汉语增强依存句法自动转换研究

余婧思[1,2,3]，师佳璐[1,2,3]，杨麟儿[1,2,3*]，肖丹[4]，杨尔弘[1,3]
[1]北京语言大学 国家语言资源监测与研究平面媒体中心
[2]北京语言大学 信息科学学院
[3]北京语言大学 语言资源高精尖创新中心
[4]信阳学院 文学院
yujingsi1107@gmail.com

摘要

自动句法分析是自然语言处理中的一项核心任务，受限于依存句法中每个节点只能有一条入弧的规则，基础依存句法中许多实词之间的关系无法用依存弧和依存标签直接标明；同时，已有的依存句法体系中的依存关系还有进一步细化、提升的空间，以便从中提取连贯的语义关系。面对这种情况，本文在斯坦福基础依存句法规范的基础上，研制了汉语增强依存句法规范，主要贡献在于：介词和连词的增强、并列项的传播、句式转换和特殊句式的增强。此外，本文提供了基于Python的汉语增强依存句法转换的转换器，以及一个基于Web的演示，该演示将句子从基础依存句法树通过本文的规范解析成依存图。最后，本文探索了增强依存句法的实际应用，并以搭配抽取和信息抽取为例进行相关讨论。

关键词： 依存句法；汉语增强依存句法；自动转换

Transformation of Enhanced Dependencies in Chinese

Jingsi Yu[1,2,3], Jialu Shi[1,2,3], Liner Yang[1,2,3], Dan Xiao[4], Erhong Yang[1,3]
[1]National Language Resources Monitoring and Research Center Print Media Language Branch, Beijing Language and Culture University
[2]School of Information Science, Beijing Language and Culture University
[3]Advanced Innovation Center for Language Resources, Beijing Language and Culture University
[4]College of Chinese Language and Literature, Xinyang University
yujingsi1107@gmail.com

Abstract

Syntactic analysis is a key step of the natural language understanding process. Affected by the rule that each word can only have one entered arc in basic dependent syntax, many functional and semantic relationship between content words cannot be indicated directly by the dependent arc and label. At the same time, there is still room for further refinement and improvement of the dependencies in the existing dependency syntax system in order to extract coherent semantic relations from them. In the face of this situation, this paper develops a guidelines of enhanced dependency syntax for Chinese based on the Stanford Dependency Syntax. The main contributions are: prepositions and conjunctions, parallel structures, syntactic alternations and special syntactics. In addition, the paper provides a converter for Chinese enhanced dependency transformation by python, as well as a web-based demo that parses sentences into

* 通讯作者
基金项目：国家语委项目（ZDI135-131）；中央高校基本科研业务费（北京语言大学梧桐创新平台，21PT04）

dependency graphs through universal syntactic dependencies and the specification of this paper. Finally, the paper explores practical applications of augmented dependency syntax such as collocation extraction and information extraction.

Keywords: dependency syntax , enhanced dependencies in Chinese , automatic transform

1 引言

句法分析是自然语言处理当中的关键技术之一，它是对输入文本的句子进行分析以得到其句法结构的过程。依存句法分析是其中的一种表示形式，它用于分析输入句子的句法结构，将词语序列转化为树状的依存结构(李正华, 2013)，来捕获句子内部词语之间的修饰或搭配关系，描述句法结构。依存句法分析广泛应用于自然语言处理的多个领域，如在搭配抽取中，通过大规模的语料进行依存句法分析，从中抽取想要的依存弧以获得具有句法关系的词对，再通过词对之间的共现频次、互信息、联合熵等统计方法来说明词对之间的相关性；再如在信息抽取中，利用依存句法分析来抽取关系三元组，进而达到信息抽取的目的。

依存句法分析在准确地反映句法关系、描述句法结构的同时，也带有一些浅层的语义表示，但语义关系还不够明确，一些实词之间的关系没有直接明确地表示出来，且缺乏对句法转换的抽象。此外，一些依存标签被用于多种情况，难以区分，在自然语言理解的下游任务，如信息抽取、文本挖掘、语义分析中，就需要投入许多工作来处理语法树。因此，研究人员在依存句法的基础上提出了增强依存句法，来满足依存句法反映语义信息的需求。目前，增强依存句法在英语上已获得了有益的探索，并在信息抽取、关系抽取上得到了应用，但在汉语中还未见相关研究。

本文在斯坦福依存句法规范的基础之上，制定了增强的依存句法规范，从利于搭配抽取和自然语言理解的角度重新构建依存图，将实词之间的语义关系显性地展示出来，并统一句式转换中的依存句法关系，以便于进一步的研究和应用。

