	▲	非……莫……	▲
莫	莫不	▲	▲	▲	▲	▲	▲
别	别不	别没	▲	▲	▲	▲	▲

2 双重否定结构的成立条件与识别策略

通过初步考察,发现,双重否定结构“不 V1……不 V2”的成立条件最为复杂,需要同时满足以下三个条件:

- (1) “不 V1”与“不 V2”构成述宾关系;
- (2) “V1”动词有限制(为非叙实动词);
- (3) “不 V1”的否定焦点在“不 V2”上。

除了“不是……不……”需要区别语义否定与语用否定外,其他双重否定结构的成立条件都可被囊括在上述三个条件中,只是部分细节存在差异。因此,我们先对“不 V1……不 V2”与“不是……不……”的成立条件与识别策略进行详细分析,再在此基础上,对其他双重否定结构进行讨论。

2.1 “不 V1……不 V2”双重否定结构的成立条件与识别策略

2.1.1 第一个条件:“不 V1”与“不 V2”构成述宾关系

“不 V1”与“不 V2”的结构类型有并列、主谓、紧缩、述宾等。在各类结构类型中,只有述宾结构的“不+V1+(……)+不+V2”存在表示双重否定的可能。具体讨论如下:

并列结构的“不 V1……不 V2”,指“不哭不闹”、“不高不低”这一类表达。袁毓林^[15]指出,并列结构“通常不能通过直接在这种谓词性并列结构的前面加上‘不、没有’等否定词来构成否定式,而是要在这

① 其中,三角形表示该组合不出现/极少出现于实际语料中。

种并列结构的各个直接成分之前分别加上‘不、没有’等否定词。”因此,并列结构的“不 V1……不 V2”只是其肯定式“V1V2”的单重否定,不属于双重否定结构。如“不哭不闹”不等于“哭闹”。

主谓结构的“不 V1……不 V2”语料数量很少,指“不隐藏不代表泄露”“不买票不是我的决定”这一类表达。在该类结构中,“不 V1”只是一个命题陈述,是交流中的旧信息。“不 V1”的“不”与“不 V2”的“不”并没有语义上的关联,不构成“双重”否定的结构。如“不隐藏不代表泄露”不等于“隐藏代表泄露”,“不买票不是我的决定”不等于“买票是我的决定”。

紧缩结构的“不 V1……不 V2”指“不给钱不办事”“不买票就不让进”这一类表达。紧缩结构虽然在语义上有条件性,但在句法上仍是并列关系,前后不构成从属结构。关于紧缩条件类的结构是否为双重否定未有定论。本文主要从形式语义学的角度对其进行讨论。

以“不给钱不办事”为例。“给钱办事”语义为“如果给钱,那么办事”。P 命题可以分解为 q_1 “给钱”, q_2 “办事”,逻辑式为蕴含式 $q_1 \rightarrow q_2$,它的等值式为 $\neg q_2 \rightarrow \neg q_1$ 。而“不给钱不办事”语义为“如果不给钱,那么不办事”,逻辑式应为蕴含式 $\neg q_1 \rightarrow \neg q_2$ 。从下列真值表(表 2)可以看出, $q_1 \rightarrow q_2$ 与 $\neg q_1 \rightarrow \neg q_2$ 的语义真值不一致,不符合“ $\neg \neg P = \Rightarrow P$ ”的标准。因此从形式语义学来看,紧缩语义结构不是双重否定结构。

表 2 紧缩结构语义真值表

q_1	q_2	$\neg q_1$	$\neg q_2$	$q_1 \rightarrow q_2$	$\neg q_2 \rightarrow \neg q_1$	$\neg q_1 \rightarrow \neg q_2$
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	F
F	F	T	T	T	T	T

