

## 依存句法分析的回顾与发展

杨 牧<sup>1</sup>, 蔡言胜<sup>2</sup>

(南开大学 汉语言文化学院, 天津 300350)

**摘 要:** 依存语法的基本思想是探讨词与词之间的依存关系。依存句法分析以依存语法为理论来源, 以算法为实现手段, 在语言研究和实际应用中均具有一定的价值。采用 CiteSpace 软件, 对 Web of Science 核心数据库 1985—2020 年所收录的相关文献进行可视化分析, 研究显示, 依存句法分析研究的发文量呈递增趋势, 研究内容聚焦于语义理解和算法设计, 研究主体为计算机学界和语言学界。

**关键词:** 依存语法; 依存句法分析; 可视化分析

### 一、引言

语言学研究的趋势之一是越来越强调用数据说明问题。得益于声学技术的发展, 语音学可以更加精密地采集数据。结合统计学的分析方法, 语音学研究具备了成熟的定量分析能力。而传统的语法定量研究主要是统计某类语法单位或特定格式的出现频次, 解释能力是有限的, 因此, 语法研究必须采取新的手段。自然语言处理是应用语言学的重要方向之一。自然语言处理直接面向应用, 这就要求必须有合适的方法批量处理语言, 准确找到所需信息, 之后的所有操作都是建立在这个基础之上的。在具体实践中, 依存语法被证明是合适的理论, 学者们根据依存语法的基本思想建立起了比较成熟的句法分析方法。可以说, 依存

语法的理论和分析方法, 无论是对语言学的本体研究还是应用研究, 都提供了很大的帮助。

依存语法是一种基于词与词关系的形式语法。Robinson 曾给出四条公理<sup>[1]</sup>: 1. 一个句子只有一个成分是独立的; 2. 句子中的其他成分直接依存于某一成分; 3. 任何一个成分都不能依存于两个或两个以上的成分; 4. 如果成分 A 直接依存于成分 B, 而成分 C 在句子中位于 A 和 B 之间, 那么, 成分 C 或者依存于 A, 或者依存于 B, 或者依存于 A 和 B 之间的某一成分。这一理论初看未免有些抽象, 下面, 我们就以“这是一个好例子”为例来说明这四条公理。该例句的依存树图如图 1(左)所示、该例句的依存有向图如图 1(右)所示:

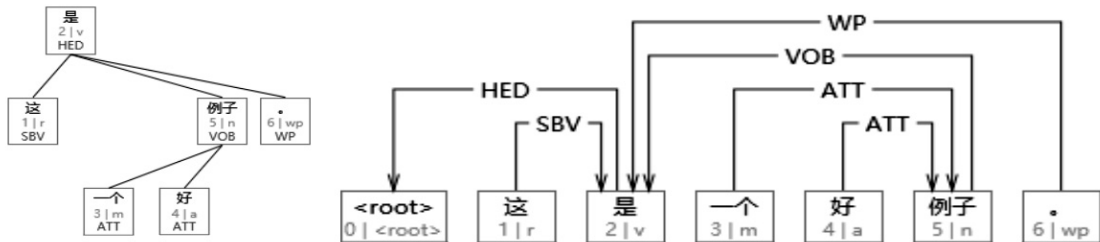


图1 “这是一个好例子”依存树图(左)和依存有向图(右)<sup>①</sup>

作者简介: 1. 杨 牧, 男, 南开大学汉语言文化学院硕士研究生;