2 相关研究

斯坦福依存句法框架中提出了几种对句法结构进行面向语义修改的方案，引入了Collapsed Dependencies 和 CCprocessed Dependencies 两种形式(de Marneffe and Manning, 2008)。Collapsed Dependencies 折叠了涉及介词（包括功能类似于介词的多词结构）、连词以及关系从句所指信息的依存关系，从而得到实词之间的直接依存关系，这对于关系抽取应用很有用。此外，该方案还考虑了其他依存关系，如关系子句及其先行词、xsubj 关系和 pobj 关系，甚至破坏了树结构，将依存关系结构转换为有向图。CCprocessed Dependencies 在 Collapsed Dependencies 的基础上，增加了并列词的传播，即当句中存在并列连词时，一个并列词的依存关系可以传播到其他并列词。这样，通过额外增添的和增强的关系，实词之间的关系更加明显，多数涉及实词之间关系的系统通常会采用这两种形式。

通用依存项目（Universal Dependences，简称UD）在第一个版本（UD v1）中(Nivre et al., 2016)同样提出了增强依存（Enhanced Dependencies）的概念，它增加额外的依存关系来表示先行词与关系从句中某个成分之间的主语关系，并在并列词之间传播关系。Schuster 和 Manning(2016)详细描述了增强英语UD（enhanced English UD），并介绍了更适用于自然语言理解任务的增强++表示（enhanced++ representation），对量名词短语和轻名词结构、多词介词、并列的介词或介词短语、关系代词的表示作了改进，并提供了转换器，实现了从基础依存句法（Basic Dependencies）到增强英语UD图和增强++英语UD图的转换。UD V2(Nivre et al., 2020)在先前研究的基础上，定义了五种增强类型：1. 省略谓语的空节点；2.并列项的传播；3. 控制和提升主语；4.关系代词；5. case 信息。

Candito 等人(2017)给出了更进一步的改进，他们沿着两个方向来丰富增强依存框架：扩展非限定性动词的论元依存类型（包括分词、控制名词和形容词、非限定动词以及更多不定式动词的情况）、中和句法转换（包括被动语态，中间被动语态，非人称和使役）。Nivre 等人(2018)评估了向UD现有树库添加增强依存句法的两种跨语言技术，分别是为英语开发的基于规则的系统和在芬兰语、瑞典语和意大利语上训练的数据驱动系统，结果表明，这两种系统都足够精确，可以在现有的 UD 树库中引入增强依存关系。

由于英语增强 UD 的转换不支持 Python，且覆盖范围有限，Aryeh 等人(2020)制定了 BART 表示，引入了覆盖范围广的、数据驱动的、语言学上合理的增强依存转化集，包括四种结构的增强：嵌套结构、并列结构、句式转换以及以事件为中心的表示，该转化集使事件结构和许多词汇关系更加明确。此外，他们提供了一个易于使用的开源 Python 库 pyBART[0]，用于将英语 UD 树转换为增强 UD 图或 BART 表示。该库可以作为一个独立的包工作，也可以集成在一个 spaCy 流水线中。当在信息抽取任务中进行评估时，使用增强依存分析结果，可以通过更少的训练样本得到更多的信息，因此BART 表示比增强 UD 产生更高的提取分数。

3 增强的依存句法规范

本文基于斯坦福依存句法，在借鉴英文增强依存句法思想的基础上，制定了增强依存句法标注规范。该规范通过修改依存标签、添加弧或节点的方式，将依存句法树转换为可以表示更多信息的依存句法图，显性地展示实词之间的语义关系，从而更有利于自然语言处理下游任务的应用。

3.1 介词和连词的增强

介词和连词是构造句子时较为常用的词类，对于句意的理解有很大的影响，当一句话中介词或连词发生改变时，句意可能会发生巨大的改变。例如在“我给小王讲了个故事”和“我替小王讲了个故事”这两句话中，只有介词“给”和“替”发生了改变，但句意却完全不同，在前一句话中，动作“讲”的对象是“小王”，而后一句话中，“讲”的对象并没有在句中出现。

在自然语言理解任务中，由于依存句法还带有一定的语义信息，因此常常通过依存句法来识别和提取所需信息，但是，当句中含有介词或连词时，基础依存句法不能完全满足自然语言理解任务中直接通过词之间的依存弧提取信息的需求，因此，需要对介词和连词来进行增强，以更好地适应自然语言理解及其下游任务。

介词的增强 在基础依存句法规范中，当一个介词短语修饰其他实词时，依存弧通常连接在介词短语中的实词和被修饰词上，增强依存句法规范要求把介词添加在该弧的依存标签上，原标签与添加的介词中间用“_”连接，如图1中将该依存弧的依存标签修改为“nmod:prep_向”。这有助于消除介词短语修饰时的歧义，促进实词之间关系的提取，特别是当只通过两个节点之间的依存弧来提取信息时，增强后的依存句法包含的信息更多，更有利于语义理解。