最后述宾结构的“不 V1……不 V2”指“不觉得不好”“不认为你不来”等 V1 为述语,“不 V2”为宾语的结构。“不 V1”与“不 V2”构成从属结构,存在“否定+否定 $=\Rightarrow$ 肯定”的可能。例如,“不认为他明天不会来”可以理解为“认为他明天会来”,“不觉得这件事不妥”可以理解为“觉得这件事妥”,等等。由此,我们可以得出“不 V1……不 V2”构成双重否定结构的第一个条件为:“不 V1”与“不 V2”构成述宾关系。

2.1.2 第二个条件: V1 为非叙实动词

述宾结构的“不 V1……不 V2”中只有一部分成员,其“不 V1”对“不 V2”有管辖作用,属于双重否定结构,其他成员仍只表示单纯的否定。试看下例:

1a. 我不知道他不来。

1b. * 我知道他来。

2a. 我不幻想他不来。

2b. * 我幻想他来。

3a. 我不认为他不来。

3b. 我认为他来。

通过例句可以发现,当 V1 为“认为”时“不 V1……不 V2”可以理解为“V1……V2”,而当 V1 为“知道”“幻想”时却不能。同样是动词,“知道”“幻想”“认为”对“不 V2”的影响却存在着差别。经过研究,我们发现,“不 V1”对“不 VP”是否有管辖作用与 V1 的语义有关,具体来说与 V1 的叙实性有关。

李新良^[16]将叙实性定义为“叙实性是动词的一种语义功能,即动词预设其宾语小句真值的能力。具体来说,肯定式和否定式都预设其宾语小句为真的动词是叙实动词……肯定式和否定式都不预设其宾语小句为真,也不预设其宾语小句为假的动词是非叙实动词……肯定式和否定式都预设其宾语小句为假的动词是反叙实动词”。对于叙实动词和反叙实动词来说,由于其预设固定,无论主句有无否定,宾语小句的真值都不变,因此主句无法影响宾语小句的真值,不构成“ $\neg \neg P = \Rightarrow P$ ”。而对于非叙实动词(如:认为)来说,由于非叙实动词对宾语小句并没有预设,在述宾结构中,主句中的 V1 可以对宾语的真值造成影响,存在“ $\neg \neg P = \Rightarrow P$ ”的可能。因此,我们可以得出“不 V1……不 V2”构成双重否定结构的第二个条件为: V1 为非叙实动词。

2.1.3 第三个条件:“不 V1”的否定焦点包含“V2”

除上述两个条件外,结构中否定焦点的情况也会对“不 V1……不 V2”是否为双重否定造成影响。试看下例:

4a. 我不认为他不来。

4b. 我认为他来。

5a. 我不认为他故意不来。

5b. * 我认为他故意来。

6a. 我不相信他不喜欢我。

6b. 我相信他喜欢我。

7a. 我不相信他不喜欢我到了看见我就恶心的地步。

7b. *我相信他喜欢我到了看见我就恶心的地步。

在例句中,4a、6a 可以转换为 4b、6b,而 5a、7a 却不能转换为 5b、7b。这主要与否定的焦点有关。袁毓林^[17]指出“有的成分表达的是句子的预设意义,属于旧信息,事实上它们的意义在否定的情况下仍然得以保持;有的成分表达的是句子的焦点意义,属于新信息,它们是真正被否定的。”当“不 V1”的否定焦点未落在“不 V2”上时,“不 V1”对“V2”不造成否定,不能构成“ $\neg \neg P$ ”结构,因此无法满足“ $\neg \neg P = \neg \neg P$ ”的条件,不属于双重否定。

如“我不认为他故意不来”中的“不 V1”否定的是“故意”而不是“不来”,其中“不来”是预设成分,属于旧信息。无论是“认为他故意不来”还是“不认为他故意不来”,语义都是“他不来”。“V1 认为”的否定无法影响到“不来”的真值,无法构成“否定+否定”的语义结构。因此,我们可以得出“不 V1……不 V2”构成双重否定结构的第三个条件为:“不 V1”的否定焦点落在“不 V2”上。