2. 蔡言胜, 男, 南开大学汉语言文化学院副教授, 文学博士。

①图 1 中的“root”实际对应的是头词“是”。

从依存树图中,可以看出词语之间的层次关系;从依存有向图中,则更有利于看出依存关系的类型,以及依存关系两端词语的地位,即支配与从属。同时,上文提到的四条公理在例句中都有所体现。其中,例句中的“是”不依存于其他成分,因此,它是独立的,并且句中只有“是”一个词语是独立的。这符合公理1。除“是”之外的所有词语都直接依存于其他词语,如“这”直接依存于“是”。这符合公理2。图1中只存在向下的分叉,不存在向上的分叉。这说明一个词语可以有多个从属词,但只能有一个支配词,如“例子”有两个从属词“一个”“好”和一个支配词“是”。这符合公理3。“一个”直接依存于“例子”,处在中间的“好”依存于“例子”。这符合公理4。简言之,依存语法认为,词语之间的地位是不平等的,一方从属于另一方。这种不平等的关系就是依存关系。

## 二、依存语法研究简述

### (一) 依存语法的产生和发展

严格来讲,法国语言学家吕西安·泰尼埃的遗作《结构句法》于1959年的发表,标志着依存语法的正式诞生。这个时间虽然略晚于乔姆斯基的《句法结构》(1954),但也引起了以德国学者为代表的语言学家的关注。从泰尼埃的论述中,学者们认识到了依存语法和短语结构语法的本质区别,这在欧洲尤其是德国引发了运用依存语法理论解决问题的热潮。值得注意的是,生成语法此时已经统治了美国语言学界,但仍有学者将关注的目光投向依存语法。Hays正式提出了“依存”和“依存语法”两个术语,并且形成了一种完全基于依存关系的句子结构分析方法<sup>[2]</sup>。

泰尼埃与Hays是今天公认的现代依存语法的先驱,在两位学者之后,依存语法理论的发展势头十分迅猛。其中,产生广泛影响的主要有四家,即理查德·哈德森的“词语法(Word Grammar)”理论、Mel'čuk的“意义—文本理论(Meaning-Text Theory)”、Petr Sgall等人的“功能生成描述(Functional Generative Description)”理论、Stan Starosta的“词格(Lexicase)”理论。

“词语法”理论认为,语法就是语言中所有的词构成的网络,语言中不存在短语这一级单位,词与词通过依存关系组织在一起<sup>[3] (P95)、[4]、[5] (P117)</sup>。该理论指出,语言的各个层级之间没有明显的界限,语言的各个子系统相互交织,为了处理这种情况,“词语法”选择运用“关系(relation)”连接“实体(entities)”

以回避层级问题。其中,哈德森用来说明依存关系的“依存有向图”得到普遍认可与广泛应用。“意义—文本理论”一开始就面向机器翻译,其关注点是在于意义的表达和理解<sup>[3] (P95)、[6] (P43)</sup>。意义和文本的关系是多对多的,相同意义可以由不同的文本来表达,相同文本也可以表达不同意义。在Mel'čuk看来,语言的生成比语言的理解更值得关注,因此,相比于句法分析,该理论更关注语言的生成。“意义—文本理论”是目前最具影响力的依存语法理论,并且已广泛应用于自然语言处理上。“功能生成描述”理论同样与自然语言处理密切相关<sup>[3] (P96)、[7]</sup>。该理论把语言分为四个层次:词汇层、形态层、表层句法层、深层语法层。词汇层指的是原始文本;形态层是过滤形态之后的文本;表层句法层就是通常所说的句法层;深层语法层则是语义层。围绕这一理论,Petr Sgall等人建立了目前最大的依存树库——布拉格依存树库(Prague Dependency Treebank)和最大的面向应用的配价词典——捷克语动词配价词表(The Valency Lexicon of Czech Verbs)。“词格”理论也是一种依存理论,但在Stan Starosta去世后,其影响力逐渐衰落<sup>[3] (P96)</sup>。

