

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

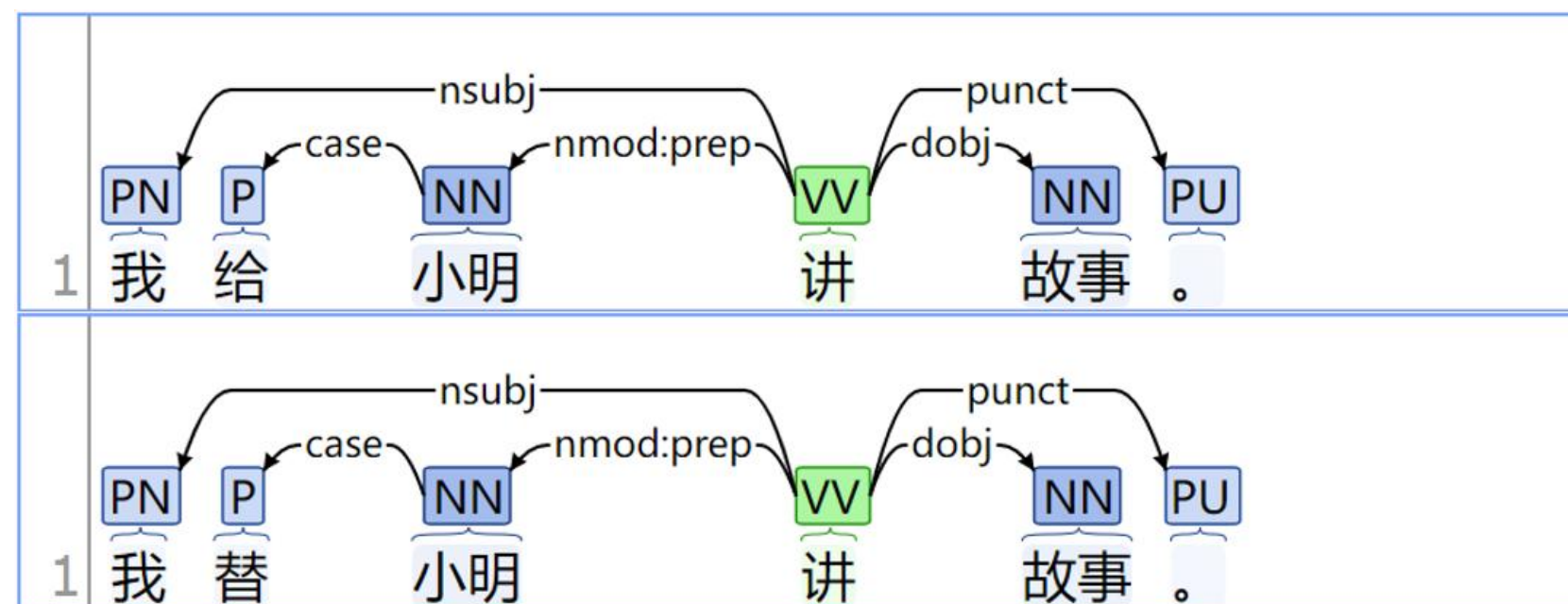
³北京语言大学 语言资源高精尖创新中心

⁴信阳学院 文学院

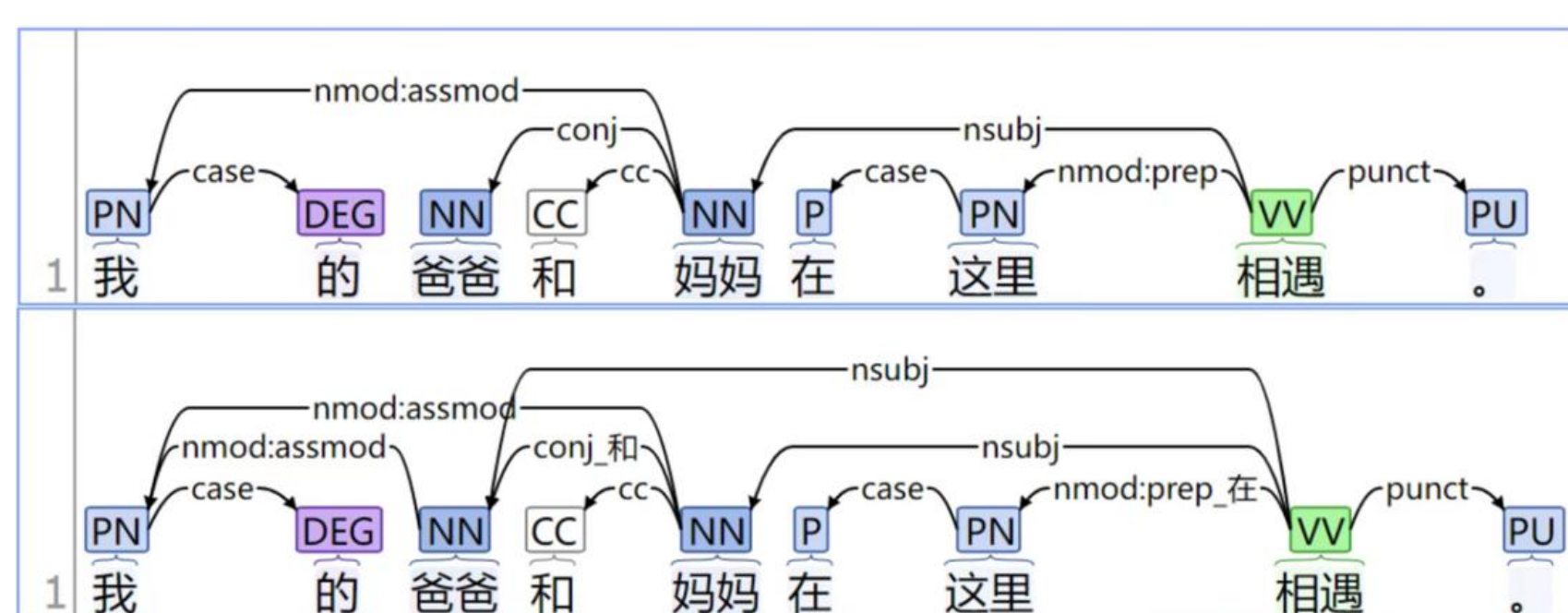


研究背景及意义

- 汉语依存句法表示了词与词之间的语法关系，但缺乏许多存在于句中实词之间的语法和语义关系。