2.2 “不是……不 V2”双重否定结构的成立条件与识别策略

如本节开头所述,“不是……不 V2”若要表示双重否定,除需满足上述“不 V1……不 V2”的成立条件外,还需保证“不是”为描述性真值否定(descriptive truth-functional negation,又称语义否定),而不是元语否定(non-truth-functional negation,又称语用否定)。“所谓元语否定,就是用元语言对对象语言所描述的非真值语义的否定,……是一种非真值意义否定;与之相对应的是真值否定,否定的是句子的真值条件(truth condition)”^[5]“元语/语用”否定常常是引述性否定,是对之前对话中已出现的内容的否定。例句如下:

8a. “可现在杀他不容易啊。”有人说。不是不容易,是根本不可能。

8b. * “可现在杀他不容易啊。”有人说。是容易,是根本不可能。

例句中否定形式的 8a“不是不容易,是根本不可能”不能理解为相应的肯定形式 8b“是容易,是根本不可能”。这是因为例 8 中的“不是”是语用否定,并不对语义真值产生影响,不构成“否定+否定”的语义结构,无法通过两次否定表示肯定。因此,为了识别“不是……不 V2”中的双重否定结构,首先需要区分“不是”是语义否定还是语用否定。为此,本文

对大量语料进行了考察后发现,“不是”的否定情况具体有以下三种:

(1) 当“不是”的上文中没有出现“不是”否定的观点时,“不是”只能是语义否定,而不能是语用否定。例如:

9a. 甲:你明天来不来参加生日聚会?

9b. 乙:不是我不乐意,而是我这边实在没时间。

语义否定✓ 我是乐意,(但)是我这边实在没时间。

语用否定× “不乐意”这个表述不恰当,是我这边实在没时间。

(2) 当“不是”的上文中出现了“不是”否定的观点,且“不是”所在句的后文与被否定观点的真值一致,则“不是”只能是语用否定,而不能是语义否定。例如:

10a. 甲:你不乐意明天来参加生日聚会?

10b. 乙:我不是不乐意,而是超级反感。

语义否定× 我是乐意,是超级反感。

语用否定✓ “不乐意”这个表述不恰当,(应该)是超级反感。

(3) 当“不是”的上文中出现了“不是”否定的观点,且“不是”所在句的后文与被否定观点的真值不一致,则“不是”既可看作语义否定,又可看作语用否定。对于这种情况,从计算机处理的角度出发,我们可以统一将其处理为“不是”表示语义否定。

11a. 甲:你不乐意明天来参加生日聚会?

11b. 乙:不是我不乐意,而是我这边实在没时间。

语义否定✓ 我是乐意,(但)是我这边实在没时间。

语用否定✓ “不乐意”这个表述不对,(应该)是我这边实在没时间。

综上所述,只有第二种情况“不是”才会是语用否定。而其余情况,皆可被计算机视为语义否定。

根据上述条件,我们可以通过计算情感真值的方法来判断“不是”的否定类别。在文本里,大多数表示语用否定的“不是”都只出现在“不+是+不+ x , (而)是+ y ”的结构中。因此,我们可以拟定识别策略如下:

提取“不+是+不+ x ”中的“ x ”,并匹配“ y ”与“不+ x ”的情感真值。若“不+是+不+ x , (而)是+ y ”中“ y ”的情感真值与“不+ x ”不一致,则句中的“不是不”属于双重否定结构;若一致,则不属于双重

否定结构。

2.3 双重否定结构的成立条件

以上即为“不 V1……不 V2”与“不是……不 V2”双重否定成立条件的基本情况。本文所有双重否定结构的成立条件皆在上述成立条件的基础上进行调整。我们按照“组合类”“构式类”“黏合类”三个

大类类别^①，对第 2 节中所提出的 25 小类双重否定结构的成立条件进行了详细梳理，在此过程中，本文还结合成立条件，从 25 小类中梳理出了双重否定常用结构或构式 132 个。与芜崧^[2]的结果相比，我们的分类更系统，类型更丰富，覆盖更全面，所涉及的常用双重否定结构格式约为芜崧的 5 倍。具体结果如图 3 所示。