### (二) 自动句法分析的研究

与其他理论相比,依存语法的最大优势是在于从它诞生起就和应用联系在一起,而自动句法分析则是这些应用的基础。早期的依存分析是基于规则的分析,首先是依据于语言学家归纳出的句法规则建立起语法知识库,然后根据语法知识库,对文本进行分析。语言学家们有意识地运用语言理论,积极地尝试句法分析,开了自动句法分析的先河。Hays曾提出一种识别句子合法性的句法识别器,并启发了后来CYK算法的诞生<sup>[2]</sup>;Menzel则提出面向依存分析的“加权约束依存语法”<sup>[8]</sup>;Gitguet & Vergne依据泰尼埃的理论,提出一种依存分析和组块分析相结合的句法分析器<sup>[9]</sup>。这些方法对依存分析的应用提供了有力的工具,扩大了这一理论的影响。但是问题也显而易见,语言中本来就存在着语法理论不好解决或解决不了的问题。为了提高准确性,学者们继续探索更合适的理论或者是诉诸于其他方法。

随着大规模语料库的建立,海量的语料为依存分析提供了数据来源,基于统计的依存分析开始产生。这种方法把语料库当作语言知识的全部来源,通过训练能够得到正确率较高的句法分析器。Yamada & Matsumoto采用“支持向量机(Support Vector

Machine)”，训练出接近当时最好的短语结构分析器性能的依存分析器<sup>[10]</sup>；McDonald 等采用“生成树算法”，有效地进行了非标记依存关系的句法分析<sup>[11]</sup>；Sagae 基于儿童语言转写语料库，开展了依存分析研究<sup>[12]</sup>。同时，还出现了将规则与统计方法相结合的依存分析，学者们在基于规则的依存分析中加入了统计方法，这种方法有助于解决一些原来不好处理的棘手问题。Gala 利用施乐公司的增量深层句法分析系统(Xero Incremental Deep Parsing Sysytem)所开发的分析器，具有从万维网自学习的能力<sup>[13]</sup>；Schneider 的 Pro3Gres 分析器，一方面依托于语言学家所归纳的语法规则，另一方面把来源于宾州树库的词汇化统计数据作为经验依据，有效提高了分析器精度<sup>[14]</sup>。

### （三）汉语依存分析的研究

互联网技术的蓬勃发展，语言学理论的逐渐成熟，催生了一场计算语言学的“革命”，从事于中文信息处理的学者们也意识到依存语法在句法分析上的巨大潜力。

黄昌宁等介绍了一种基于语料库的依存分析<sup>[15]</sup>；周明、黄昌宁提出了一种基于规则和统计的汉语依存分析模型<sup>[16]</sup>；刘伟权等初步建立起汉语依存关系的层次体系<sup>[17]</sup>；Zhou 结合浅层短语结构分析和深层依存分析所研制的分析方法，已应用于汉日机器翻译<sup>[18]</sup>。值得注意的是，在以依存分析为主题的 2006 年、2007 年 CoNLL (Conference on Computational Natural Language Learning) 中，汉语的依存分析精确度和英语、意大利语等印欧语言同属于高分区。由此可见，汉语依存分析研究虽然晚于国外，但也取得了不少成果，并且这些成果很多已经应用于实践。

需要指出的是，如今已经产生了一些成熟的面向汉语的句法分析工具。其中，NLTK (Natural Language Toolkit) 是最经典的自然语言处理工具包，在 Python 上可以实现词性标注、依存分析等任务。同时，NLTK 还自带大量英语语料，在引入分词后的汉语语料时，也适用于汉语的依存分析。LTP (Language Technology Platform) 是哈尔滨工业大学研发的自然语言处理基础技术平台，加载训练后的模型，能够实现分词、词性标注、依存句法分析、语义角色分析等功能。LTP 支持在 Python 中调用 pyltp 库和网页直接使用两种方式。SpaCy 诞生于 2014 年，它是工业级强度的自然语言处理包。在 Python 中调用 SpaCy 库，

可以实现自然语言处理的各种基础操作、信息提取和深度学习预处理。此外，还有一些基于 Python、Java 或其他编程语言的工具包，都可应用于汉语依存分析，只是会在精确度上有所差异。语体、文本平均句长、语法歧义数量等因素不同，训练出的模型也会不同，使用者可以加载自己训练好的模型以适应自己的需求。