- 现有的汉语依存的依存标签难以表达长距离的语义关系。

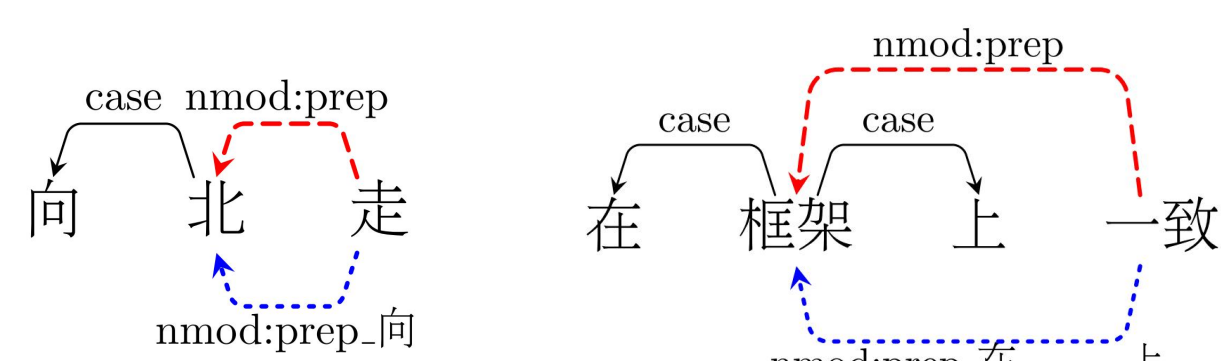


- 前人在依存句法的基础上提出了增强依存句法，来满足依存句法反映语义信息的需求。
- 我们基于斯坦福依存句法，在借鉴英文增强依存句法思想的基础上，制定了汉语增强依存句法规范。
- 该规范通过修改依存标签、添加弧或节点的方式，将依存句法树转换为可以表示更多信息的依存句法图，显性地展示实词之间的语义关系，更有利于自然语言处理下游任务的应用。