大类	双重否定结构格式小类	双重否定结构常用格式	具体限制情况	语料具体示例
组合类	1. 不+助动词+不+VP	不得不+VP 不能不+VP 不可不+VP …共计 23 个	助动词范围：能、能够、可能、会、可以、应该、应、应当、一定、要、得、愿意、愿、肯、可、想、要、敢、该、当、准、许、容	不得不四处筹钱 不得不惭愧地沉默不语 不能不从我国的国情出发 不可不仔细用功
	2. 不+是（+…）+不+VP	不是（+…）+不+VP 共计 1 个	1. “不是”与“不”必须是述宾关系，“不是”后的“不 VP”为否定焦点； 2. 如果出现在“不+是+x，（而）是+y”结构中，若 y 的情感真值与“不 x”的情感真值不一致，则“不是不”为双重否定结构，否则，不为双重否定结构。	不是不读书 不是不重要 不是不愿意丰富自己的生活 不是审计师不小心
	3. 不+V1（+…）+不+VP	不认为（+…）+不+VP 不觉得（+…）+不+VP 不希望（+…）+不+VP …共计 11 个	1. 不+V1+不+VP 为述宾结构 2. V1 属于非叙实动词，且 V1 后的“不 VP”为否定焦点。V1 具体为：认为、说、感到、觉得、允许、同意、相信、愿意、希望、考虑、打算。	不觉得不尊重 不认为不适当 不觉得不应该抢 不认为自己的点子不好 不希望有个不愉快的开始
	4. 不+助动词+V1（+…）+不+VP	不会感到（+…）+不+VP 不会觉得（+…）+不+VP 不应该说（+…）+不+VP …共计 15 个	1. 助动词范围 2. 不+V1+不+VP 为述宾结构 3. V1 属于非叙实动词，且 V1 后的“不 VP”为否定焦点。	不能说不好 不能说不可以 不会觉得不愉快不
	5. 不+助动词+没/没有+NP/VP	不可没/没有+NP/VP 不能没/没有+NP/VP 不可能没/没有+NP/VP …共计 23 个	同 1	不得没规矩 不可没有梦想 不能没带耳机和钱
	6. 不+是（+…）+没/没有+NP/VP	不是（+…）+没/没有+NP/VP 不是（+…）+没有+NP/VP 共计 2 个	同 2	不是他没有这个能耐 不是男人没有责任 不是人民还没有战胜这些灾害
	7. 不+V1（+…）+没/没有+NP/VP	不认为（+…）+没/没有+NP/VP 不觉得（+…）+没/没有+NP/VP 不希望（+…）+没/没有+NP/VP …共计 11 个	同 3	不认为院长没有注意到手术中发生的过失 我不认为自己就没有生存力 我不相信没有主持公道的地方 你不办，我还不相信没有人办了。
	8. 不+助动词+V1（+…）+没/没有+NP/VP	不能认为…+没/没有+NP/VP 不会觉得…+没/没有+NP/VP 不能说…+没/没有+NP/VP …共计 15 个	同 4	不会觉得这世界没有他的容身之地 不能认为本案已经没有价值。 不能说他没有责任。
	9. 没/没有+NP/VP+不+VP/AP	没有人不+VP/AP 没有什么不+VP/AP 没有一个不+VP/AP 共计 3 个	#	没人不喜欢 没人不辛苦 没有人不会走 没有一个学校不采用应试教育

图 3 双重否定结构格式与成立条件示例图

① 组合类：否定词与否定词之间不相连，扩充了其他成分；构式类：特定的否定词与否定词之间组成构式，结构复杂，形式固定；黏合类：否定词与否定词之间没有其他成分，二者紧连。