## 三、依存分析的主要应用范围

### （一）文本理解

机器无法像人类一样直接理解文本，文本的批量处理是建立在解构文本的基础之上的，这就很可能会出现理解歧义或理解偏差。比如，“张三的父亲是谁？”这句话经过分词后，可以得到“张三”“的”“父亲”“是”“谁”五个词语。在不考虑句法关系的情况下，机器会得到两种解读：一种是正确理解，询问“张三的父亲”是哪个人；另外一种则是错误理解，询问“张三”是谁的父亲。“张三的父亲是谁？”的依存树图可如图 2 所示：

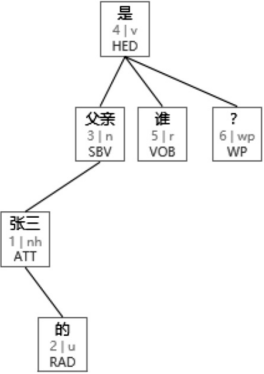


图 2 “张三的父亲是谁？”依存树图

从图 2 可以看出，对“张三的父亲是谁？”的理解只有一种。“谁”和“的”“父亲”并不存在依存关系，因此，第二种理解是错误的。由此可见，依存分析能够显著提高文本理解的正确率。

### （二）事件抽取

文本的关键信息是句子所要表达的事件，其中，谓词最能够体现事件的性质、状态、属性或动作。经过依存分析，句子可以被整合为具有依存关系的树结构，提取核心谓词所对应的节点就可以把事件抽取出来。在事件抽取时，通常需要关注的节点是核心谓词、与核心谓词并列的谓词、核心谓词的宾语。这里不妨以“张三今天吃了馒头，喝了啤酒”为例加以说明，其依存树图可如图 3 所示：



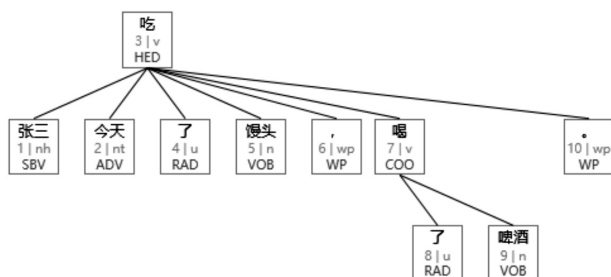


图3 “张三今天吃了馒头，喝了啤酒”依存树图

从图3可以看出，“张三今天吃了馒头，喝了啤酒”中的核心谓词为“吃”，宾语为“馒头”；与核心谓词并列的谓词为“喝”，宾语为“啤酒”。那么，该句的事件就是“吃馒头”和“喝啤酒”。加上与谓词具有不同类型依存关系的节点，我们还可以得到事件的主体和时间。由此可知，事件抽取不仅能够有效处理结构化和非结构化的文本数据，快速地获得文本的关键信息，而且还能够根据抽取出来的结构化数据，生成我们所需要的信息。

### （三）情感分析

情感分析也称“意见挖掘”“倾向性分析”，它的主要任务是判断文本的主观态度、评价、感情色彩。其中，最为常见的是对评论的情感分析，通过分析结果，研究者可以有针对性地调整产品计划。其基本流程是首先把句子依存分析为词语级别，接着根据情感词典得出词语的情感得分，然后处理否定逻辑和转折逻辑，得分加权求和即可得到整个句子的情感色彩。

### （四）机器翻译

机器翻译是自然语言处理的主要课题之一。如前所述，依存分析在“意义—文本理论”中扮演着重要角色。其基本流程是首先把A语言依存分析为树结构，再把词语翻译为B语言；接着根据句法结构，把翻译为B语言的词汇组织成合乎B语言语法的句子。前一过程与依存分析密切相关，后一过程则与语言生成密切相关。可以说，依存分析在机器翻译过程中起到的是理解语言的作用。