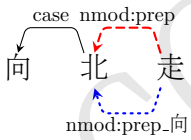

图 1: 介词的增强标注示例

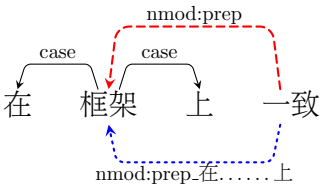

图 2: 框式介词的增强标注示例

除了此类单独出现的介词，汉语中还有一类特殊的介词，即框式介词。刘丹青在《汉语框式介词》一书中最早引入“框式介词”的概念，认为“框式介词是由前置词和后置词构成的使介词支配的成分夹在中间的一种介词类型”(刘丹青, 2002)。在增强依存句法中，用依存弧连接框式介词短语与被框式介词短语修饰的实词时，依存标签中也要把框式介词的两个部分都加上，两个部分中间用省略号连接。如图2中的依存标签“nmod:prep_ 在……上”。

除了 nmod:prep，在被分析为 advcl:loc 等的从句当中，如果从句中存在标签为 case 的依存弧，则在增强依存弧中也要将该弧指向的词添加在连接主句和从句的依存弧 advcl:loc 上。

连词的增强 并列结构是人类语言中最原始最普遍的一种结构式，并列连词可以连接词、短语或小句之间的并列。基础依存句法中用依存标签为 conj 的依存弧来连接句中并列的部分，用依存标签为cc的依存弧连接并列连词与并列项的其中一项。

在增强依存句法规范中，通过在依存标签 conj 上添加依存弧 cc 所指的并列连词，可以使得并列项之间的语义关系更加明晰，特别是当句中出现多个并列连词时，并列结构之间的并列

[0]https://pybart.apps.allenai.org/

类型就会更加明确，如图3，将依存标签修改为“conj_和”“conj_或者”，这三组并列结构中并列项之间的关系可以一目了然，计算机在提取并列项间的语义信息时也更加便利。

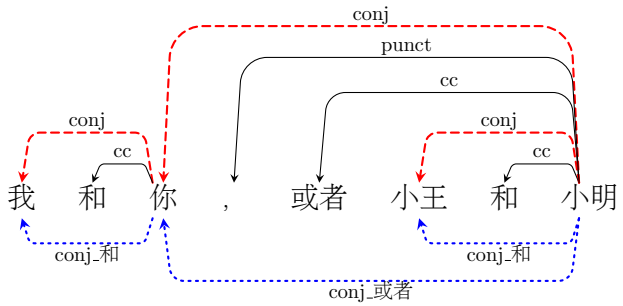

图 3: 连词的增强标注示例

3.2 并列结构的传播

在基础依存句法规范中，多个并列项之间由其中一个并列项作为父节点，来连接其他的句子成分，如主语、宾语。但从语义上来说，并列项之间通常是共享这些句子成分的。因此，在增强依存句法规范中，两个并列的结构共享其父节点和子节点，且依存标签相同。

并列成分的传播 主语、谓语、宾语、时间地点状语等成分在句中都可能由并列结构来承担，在基础依存句法中，只标出其中一个并列项与其支配词和从属词间的依存句法关系，在增强的依存句法图中，需要将并列结构中的其他项与支配词或从属词间的依存关系也表示出来，如图4为并列谓语的增强。

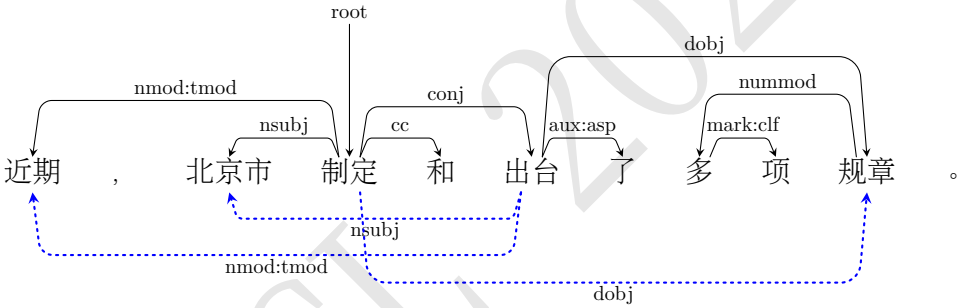

图 4: 并列谓语的传播标注示例

偏正短语中，并列修饰语、状语或中心语也需要传播其支配词或从属词。如图5为并列修饰语修饰中心语的情况，如图6为单个状语修饰并列中心语的情况。这时，在增强依存句法中，就需要补出未被标出修饰关系的修饰语、状语与中心语之间的依存弧。