大类	双重否定结构格式小类	双重否定结构常用格式	具体限制情况	语料具体示例
组合类	10. 没/没有+NP+没/没有+VP	没/没有+NP+没/没有+VP 共计 4 个	/	没有人没完成任务 没有白羊座没有见过他 没有一个人没有写过好作品
	11. 无+NP+不+VP/AP	无一不、无人不、无一个不 共计 3 个	/	无人不升官 无人不受奖 无一人不会唱
	12. 非+……+没/没有/不+VP/AP/NP	并非/绝非+……+没有/没/不+VP/AP/NP 共计 6 个	/	并非他没有尽职 并非对生活没有要求 并非分部不精当
句式类	13. 不……不行/不好	不……不行、不……不好 共计 2 个	“不……不行/不好”为句式短语，在一个分句中出现。	不降价不行、不孝顺不行、不卖不行、不借不好
	14. 非……不可/不行	非……不可、非……不行 共计 2 个	“非……不可”为句式短语，在一个分句中出现。	非维新不可、非彻底改革不可、非看赛马不行、非阿永不行
	15. 非……莫属	非……莫属，共计 1 个	“非……莫属”为句式短语，在一个分句中出现。	非马云莫属、非他莫属、非邓丽君莫属、非机器人莫属
粘合类	16. 没不……	没不…… 共计 1 个	“没”、“不”皆为独立的否定词	没不严肃、没不成的、没不让你追求
	17. 没有不……	没有不…… 共计 1 个	“没有”、“不”皆为独立的否定词	没有不对、没有不同意、没有不失败的
	18. 不无……	不无…… 共计 1 个	“不”、“无”皆为独立的否定词	不无道理、不无裨益、不无参考意义、不无关系
	19. 无不……	无不…… 共计 1 个	“无”、“不”皆为独立的否定词	无不佩服、无不为之动容、无不写、无不悲伤欲绝
	20. 非不……	非不…… 共计 1 个	“非”为“并非”、“绝非”等否定词，“不”为独立的否定词	并非不懂政治、并非不知道、并非不合理、绝非不让人接近
	21. 非没……	非没…… 共计 1 个	“非”为“并非”、“绝非”等否定词，“没”为独立的否定词	并非没意见、并非没听到、并非没感觉、并非没钱
	22. 非没有……	非没有…… 共计 1 个	“非”为“并非”、“绝非”等否定词，“没有”为独立的否定词	并非没有依据、并非没有想法、绝非没有道理、绝非没有可能
	23. 非无……	非无…… 共计 1 个	“非”为“并非”、“绝非”等否定词，“无”为独立的否定词	并非无次序、并非无计划、绝非无好处、绝非无意义
	24. 别不……	别不…… 共计 1 个	“别”、“不”皆为独立的否定词	别不信、别不吭声、别不满意、别不玩
	25. 别没……	别没…… 共计 1 个	“别”、“没”皆为独立的否定词	别没良心、别没钱
合计	25 小类	132 个常用结构/句式		

图 3 （续）

2.4 双重否定结构的识别策略

针对上述成立条件,我们设计了相应的计算机识别策略,具体如表 3 所示。

表 3 双重否定结构识别策略表

序号	成立条件	识别策略
1	否定词 1 所在结构与否定词 2 所在结构构成述宾关系	依存句法分析,检测句法结构
2	否定词 1 所在结构对动词有限制	根据情况建立词表(如,助动词词表)+字符串匹配
3	否定词 1 的否定焦点在否定词 2 所在结构上	依存句法分析,检测句法结构+字符串匹配,排除否定焦点在“否定词 2 结构”修饰语上的情况

续表

序号	成立条件	识别策略
4	“不是”为语义否定	构建情感词典,计算情感真值+字符串匹配

3 双重否定自动识别程序的建立

为了使计算机能够识别助动词、非叙实动词、情感真值,本文对助动词、非叙实动词与情感词进行了梳理,在常用的基础词表中补充了助动词词表、非叙实动词词表与情感词表。具体如下:

常用助动词词表(23 个):

能、能够、可能、会、可以、应该、应、应当、一定、要、得、愿意、愿、肯、可、想、要、敢、该、当、准、许、容

常用非叙实动词词表(11个):

