Stanford NLP分享

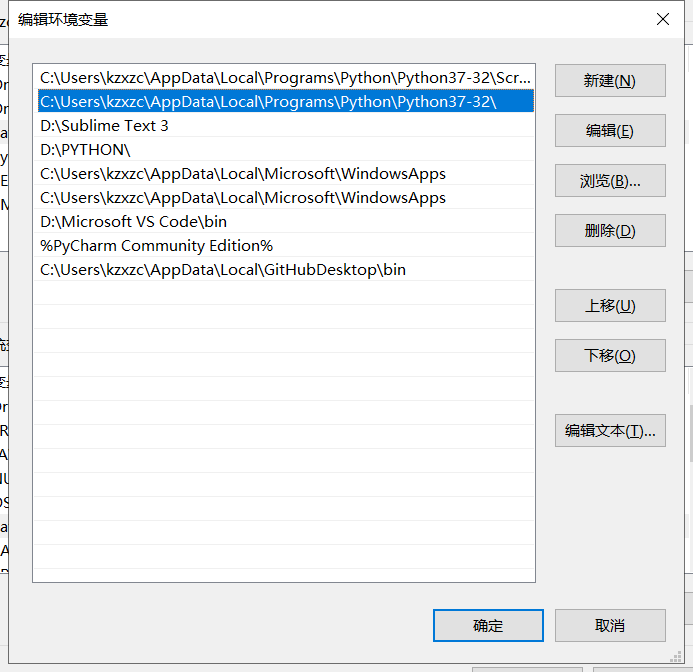
一、准备工作：

1. 下载Stanford CoreNLP， Chinese语言包

<https://stanfordnlp.github.io/CoreNLP/>

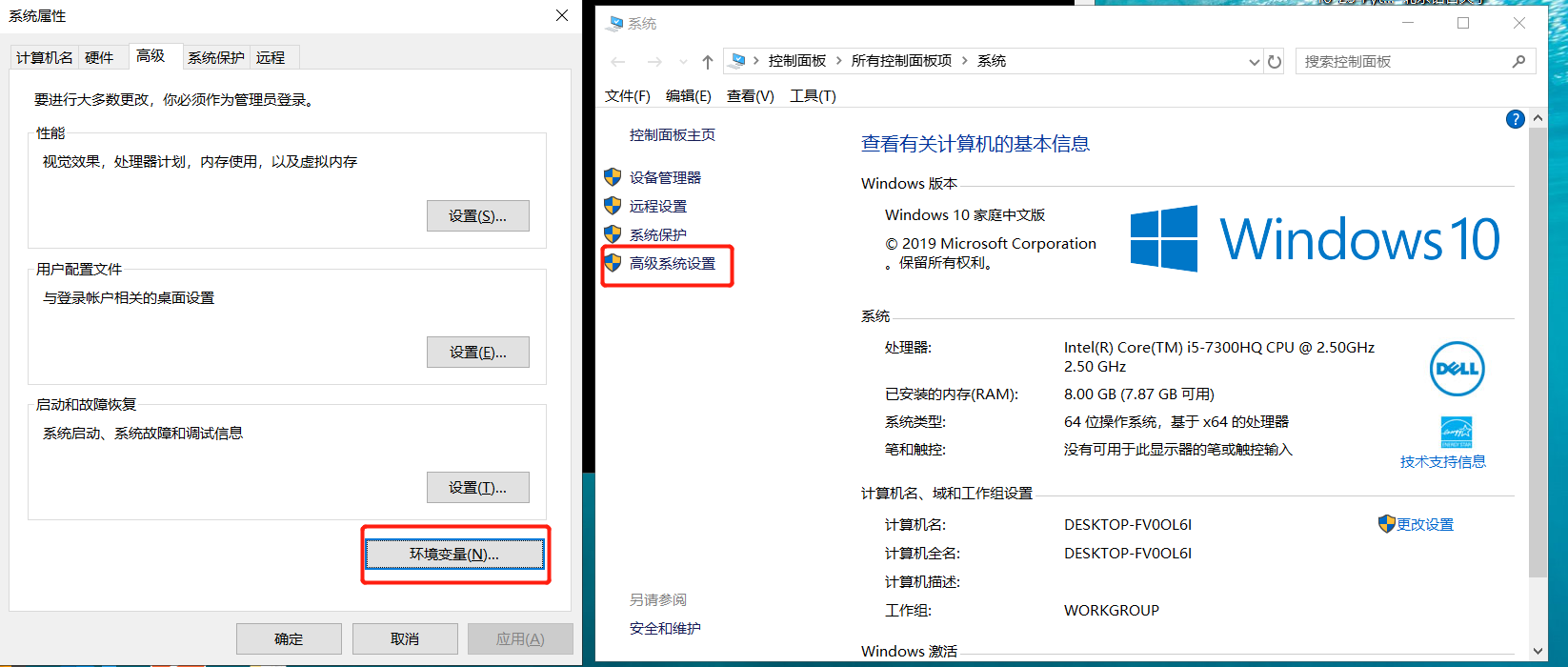
2. 下载安装jdk1.8

3.环境配置