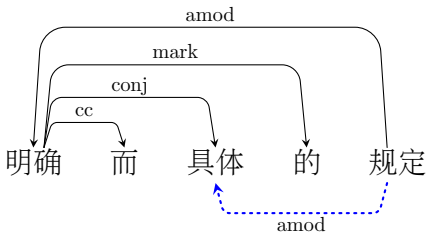

图 5: 并列修饰语的传播标注示例

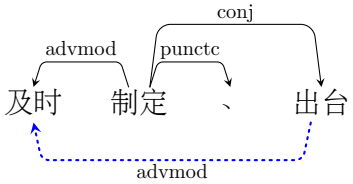

图 6: 并列中心语的传播标注示例

同位语的传播 由于同位语所指代内容相同，在句中承担的句子成分也相同，因此，本文把它看成是一种特殊的并列形式。在基础依存句法中，同位语之间用依存弧 appos 连接，其他句法成分连接在同位语的后一部分上。在增强依存句法中，需要将句中实词与同位语后一部分之间的依存关系，通过增加弧的方式添加在同位语的前一部分上，如图7。

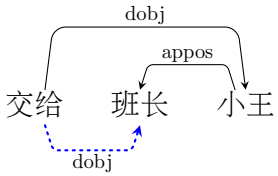

图 7: 同位语的传播标注示例

3.3 句式转换

依存句法是从句子的表层语法来进行分析的，缺乏对句式转换的抽象。同样的语义，采取不同的表述方式，实词之间的依存关系就可能会发生变化。如被动句“书被小王拿走了”和主动句“小王拿走了书”，这两句的句意是完全相同的，但由于句子形式改变，“书”和“小王”之间的依存关系也不同，被动句中，它们之间的关系用 nsubjpass 来表示，意为被动主语，而主动句中，他们之间则用表示宾语的 dobj 来连接。

上述情况对自然语言理解及其下游任务造成了一定的局限，为了使语义分析更简单，本文利用规则统一了句式转换，借助深层的语义关系将不同句式中相同实词间的关系用同样的依存关系来表示。

被动句的转换 在基础依存句法规范中，被动主语，一般为意义上的受事，通常用依存标签为 nsubjpass 的依存弧与谓语连接，意义上的施事主语仍用表示主语的 nsubj 标签与谓语连接，而在被动句转换后的主动句中，施事主语在主动句中形式上做主语，被动主语则作为主动句中的宾语。

为了将被动句与主动句中实词间的依存关系统一，本文采用更为常用的主动句中的依存关系作为标准，即被动主语与谓语之间的依存关系为 dobj。因此，在增强依存句法中，添加一条弧从句中的谓语指向被动主语，依存标签为 dobj，如图8。

此外，修饰成分是被动短语的偏正短语，在基础依存句法规范中，依存弧从中心语指向被动短语中动词，依存标签为 acl，这种表被动的短语在语义上，其中心语通常是被动短语中动词的受事，在转换后的主动句中，中心语是该动词的宾语。因此，在增强依存句法中，添加一条依存弧从被动短语的动词指向中心语，其依存标签为 dobj，如图9。

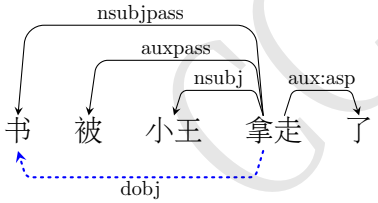

图 8: 被动句的转换标注示例

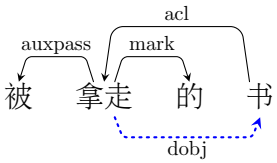

图 9: 被动短语的转换标注示例

有一种比较特殊的被动句，其动词是认作或任选义的动词，如句子“小王被选为班长”、“被誉为‘中国国酒’的茅台酒”，其转换为主动句式为“选小王为班长。”、“誉茅台酒为‘中国国酒’。”，将动词与其后的“为”拆分开来。但在基础依存句法当中，“选为”、“誉为”被当作一个词，难以拆开。面对这种情况，本文尊重了原本的分词及词性规范，在增强的依存句法当中，对此类动词不做特殊考虑。

“把”字句的转换 “把”字句是汉语特有的一种句式，其句式语义主要是主语对动词的受事作了某种处置。“把”是一个介词，它将原来充当动词宾语的受事成分提到动词之前，因此，“把”字句可以通过句式转换将“把”引导的宾语还原到动词宾语的位置。如图10中，“我把苹果吃了。”可以转换为“我吃苹果。”，因此在增强依存句法中增添了一条依存弧从该动词指向“把”引导的宾语，依存标签为 dobj。