认为、说、感到、觉得、允许、同意、相信、愿意、希望、考虑、打算(袁毓林^[17]、李新良^[16])

情感词方面,我们结合知网 HowNet 情感词典、台湾大学 NTUSD 简体中文情感词典与清华大学李军中文褒贬义词典,设计了情感词表,共收纳正

面词语 10 323 个,负面词语 9 411 个。

根据上述识别策略,我们设计编写了双重否定结构自动识别程序。通过该程序,对含有两个否定词的语料文件进行自动识别实验,输出其中存在的双重否定句以及相应的双重否定结构^①,程序输出结果示例如图 4 所示。

此句话为双重否定句,双重否定结构为“无不”,语气为--语气强,加强肯定语气--: *****19:.....尽的义务。世界各国的义务教育年限有长有短,培养目标的提法各有不同,但是无一不是为了培养

此句话为双重否定句,双重否定结构为“不+助动词+不”,语气为--语气强,加强肯定语气--: *****22:.....切社会的生存和发展的基础,究竟应当把年轻一代培养成怎样的人,不能不受现实社会生

此句话为双重否定句,双重否定结构为“不+助动词+不”,语气为--语气强,加强肯定语气--: *****30:.....由于基层财政困难,一些农村学校的校长为了保证学校的正常运转,不得不四处筹钱,被

此句话为双重否定句,双重否定结构为“不+助动词+不”,语气为--语气强,加强肯定语气--: *****38:.....中并非不能培养学生的素质,同时应试教育的强制性好,可以使学生的不得培养一些素质!

此句话为双重否定句,双重否定结构为“不+助动词+不”,语气为--语气强,加强肯定语气--: *****39:.....门的练习题、测验卷、考试宝典等,教师逼着买,家长主动买,学生不得不买,于是学生

此句话为双重否定句,双重否定结构为“不+助动词+不”,语气为--语气强,加强肯定语气--: *****40:.....虽然教育管理者口头上不得不说要大搞素质教育,但是他们的内心深处依然深深热恋着

此句话为双重否定句,双重否定结构为“不+助动词+非叙实动词+不”,语气为--语气强,加强肯定语气--: *****60:.....样轻轻抹去”在如此艰难的情况下,马克思对自己情绪的把握和管理不能说不

图 4 程序输出结果示例图

4 双重否定自动识别实验

我们从 CCL 语料库中,按照各类结构的情况,进行了同等分布提取(即根据各结构在 CCL 语料库中的语料数量比例进行提取),收集了 100 000 条初始语料。通过程序,对上述语料进行筛选,排除了所有未被分句、长度超过 150 字的句子,最终获得测试语料 96 281 句。

4.1 初始实验

我们使用双重否定结构自动识别程序对 96 281 句语料进行识别,获得了 8 640 句计算机判断含有双重否定结构的句子。由于人力有限,我们无法对 8 640 个句子全部进行人工校验。因此,为计算精确率,我们从上述句子中随机抽取了 1 000 句进行检测。经校验,1 000 个句子中判断正确的句子为 992 句,精确率约为 99.20%;在 CCL“否定词+否定词”语料中,双重否定句的比例约为 8.90%。

为进一步验证,我们从 96 281 句语料中随机抽取了 1 000 句语料进行检测。通过人工校验,上述 1 000 句测试语料中存在 92 句含有双重否定结构的句子。我们将上述 1 000 句测试语料输入双重否定自动识别程序,程序从中识别出了 90 句含有双重否定结构的句子。根据计算,初始实验中程序的召回率约为 97.83%。CCL“否定词+否定词”语料中

双重否定句的比例在 9%左右。

4.2 正式实验

初始实验表明,上述 96 281 句语料中双重否定结构与非双重否定结构的比例差距很大(约为 9:91),因此为了更好地计算程序的精确率与召回率,我们人工构建了 2 000 句语料,其中 1 000 句为含有双重否定结构的正例,1 000 句为不含双重否定结构的负例。