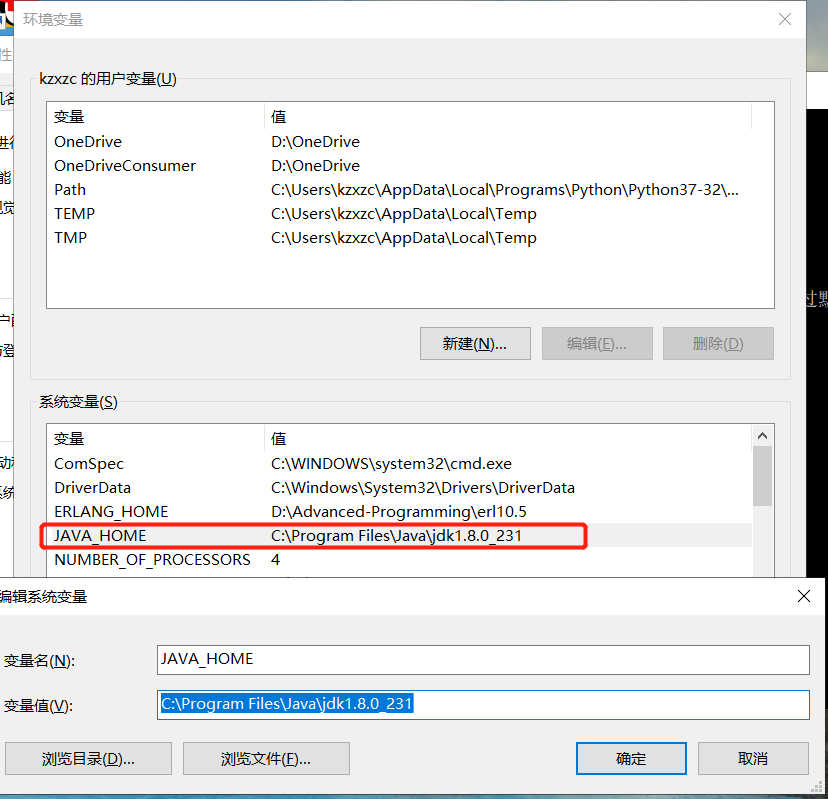
安装python时有 “Add Python to PATH” 选项

我的电脑——属性——环境变量



在系统变量下面配置

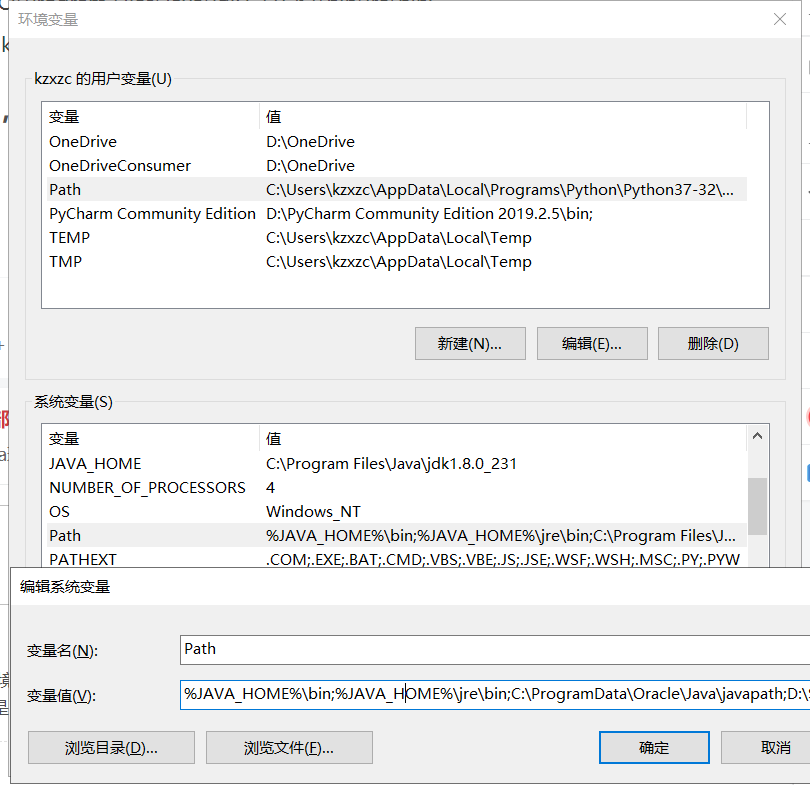
1. JAVA\_HOME:你自己的jdk的路径



1. 在Path中加入变量值%JAVA\_HOME%\bin;%JAVA\_HOME%\jre\bin;