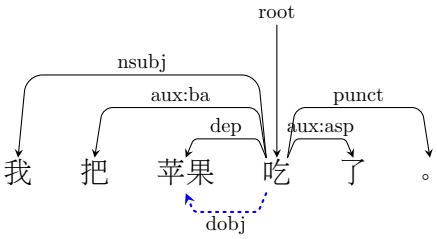

图 10: “把”字句的转换标注示例

形容词修饰语的转换 在偏正短语中，形容词短语来修饰名词中心语，那么这个偏正短语可以转换为以该中心语为主语、以该形容词为谓语的主谓短语，如图11中，“一个漂亮的女孩”可以转换为“女孩漂亮”。为了更好地捕获这些语义信息，在增强的依存句法中，为句子增添了一条从该形容词修饰语指向中心语的依存弧，依存标签为表示主语的 nsubj。

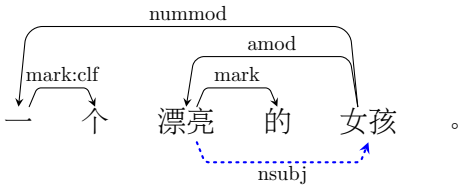

图 11: 形容词修饰的转换标注示例

动词短语修饰语的转换 一个动词短语来修饰名词中心语，如果在动词短语中，该动词不含宾语的话，那么中心语可能为该动词的受事。如图12，在语义上，“饭”是“做”的受事，那么该句可以转化为“妈妈做饭”，此时，“饭”是“做”的宾语。因此，在增强依存句法中，要增加一条依存弧由动词短语中的动词指向中心语，依存标签为 dobj。

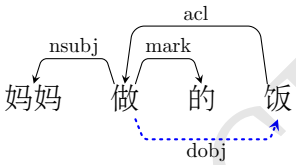

图 12: 动词短语修饰的转换标注示例

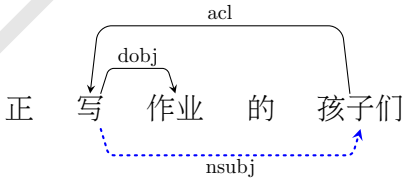

图 13: 动词短语修饰的转换标注示例2

如果修饰名词中心语的动词短语有宾语但不含主语的话，那么这种偏正短语也可能能够转化为一个中心语作主语、动宾短语作谓语和宾语的句子，如图13中“正写作业的孩子们”可以转化为“孩子们正写作业”，此时，“孩子们”为“写”的主语。那么，在增强依存句法中，需要增添从修饰语中的动词指向中心语的依存弧，其依存标签为 nsubj。

在基础依存句法当中，如果动词修饰语既不包含 nsubj 弧，也不包含 dobj 弧，那么其中心语可能是转化后句子的主语，例如“漂泊的游子”转化为“游子漂泊”，也可能是宾语，例如“设置好的页面”转化为“设置页面”，也可能存在修饰语中谓词是动宾结构，但在分词时未拆开的情况，例如“在外打工的父亲”中“打工”被看作是一个词，这些情况本文暂不予考虑。

3.4 特殊句式的增强

兼语句 兼语句是由兼语短语作谓语的句子，其谓语中第一个动词的宾语也是后一谓词的主语(年玉萍, 2003)，这个词就叫做“兼语”。例如在“老师通知我开会”一句中，“我”既是“通知”的宾语，也是“开会”的主语。在基础依存句法当中，受限于一个节点只能有一条入弧的规则，只标注了第一个动词和兼语之间的宾语关系，而没有标注出后一谓词与兼语之间的主语关系。因此，在增强依存句法中，需要增添一条依存弧由后一谓词指向兼语，依存标签为 nsubj，如图14。

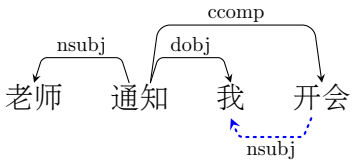

图 14: 兼语句的增强标注示例

连动句 连动句是现代汉语里一种特殊的句法结构，指的是谓语由两个或两个以上动词构成，在动词短语中间没有停顿，也没有关联词语，两个动词短语共用一个主语的句子(刘月华et al., 2001)。如在句子“外商来华投资。”中，“来华投资”是连动短语，它们的主语都为“外商”。但在基础依存句法中，只标注出第一个动词和主语之间的依存关系，因此，在增强依存句法中，应添加一条依存弧由连动短语中的其他动词指向主语，依存标签为 nsubj，如图15。