4.2.1 实验结果

我们将语料输入双重否定结构程序,具体结果如表 4 所示。

程序识别的精确率约为 98.80%,召回率约为 98.90%, F_1 值约为 98.85%,实验结果较王勇等^[8]、封洋^[10]的结果有较明显提升。需要说明的是,本实验为封闭测试,数据方式为抽样,且文章对于双重否定结构的判断均来自于作者本人,因此结果可能存在一定的偏差。后续我们将投入更多的时间与人力,来获取更为准确的数据。

4.2.2 实验结果分析

通过分析,可以发现程序识别与召回出错的主要原因与句子的分词与句法分析错误有关。由于分

^① 在程序中,我们还做了双重否定结构的语气识别,由于篇幅原因,本文暂不对其进行讨论。

表 4 实验结果数据表

	1 000 正例	1 000 负例	总精确率 $P/\%$	总召回率 $R/\%$	F_1 值/ $\%$
	识别出的双重否定句数				
本文程序	989	12	98.80	98.90	98.85
王勇等(2014)	777	10	98.72	77.70	86.96
封洋(2016)	1 000	1 000	50.00	100	66.67

词与句法分析等基础处理工具的原因,程序对一些句子的句法判断失误,使得一些本应被判为并列关系、因果关系的成分,被误判为述宾关系,从而导致整个结构的判断错误。示例如下:

34743: ……不大紧;有的急于求成把将来要办的事情,拿到今天来办,由于条件不允许迟迟开展不了。

(程序识别结果:双重否定结构为“不+非叙实动词+……不”)

(实际情况:“条件不允许”与“迟迟开展不来”是因果并列关系)

6893: ……要的网络公司均未能达到阿尔诺的预期。为此,去年 6 月底,阿尔诺不得不刹车。他说,他要考虑“战略调整”。

(程序分词:阿尔诺不得不/刹车)

(实际分词:阿尔诺/不得不/刹车)

当我们输入人工修改后的分词与句法分析结果后,程序的错误得到纠正,精确率与召回率皆可达到 100%。

5 双重否定自动识别程序的建立

为测试上述三个条件在双重否定结构识别过程中的作用,进一步验证语言学知识在双重否定结构自动识别过程中的贡献,我们将人工构建的 2 000 句语料作为输入,测试在取消某一条件后,双重否定结构识别程序的识别情况与召回情况。具体结果如表 5 所示。

表 5 双重否定结构成立条件测试结果

	1 000 正例	1 000 负例
	识别出的双重否定句数	
原始数据	989	12
取消“述宾结构”条件限制	1 078	460
取消“非叙实动词”条件限制	1 031	29
取消“否定角度”条件限制	989	13

可以看到,“构成述宾结构”与“动词为非叙实动词”对整个双重否定结构的识别造成较大影响。尤其在召回率方面,相较于我们提供的标准正确数据,没有“构成述宾结构”条件约束的程序将额外召回约 37 倍的错误结构(460 : 12),而没有“动词为非叙实动词”条件约束的程序也将额外召回 1.4 倍的错误结构(29 : 12)。

“否定焦点”条件对双重否定结构识别的影响最为轻微。我们认为,这可能是因为人们在实际语言交流中很少会使用非常复杂的句子,例如:“我不相信他不喜欢我到了看见我就恶心的地步”。当我们扩大检测数据,对 96 281 句原始语料进行程序测试时,缺少否定焦点约束的程序将会比标准程序额外召回 140 句双重否定结构,进一步说明否定焦点会对双重否定结构识别程序造成影响,只是由于语料中复杂的句子很少,导致影响较轻微。这种情况也符合我们日常表达的经济性原则。

6 结语

本文以“ $\neg \neg P \Rightarrow P$ ”为标准,借助计算机程序与 CCL 语料库,对现代汉语中所有的“否定词+否定词”结构进行了遍历研究,实现了以下目标:

(1) 将双重否定结构按照格式分为了 3 大类, 25 小类,常用双重否定结构或构式 132 个,进一步地揭示出了双重否定结构的全貌;