### （五）树库搭建

上文曾经提及一些基于语料库的依存句法分析的研究，实际上，依存分析同样能有效帮助搭建语料库。在早期语料库搭建时，许多工作需要人工完成，如分词、词性标注、句法分析等。在引入句法分析技术后，这类工作已经可以由机器自动完成。依存分析尤其有助于树库的搭建，树库不同于普通语料库，它不仅

够储存句法分析的结果，而且能对语言研究和自然语言处理起到辅助作用。虽然自动分析的正确率略显不足，但基本能够满足使用的需要。如果需要更高的准确度，可以将自动分析的结果加以人工校正。

## 四、研究现状及趋势

本文采用CiteSpace软件，对Web of Science核心数据库所收录的文献进行检索，主题为“dependency parsing”，时间跨度为1985年1月1日到2020年12月31日，共获得1339篇文献。基于文献分析所得数据，下面，主要从年度发文量、关键词、学术热点和研究趋势、文献来源四个方面，对依存句法分析的研究现状进行分析。

### （一）年度发文量

我们对1985～2020年依存分析研究的年度发文量进行了统计，并依据Web of Science的“引文报告”功能绘制出分布图。具体如图4所示：

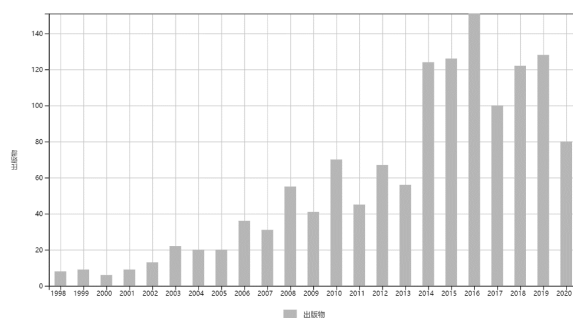


图4 依存分析研究年度发文量分布图

从图4可以看出，1985—1997年，Web of Science中并没有出现依存分析研究文献，结合上文的简述，可以得知，这一时期依存分析尚处于起步阶段；1998—2002年间，每年有10篇左右的相关文献被收录，这说明依存分析已引起学界注意，但研究成果相对匮乏，发展较为缓慢；2003—2013年间，依存分析研究进入新阶段，每年发文量均在20篇以上；从2014年开始，发文量迅速增多，表明依存分析已成为研究热点。

之所以会出现上述现象，主要是与技术手段、应用需求等因素有关。在起步阶段，由于技术手段尚不成熟，相关成果大多是出现在学术会议上或直接服务于应用，而学术期刊则基本没有刊发；2003—2013年间，随着技术手段的不断成熟、应用需求的逐步扩大，学界对依存分析的关注度也持续提升；2014年之后，技术手段更为成熟，研究热度指数继续上升，依存分

析的发文量也得以显著增加。

## （二）关键词共现知识图谱分析

在 CiteSpace 软件中，首先选择相关研究文献，设置时间切片为一年；接着选择“Keyword”，将“g-index”参数设置为 25；我们共得到节点 465 个，连线数 1363 条。然后设置节点显示条件为“By Freq”，Threshold 为 10（显示频次大于 10 的节点），可以得到关键词共现知识图谱。具体如图 5 所示：



图 5 关键词共现知识图谱

在去除检索词“dependency parsing”之后，最大节点为“parsing（句法分析）”和“natural language processing（自然语言处理）”，与它们密切相关的还有“dependency（依存）”。句法分析是自然语言的主要任务，如此高的共现率，反映出依存分析在句法分析领域内的显著地位。

数量和频次都占绝对优势的节点是语义理解类节点，从图 5 可以看出，语义理解类节点主要有“information（信息）”“comprehension（理解）”“sentence comprehension（句子理解）”“complexity（复杂性）”“sentiment analysis（情感分析）”，同时，这些节点大致呈现出聚合关系。这说明，依存分析聚焦于语义的理解，句法分析的目的之一是使计算机能够理解自然语言，而依存分析的主要应用也都是基于语义理解作出的。