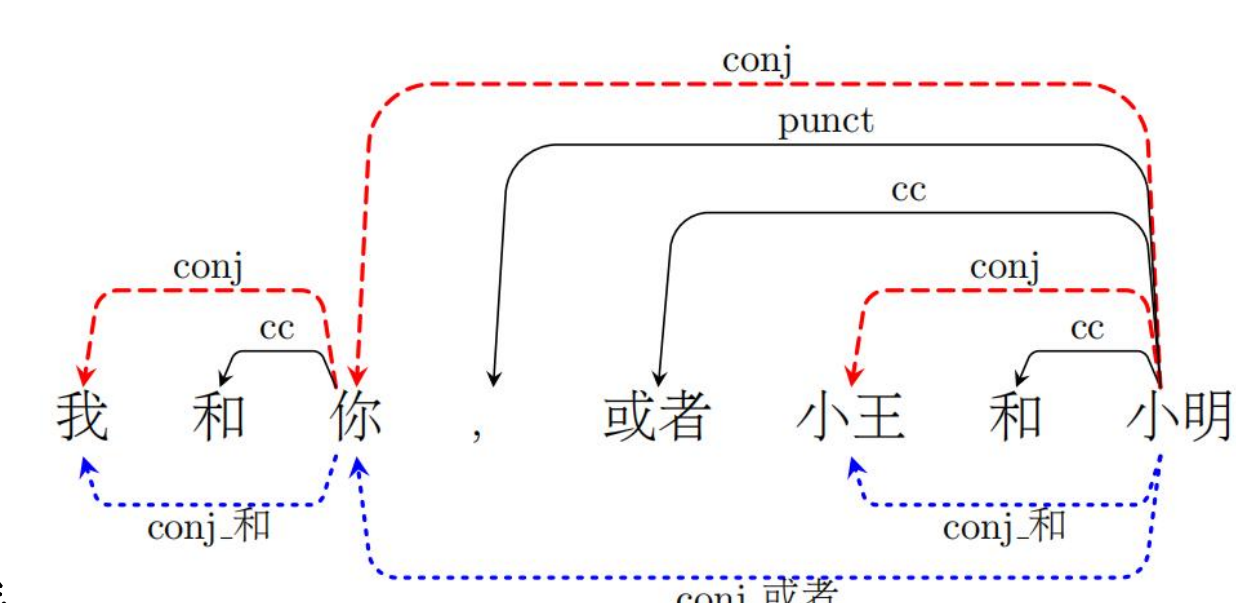
增强的依存句法规范

- 介词和连词的增强

- 介词的增强** 在依存弧上添加具体介词，这些依存弧连接了介词短语的实词和被修饰词，原标签与添加的介词中间用“_”连接。

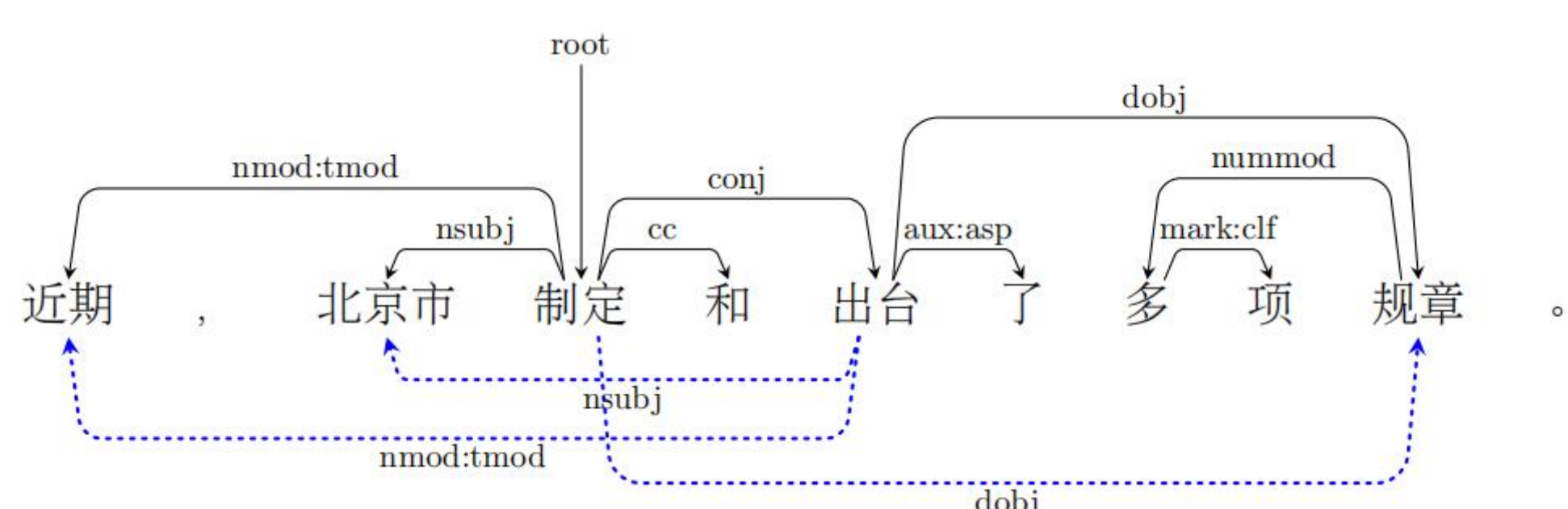


- 连词的增强** 通过在依存标签conj上添加依存弧cc所指的并列连词，使得并列项之间的语义关系更加明晰。