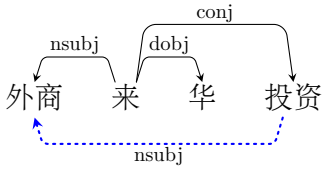

图 15: 连动句的增强标注示例

省略句 中文是一种话题驱动语言，为了表达的连贯性和简洁性，句子中常常省略某些语言成分，即句子存在缺省，本文讨论对句子中的主要结构即主语、宾语省略的增强。

含有动词性状语的句子中，存在状语中的动词和谓词共用一个主语的现象，由于汉语中的经济原则，那么状语或主句就可能省略主语。如图16 中，时间状语中省略了主语，但其实“吃饭”和“散步”的主语都为“他”。在基础依存句法中，只标出了“他”与“散步”之间的主语关系。那么在增强的依存句法中，还需要添加一条依存弧由“吃完”指向“他”，依存标签为 nsubj。

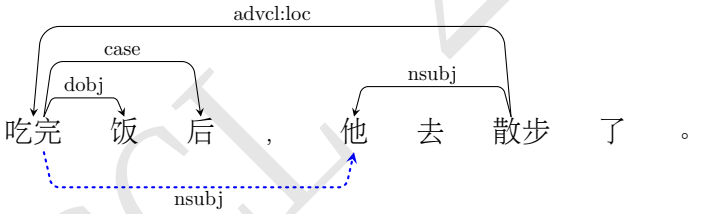

图 16: 省略主语的增强标注示例

在复句中，除了省略小句主语的情况以外，当几个小句的宾语相同时，也可能会省略小句中的宾语。如图17 中，第二个小句中没有宾语，但根据语义可知，其宾语仍为第一个小句中的宾语“小明”。因此在增强依存句法中，需要增添一条依存弧由省略宾语小句中的谓词“看见”指向其他小句中的宾语“小明”，依存关系标签为 dobj。

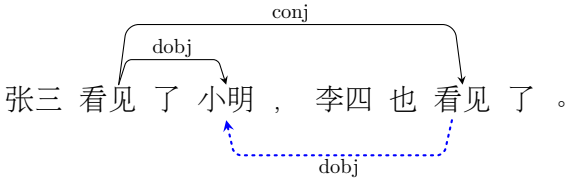

图 17: 省略宾语的增强标注示例

3.5 不确定情况的处理

在上述的规则中，也会产生一些不适应的情况，如句子“正睡觉的时候，妈妈回来了。”，如果按照增强依存句法规则，会把缺少主语小句中谓词“睡觉”的主语指向另一小句中的主语“妈

妈”，但是依照现实情况来看，“睡觉”的主语不可能是“妈妈”，其真正的主语需要联系上下文来确定。面对这些情况，本文并未放弃这几类增强规则，而是如图18，借用 Aryeh(2020)提出的 UNC=TRUE（不确定）这一概念，表示这条依存弧的正确性由用户来判断。

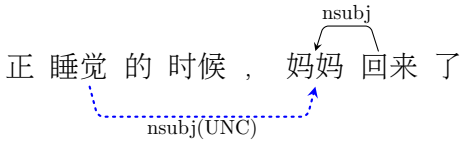

图 18: 不确定情况的处理示例1

同样的，汉语当中也存在复句中的某个小句省略了主语，但其省略的主语不是其他小句主语的情况，例如在“然而外祖母又怕都是孩子们，不可靠。”一句中，“不可靠”的主语是前一小句的宾语“孩子们”，而非前一小句的主语“外祖母”；再例如在“春游的时候，他告诉了我这件事。”一句中，“春游”的主语可能是“他”，也可能是“我”，也可能“他”和“我”都是主语，这需要根据句子的上下文来决定。此时，本文采取 Aryeh(2020)提出的概念 ALT=X，表示用户可以从中选择其一，如图19，其中 X 表示被省略主语或宾语的词在句子中的位置。

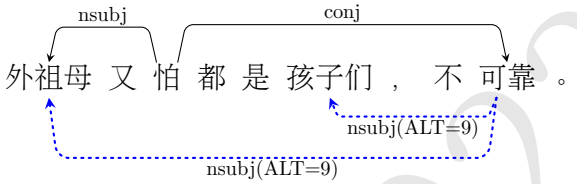

图 19: 不确定情况的处理示例2

3.6 依存句法增强器与演示平台

在斯坦福依存句法规范的基础上，本文提供了一个可以从基础依存句法分析到增强依存句法分析的转换器。在观察大量依存标注语料的基础上，寻找每类规则的规律，利用词性、依存弧的范围和指向、依存标签等约束实现了增强依存句法规范的规则转化。