值得注意的是，“constraint（约束）”也是个频次很高的节点。简单来讲，约束是一种规则，句法分析可以看作是将字符串按一定规则分析的“约束满足问题（Constraint Satisfaction Problem）”。约束满足问题主要包含三组集合：变量集合（X）、每个变量的值域集合（D）、描述变量取值的约束集合（C）。

就依存分析而言，句中的词语是变量，句子为变量集合 X；每个词语可能的支配词和依存关系则是值域，所有词语的值域集合即是 D；语法规则集合即是 C。这种方法特别适合于依存分析，它能够有效评估句子合乎语法的程度，并且能根据可用时间得到不同准确度的结果。

在图 5 中，还有一类与句法分析器密切相关的节点，如“model（模型）”“algorithm（算法）”“corpus（语料）”等。构建句法分析器一直是依存分析的核心内容，目前的主流方法是通过大量语料训练出准确率高的模型。训练过程中势必会涉及到算法，经典算法主要有朴素贝叶斯、支持向量机等，算法不同，精确度也会有所不同，因此，算法与依存分析密切相关。

此外，“brain potential（大脑潜力）”“working memory（工作记忆）”“sentence processing（句子加工）”则涉及心理和认知领域。认知是目前学术界所关注的热点问题，语言学、计算机、文学、哲学、生物学等很多学科，都开展了与认知领域的相关研究。而依存分析的主要目的是使计算机像人一样理解语言，认知研究可以为依存分析提供参考。

## （三）历时研究热点与发展趋势

突现词(burst)能够反映一段时间内的学术热点。在 CiteSpace 软件中，对“Keyword”进行分析，可以得到突现关键词。从整体上来看，在 2014 年发生了一定变化，学术热点由“information（信息）”转变为具体的“sentiment analysis（情感分析）”。具体如图 6 所示：

Top 3 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	1998 - 2020
information	1998	3.65	2008	2014	=====
dependency parsing	1998	6.73	2016	2020	=====
sentiment analysis	1998	3.36	2016	2020	=====

图 6 突现关键词图示

由于从关键词中得到的突现词较少，我们又选取了“Term”来分析所得到的突现词，以作为补充。具体如图 7 所示：

Top 6 Terms with the Strongest Citation Bursts

Terms	Year	Strength	Begin	End	1998 - 2020
language	1998	3.68	1999	2007	=====
eye movement	1998	3.46	2001	2007	=====
machine learning	1998	3.65	2009	2014	=====
parsing	1998	4.27	2011	2015	=====
sentiment analysis	1998	4.51	2016	2020	=====
deep learning	1998	3.97	2018	2020	=====

图 7 术语突现词图示

从图 7 可以看出, 1999—2007 年间的研究热点是“language (语言)”, 这一时期内的研究整体上较为宏观; 其中, 2001—2007 年的热点为“eye movement (眼动)”, 眼动现象可以体现语言不同成分的加工速度与心理表征情况。2009—2014 年间的研究热点是“machine learning (机器学习)”, 机器学习是依存分析实践的常用方式, 训练集和测试集配合算法能够训练出所需模型, 这种方法一直持续至今。2011—2015 年间的研究热点是“parsing (句法分析)”, 它已包含于主题内; 2016—2020 年间的研究热点则是“sentiment analysis (情感分析)”。2018—2020 年间的研究热点是“deep learning (深度学习)”, 深度学习是机器学习新的研究方向, 旨在通过数据使机器获得类似于人的分析能力, 以网络表示出概念, 其性能要高于传统机器学习, 借助于这种方法, 依存分析的准确率能够大大提高。

在上述基础上, 我们还绘制出 1999—2020 年依存分析的时区图, 具体如图 8 所示:

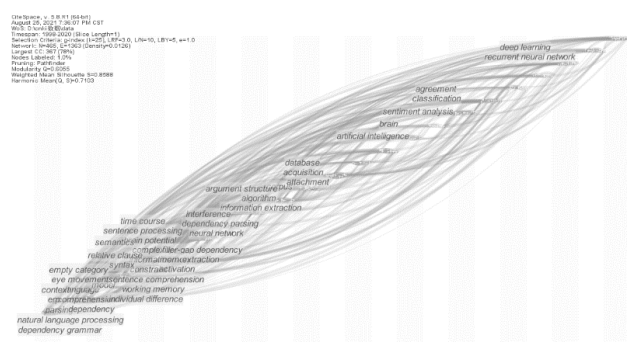


图 8 依存句法分析研究时区图

从图 8 可以看出, 其中的整体趋势和突现词图基本一致。总体上看, 依存分析从依存语法理论和自然语言处理的早期实践开始, 呈现出走向深度学习和循环神经网络的发展趋势。学界目前聚焦于算法设计, 通过更先进的算法来实现更高准确率的依存分析。

#### (四) 学科分布和国家分布

在 CiteSpace 软件中, 选择“Category”, 其他设置同上; 然后将节点标准参数设置为“By Citation”, 将 Threshold 设置为 30, 可以得到所收录文献的学科分布图。具体如图 9 所示:

从图 9 可以看出, 依存分析的最大研究主体为计算机学界, 其次是语言学界, 再次是工程学界和心理学界。自然语言处理属于计算机学与语言学这两个领域的交叉学科, 因此, 这样的结果也是符合我们的预

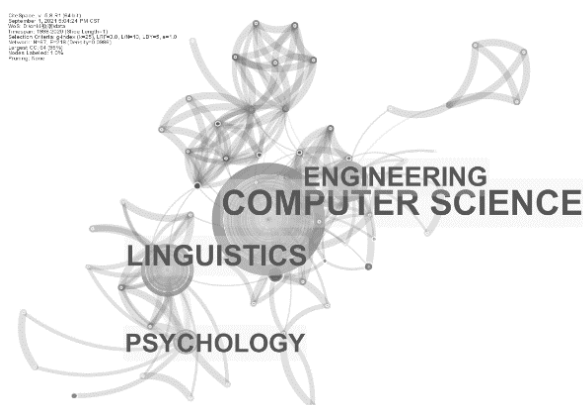


图 9 收录文献的学科分布图

期的。依存分析研究主要体现在两个方面: 一是面向于实践应用; 二是面向于理论研究。由于计算机学科占据了很大比重, 今后的依存分析研究可能会出现向计算机应用发展的趋势。

在 CiteSpace 软件中, 选择“Country”, 其他设置同上; 然后将节点标准参数设置为“By Citation”, 将 Threshold 设置为 5, 可以得到所收录文献的国家来源分布图。具体如图 10 所示:



图 10 收录文献的国家来源分布图

从图 10 可以看出, 美国占据最大的节点, 德国、日本、韩国也占有一席之地。值得注意的是, 来源于中国的文献数量仅次于美国, 这说明中国学者在依存分析方面的研究已得到国际认可, 并处在前沿位置。

综上所述, 依存句法分析的理论来源是依存语法。法国语言学家泰尼埃正式提出了现代依存语法, 之后, 在依存语法领域内发展出各种理论, 为依存句法分析提供了强有力的理论支撑。同时, 依存语法一直具有



与实践应用紧密结合的优良传统，Hays、哈德森均设计过句法分析器，Mel'čuk 也在机器翻译的实践中进一步完善了“意义—文本理论”。通过对 WOS 相关研究文献的统计和分析，可以看出，依存句法分析研究的发文量呈递增趋势，研究内容聚焦于语义理解和算法设计，研究主体为计算机学界和语言学界。早期的依存分析侧重于理论研究和认知研究，后来逐渐走向具体的算法设计、高性能句法分析器实现，其中，语义始终是依存分析的关注点。可以说，正是由于依存语法理论大量运用在依存分析实践中，才催生出成熟的自然语言处理工具，而自然语言处理工具的日益成熟，反过来又肯定了依存语法的价值。就目前的研究态势来看，自然语言处理领域正处于蓬勃发展时期，依存分析的进一步壮大也是必然的。