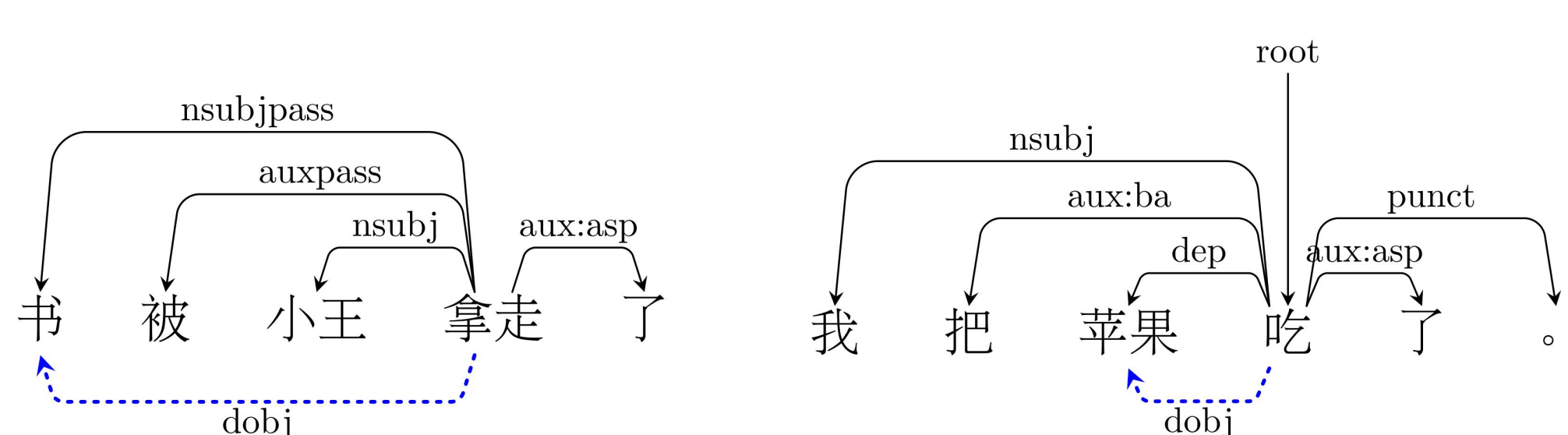
- 并列结构的传播

- 并列结构的传播** 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表示出来。

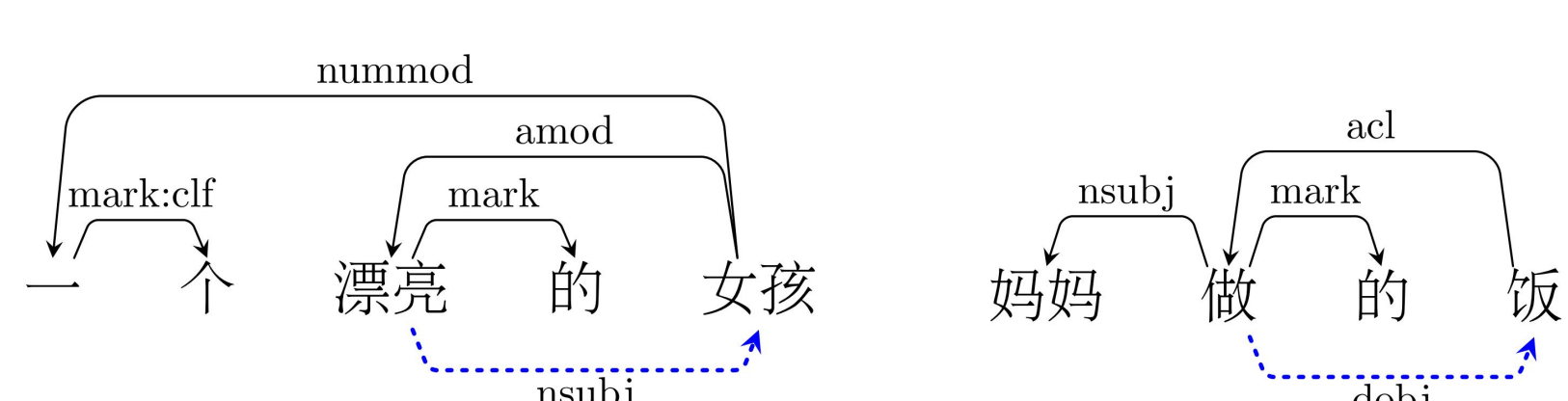


- 句式转换