此外，还提供了汉语依存句法增强转换在线平台[1]，如图20，可以将句子分析为基础依存句法和增强依存句法，并将它们可视化，便于比较和分析。

该界面分为四个部分，分别为输入句子搜索、选择示例搜索、基础句法依存演示、增强依存句法演示。用户可以在输入框中自主输入想要分析的句子，也可以在选择示例下拉框中选择，平台已经为17个汉语增强依存句法规则给出了示例演示。

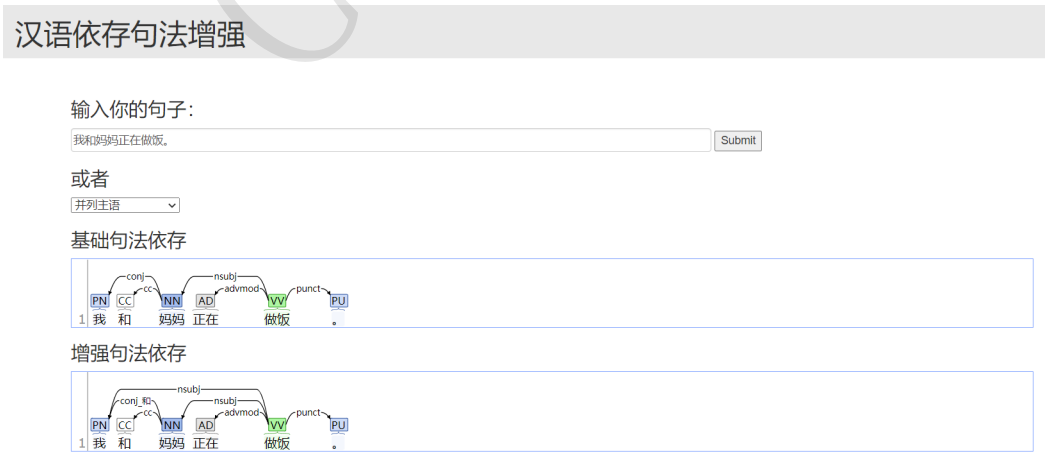

图 20: 汉语依存句法增强转换在线平台

[1]https://parser.litmind.ink

4 增强依存句法规范的实际应用

增强依存句法在基础依存句法的基础上扩充了实词间的依存关系，包含的句法和语义信息更多，在需要使用依存句法的任务中，就能更快速直接全面地获取所需要的信息。本小节从搭配检索、信息抽取两个方面来说明汉语增强依存句法规范在语料库检索中的实际应用。

4.1 搭配抽取中的应用

搭配通常是指两个或两个以上的词语所组成的一种语言表示，这种表示往往是某种语言习惯的表达(邵艳秋et al., 2019)。通过在语料库中抽取搭配，一方面便于汉语学习者检索自己所用搭配是否准确、常用，有利于学习者自学；另一方面也便于对外汉语教师和研究人员建立搭配库，通过检索某个词的常用搭配及其例句方便教学和语言本体的研究。此外，搭配也能支持自动翻译、信息检索、自动问答等应用研究。

依靠人工判断搭配费时费力，不仅主观性强，而且耗时巨大。随着计算机技术的发展，搭配抽取技术也有了长足的进步。目前，一种比较好的方法是基于依存句法分析的搭配自动抽取。通过依存弧来抽取搭配时，需要明确依存关系表示的搭配关系。例如，规定 nsubj 表示主谓搭配关系，dobj 表示动宾搭配关系，advmop:dvp 表示状中搭配关系，compound:nn 表示定中搭配关系，那么在图21所示句子中，通过基础依存句法抽取到的搭配如表1 所示。