(2) 总结得出了双重否定结构成立的三大条件;并对其进行了实验测试分析,进一步验证了语言学知识在双重否定结构中的作用;

(3) 补充了助动词表、非叙实动词表、情感词表等基础词表,编写实现了双重否定结构自动识别程序。识别的精确率约为 98.80%,召回率约为 98.90%, F_1 值约为 98.85%;

(4) 获得了 8 640 句精确率约为 99.20%的标明双重否定结构的句子,为后续建立双重否定语料库

提供了支持。具体文件烦请参见脚注链接^①。

本文还有许多不足未尽之处。例如,本文的识别程序是在规则的基础上建立的,而人为编写的规则未免有不尽之处;本文对双重否定结构的判断皆基于作者个人的语感,未免有疏漏之处;对于一些已经固化的双重否定结构,如“非……不……”“无非”“莫非”等,本文的处理还较为粗糙,有待进一步分析与探索,等等。

未来我们拟在本文的基础上,继续展开与深入,具体计划有:

(1) 建立双重否定语料库,在程序的辅助下构建数万句级的双重否定语料库;

(2) 探究双重否定结构自动识别的深度学习模型;

(3) 对双重否定结构的语用方面进行进一步研究与探索;

(4) 对双重否定结构中的构式,如“非……不可”“非……莫属”等等,进行进一步的研究与探索。

参考文献

- [1] 丁声树,等.现代汉语语法讲话[M].北京:商务印书馆,2004: 200-202.
- [2] 茆崧.双重否定句的种类与功能[J].荆州师专学报(哲社版),1987,(3): 52-57.
- [3] 叶文曦.否定和双重否定的多维度研究[J].语言学研究,2013(2): 20-31.
- [4] 方绪军.“不是不X”、“不是没(有)X”和“没(有)不X”[J].语言科学,2017,16(05): 511-521.
- [5] 何爱晶.反叙的非真值义否定和真值义肯定[J].外语研究,2019,36(04): 28-35.
- [6] LADUSAW W A. Negation and polarity items[M]. The Handbook of Contemporary Semantic Theory. Oxford: Blackwell Publishing Ltd, 1997: 321-341.
- [7] HORN L. A nature history of negation[M]. Chicago: University of Chicago Press, 1989: 311-312.
- [8] 王勇,吕学强,姬连春,等.基于极性词典的中文微博客情感分类[J].计算机应用与软件,2014,31(01): 34-37.
- [9] 吴杰胜,陆奎.基于多部情感词典和规则集的中文微博情感分析研究[J].计算机应用与软件,2019,36(09): 93-99.
- [10] 封洋.中文微博情绪分析[D].上海:上海交通大学硕士学位论文,2016.
- [11] 刘彬,袁毓林.反问句否定意义的形成与识解机制[J].语文研究,2017(04): 12-19.
- [12] 吕叔湘.中国文法要略[M].北京:商务印书馆,1956.
- [13] 王力.中国现代语法[M].北京:商务印书馆,1985.
- [14] 朱德熙.语法讲义[M].北京:商务印书馆,1982.
- [15] 袁毓林.并列结构的否定表达[J].语言文字应用,1999(03): 42-46.
- [16] 李新良.立足于汉语事实的动词叙实性研究[J].世界汉语教学,2015,29(3): 350-361.
- [17] 袁毓林.论否定句的焦点、预设和辖域歧义[J].中国语文,2000(02): 99-108.



王昱(1997—),博士研究生,主要研究领域为汉语语言学与计算语言学。
E-mail: wangyustu@pku.edu.cn



袁毓林(1962—),通信作者,教授、博士生导师,主要研究领域为理论语言学和汉语语言学,特别是句法学、语义学、语用学、计算语言学和中文信息处理等。
E-mail: yuanyl@pku.edu.cn

^① https://github.com/LANGVAGE-UNDERSTANDING/Chinese_Double_Negation