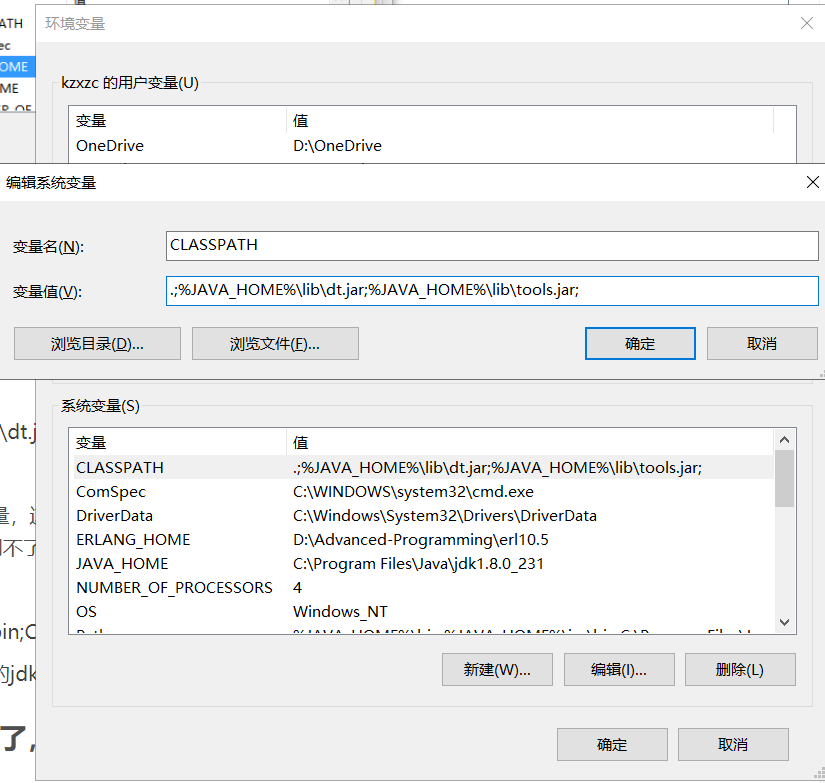
系统变量Path告诉操作系统可执行文件所在的路径

多个变量之间要用分号隔开

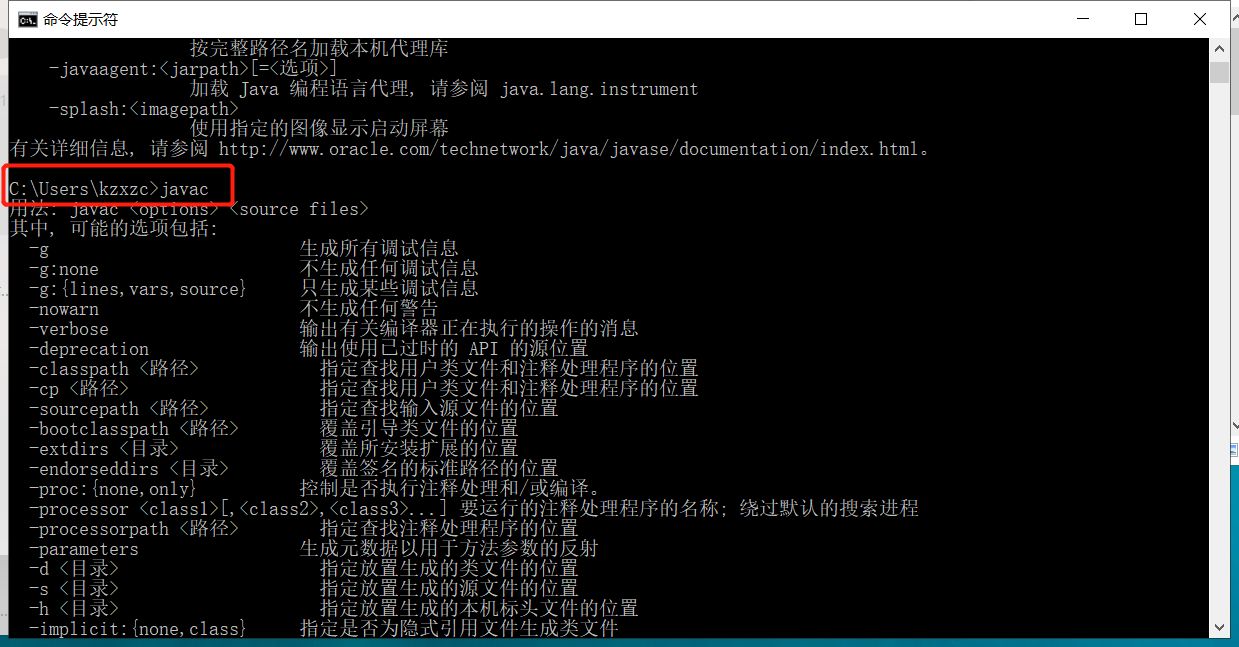