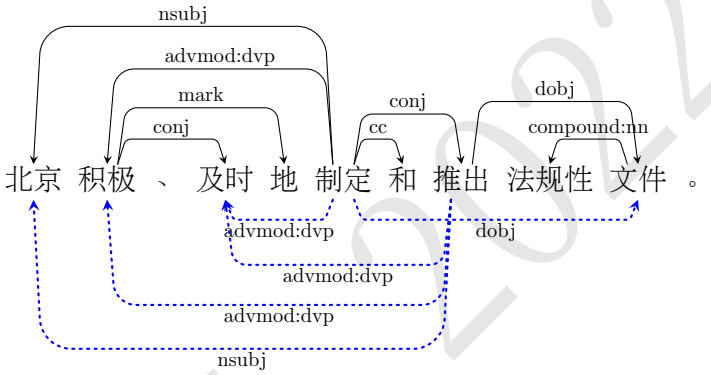

图 21: 依存句法标注示例

搭配类型	抽取结果
主谓搭配	北京制定
动宾搭配	推出文件
状中搭配	积极制定
定中搭配	法规性文件

表 1: 利用基础依存抽取到的搭配

但若对抽取到的搭配进行人工校对就会发现，由于并列情况的存在，通过基础依存句法只能抽取到并列项其中之一的搭配关系，而忽略了其他并列项的搭配。增强依存句法就能很好地解决这个问题，它把并列项之间的依存关系都通过添加依存弧的方式展现出来，用增强依存句法来抽取搭配就能找回那些被遗漏的搭配。这种全面的搭配抽取方式一方面能帮助学习者在用例句学习搭配时找到句中所有搭配，明确可使用的搭配；另一方面，可以扩大搭配库，便于后续的统计处理和研究工作，即使原始语料库较小，也能抽取出更多的搭配范式。如图21例句中，利用增强依存句法还能抽取出的搭配如表2。

搭配类型	抽取结果
主谓搭配	北京推出
动宾搭配	制定文件
状中搭配	及时制定 积极推出 及时推出

表 2: 利用依存句法重现的遗漏搭配

4.2 信息抽取中的应用

信息抽取的主要功能是从非结构化的文本中自动提取用户感兴趣的结构化事件信息，是各项自然语言处理任务例如知识图谱构建、翻译、篇章理解等应用的基石(项威and 王邦, 2020)。目前，信息抽取主要包括以下命名实体识别、指代消解、关系抽取以及事件抽取等几个方面的研究(张素香, 2007)。其中，比较常见的一种方法是利用依存句法来抽取信息。

用基础依存句法在检索平台中进行信息抽取时，如果句中存在大量信息嵌套和成分共享、抽取的信息之间有多层依存弧或存在句式转换的现象时，就需要对不同的情况建立多种抽取模式，甚至可能存在信息漏抽或抽取错误的情况。

例如“小王今年25岁，来自北京。”一句中，由于第二小句缺省主语，直接运用依存句法抽取主谓宾不能抽取到“小王来自北京”这一信息，必须对依存句法树进行一定的处理才能得到。运用增强依存句法之后，就能直接得到这些实词之间的语义关系，在信息抽取中无需花费大量的时间和精力处理句法树，这在句中存在并列结构、成分省略和转换句式时尤为明显。

5 总结

本文基于汉语基础依存句法制定了增强依存句法规范，使得句中尽可能多的实词间的语义关系更加清晰明确。此外，本文还提供了汉语增强依存句法转换的 Python 转换器以及方便进行可视化比较的 Web Demo，并给出了该规范在搭配抽取和信息抽取中的实际应用，以说明该规范在这些任务中的优势。

未来，还应进一步完善和补充汉语增强依存句法体系，以满足规模更大、句子更长、结构更复杂的语料。目前增强依存句法规范在汉语特殊句式中只考虑到了比较常见的一部分，之后还需要将判断句、倒装句等句式纳入到增强依存句法体系中来。此外，面对不确定的情况的处理，也可以更好地进行分类讨论，例如当复句中省略宾语时，如该谓语动词为不及物动词，那么不添加该谓语动词与其他小句成分间表示宾语的依存弧，因此，就需要对谓语动词进行及物和不及物的分类处理。最后，还应进一步探索其应用场景，找到更多适合其发挥的任务，挖掘其更大的优势。

参考文献

项威and 王邦. 2020. 中文事件抽取研究综述. 计算机技术与发展, 2(20):1–6.

刘月华, 潘文娱, and 故韡 . 2001. 实用现代汉语语法. 商务印书馆.

邵艳秋, 申资卓, and 刘世军. 2019. 基于依存搭配抽取技术的平面媒体语言监测研究. 山西大学学报:自然科学版, 3(42):526–533.

刘丹青. 2002. 汉语中的框式介词. 当代语言学, 4:241–253+316.

年玉萍. 2003. 谈谈兼语句. 延安教育学院学报, 1:40–42.

张素香. 2007. 信息抽取中关键技术的研究. Ph.D. thesis, 北京邮电大学.

李正华. 2013. 汉语依存句法分析关键技术研究. Ph.D. thesis, 哈尔滨工业大学.

Marie Candito, Bruno Guillaume, Guy Perrier, and DjaméSeddah. 2017. Enhanced ud dependencies with neutralized diathesis alternation. In *Proceedings of the Depling 2017-Fourth International Conference on Dependency Linguistics*.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. Stanford typed dependencies manual. Technical report, Stanford University.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan T. McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*.

Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster, and Maria Simi. 2018. Enhancing universal dependency treebanks: A case study. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajic, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis M. Tyers, and Daniel Zeman. 2020. Universal dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'20)*.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*.

Aryeh Tiktinsky, Yoav Goldberg, and Reut Tsarfaty. 2020. pybart: Evidence-based syntactic transformations for ie. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.