- 被动句/“把”字句的转换** 为了统一被动句与主动句中实词间的依存关系，我们以更为常用的主动句依存关系为标准，将被动主语与谓语的关系标为dobj。同时，将“把”引导的宾语还原到动词宾语的位置。

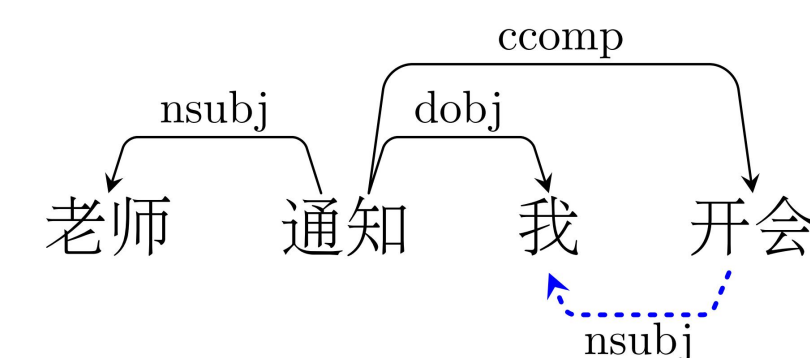


- 形容词/动词短语修饰语的转换** 将偏正短语转换为以中心语为主语、以形容词性修饰语为谓语的主谓短语。当一个动词修饰名词中心语时，如该动词不含宾语，那么中心语可能为该动词的受事。