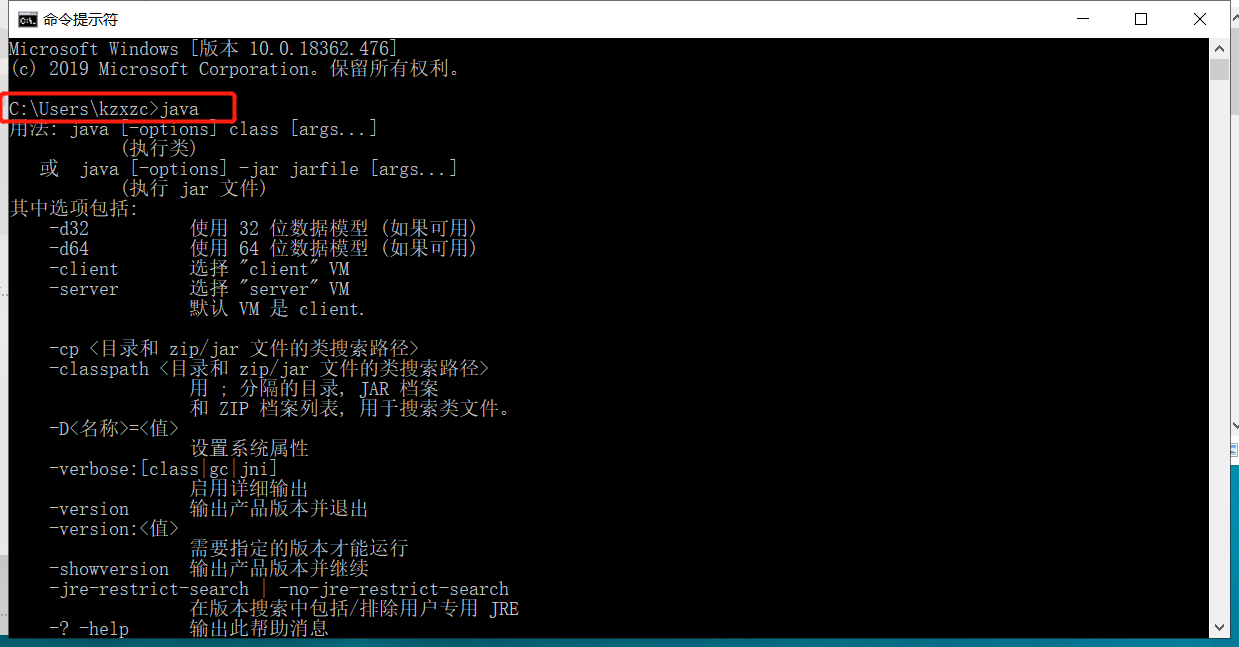
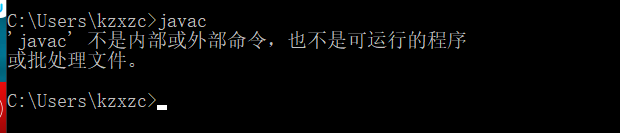