#### 参考文献：

- [1] Robinson, J. J. Dependency Structures and Transformational Rules[J]. *Language*, 1970, (2).
- [2] Hays, D. G. Dependency Theory: A Formalism and Some Observations[J]. *Language*, 1964, (4).
- [3] 刘海涛. 依存语法的理论与实践 [M]. 北京: 科学出版社, 2009.
- [4] 冯志伟, 周建. 赫德森的词语法理论 [J]. *现代语文*, 2018, (3).
- [5] Hudson, R. *Language Networks: The New Word Grammar*[M]. Oxford: Oxford University Press, 2007.
- [6] Mel'čuk, I. A. *Dependency Syntax: Theory and Practice*[M]. Albany: State University Press of New York, 1988.
- [7] 冯志伟, 周建. 布拉格学派的功能生成描述理论 [J]. *现代语文*, 2019, (7).
- [8] Menzel, W. Parsing of spoken language under time constraints[A]. *Proceedings 11th European Conference on Artificial Intelligence*[C]. 1994.
- [9] Giguët, E. & Vergne, J. Syntactic analysis of unrestricted French[A]. *Proceedings for the International Conference on Recent Advances in Natural Languages Processing*[C]. 1997.
- [10] Yamada, H. & Matsumoto, Y. Statistical Dependency Analysis with Support Vector Machines Proc[A]. *Proceedings 8th International Workshop on Parsing Technologies*[C]. 2003.
- [11] McDonald, R., Pereira, F., Ribarov, K. & Hajič, J. Non-Projective Dependency Parsing Using Spanning Tree Algorithms[A]. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*[C]. 2005.
- [12] Sagae, K. A Multi-Strategy Approach to Parsing of Grammatical Relations in Child Language Transcripts[D]. Ph.D thesis, Carnegie Mellon University, 2005.
- [13] Gala, N. Un Modèle D'analyseur Syntaxique Robuste Fondé sur la Modularité et la Lexicalisation de ses Grammaires[D]. Thèse de Doctorat en Informatique Université de Paris-Sud, 2003.
- [14] Schneider, G. Hybrid Long-Distance Functional Dependency Parsing[D]. Ph.D thesis, University of Zurich, 2008.
- [15] 黄昌宁, 苑春法, 潘诗梅. 语料库、知识获取和句法分析 [J]. *中文信息学报*, 1992, (3).
- [16] 周明, 黄昌宁. 面向语料库标注的汉语依存体系的探讨 [J]. *中文信息学报*, 1994, (3).
- [17] 刘伟权, 王明会, 钟义信. 建立现代汉语依存关系的层次体系 [J]. *中文信息学报*, 1996, (2).
- [18] Zhou, M. A Block-Based Robust Dependency Parser for Unrestricted Chinese Text[A]. *The Second Chinese Language Processing Workshop Attached to ACL2000*[C]. Hong Kong, 2000.

## Review and Development of Dependency Parsing

Yang Mu, Cai Yansheng

(College of Chinese Language and Culture, Nankai University, Tianjin 300350, China)

**Abstract :** The basic idea of dependency grammar is the dependency relationship between words. Dependency parsing takes dependency grammar as its theoretical source and algorithm as its realization method, and has value in both language research and practical application. Based on CiteSpace's visual analysis of the documents included in the WOS core database from 1985 to 2020, the number of articles published in dependent parsing is increasing, focusing on semantic understanding and algorithm design, and the main body of research is computer science and linguistics.

**Key words :** dependency grammar ; dependency parsing ; visual analysis