- 特殊句式的增强

- 兼语句** 受限一个节点只能有一条入弧的规则，基础依存句法只标注了前一谓语与兼语之间的宾语关系。在增强依存句法中，增添一条依存弧由后一谓词指向兼语，依存标签为nsubj。



- 省略句** 当动词性成分作状语时，状语中的动词和主句的谓语有时共用一个主语或宾语，由于汉语的经济原则，就可能省略相应主语或宾语，因此我们需要还原省略部分。

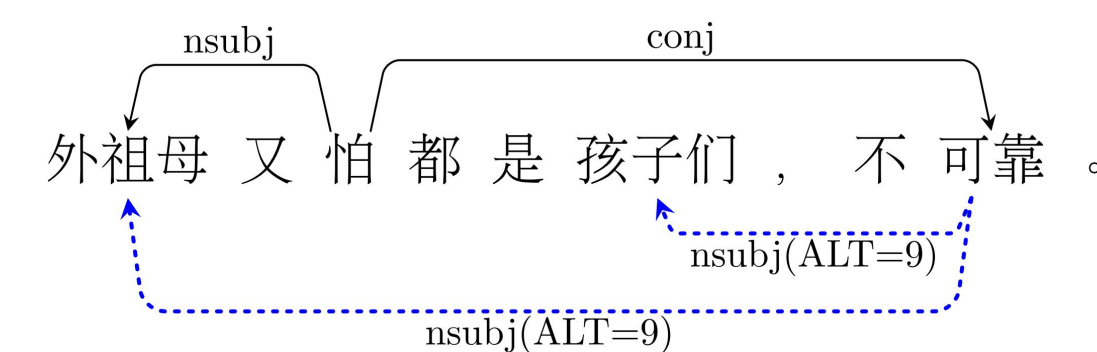


- 不确定情况的处理

- 在上述的规则中，也会产生一些不适用的情况，我们借用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。



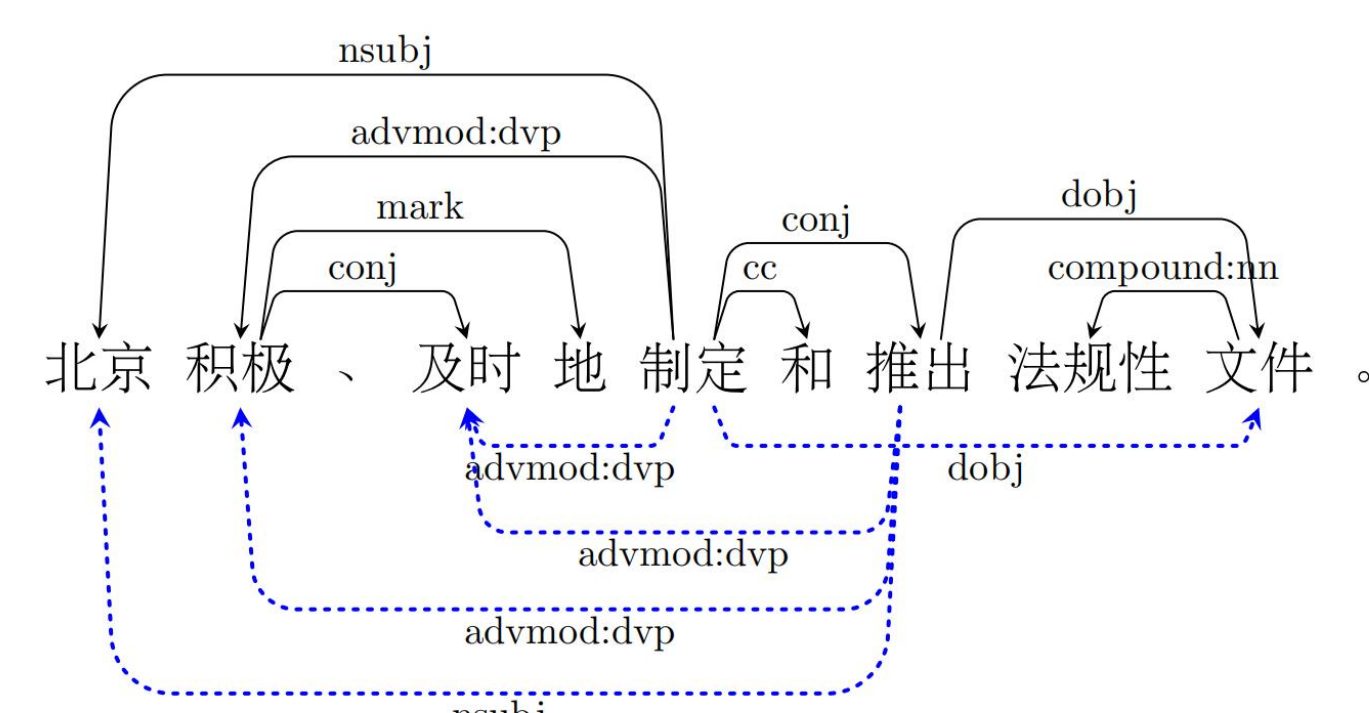
- 此外，汉语当中也存在复句中的某个小句省略了主语，但其省略的主语不是其他小句主语的情况，我们采取Aryeh (2020) 提出的概念ALT=X，表示用户可以选择其一，其中X表示被省略主语或宾语的词在句子中的位置。



增强依存句法规范的实际应用

- 搭配抽取中的应用

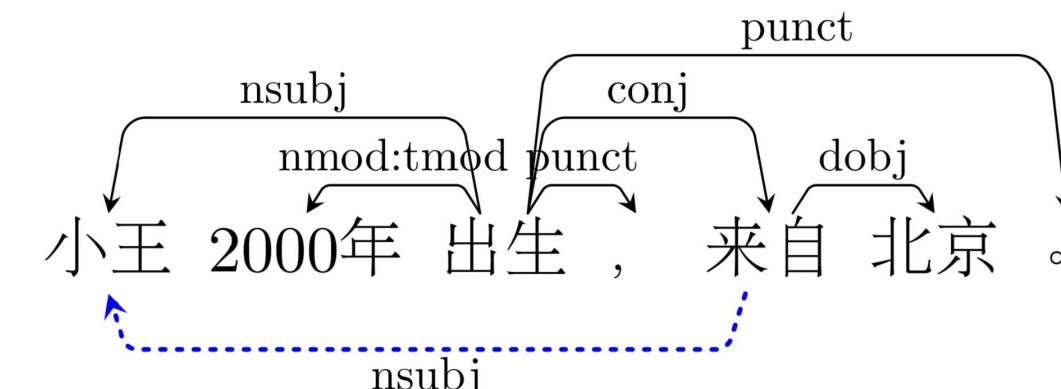
- 增强依存句法能很好地通过添加依存弧的方式将并列项之间的依存关系展现出来，找回被遗漏的搭配。
- 这种全面的抽取方式一方面能帮助学习者明确例句中可使用的搭配；另一方面，在原始语料库较小的情况下，可以扩大搭配库，抽取更多的搭配范式，便于后续的处理和研究工作。



搭配结构	基础/增强句法依存抽取结果
主谓搭配	北京制定 北京推出
动宾搭配	推出文件 制定文件
状中搭配	积极制定 及时制定 积极推出 及时推出
定中搭配	法规性文件

- 信息抽取中的应用

- 用基础依存句法在检索平台中进行信息抽取时，可能导致信息漏抽或抽取错误，如下图句子中无法获得“小王来自北京”这一信息，因此需要对不同的情况建立多种抽取模式。
- 增强依存句法能直接得到这些实词之间的语义关系，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。



总结

- 我们基于汉语基础依存句法制定了增强依存句法规范，使得句中实词间的语义关系更加清晰明确。
- 我们还提供了汉语增强依存句法转换工具包。



- Demo: <https://parser.blcuicall.org/dep/>

- GitHub: <https://github.com/blcuicall/chinese-enhanced-dependencies>