1. CLASSPATH= .;%JAVA\_HOME%\lib\dt.jar;%JAVA\_HOME%\lib\tools.jar

编译、运行Java程序时，JRE会去该变量指定的路径中搜索所需的类（.class）文件。不必须



1. Cmd中运行java和javac



Standfordnlp应用：

from stanfordcorenlp import StanfordCoreNLP

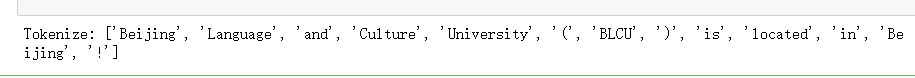
from nltk.tree import Tree

nlp=StanfordCoreNLP(r'D:\Advanced-Programming\stanford-corenlp-master\stanfordnlp')

sentence = 'Beijing Language and Culture University(BLCU) is located in Beijing!'

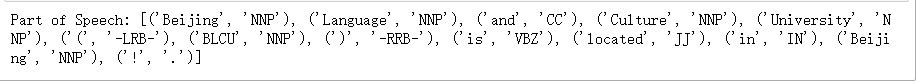
**功能一：分词**

print ('Tokenize:', nlp.word\_tokenize(sentence))



**功能二：词性标注**

print ('Part of Speech:', nlp.pos\_tag(sentence))



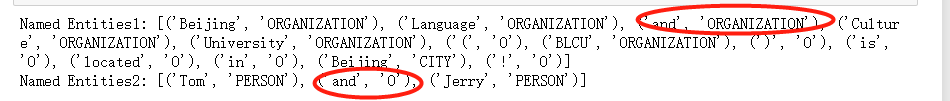
**功能三：实体识别**

sentence1 = 'Beijing Language and Culture University(BLCU) is located in Beijing!'

sentence2 = 'Tom and Jerry'

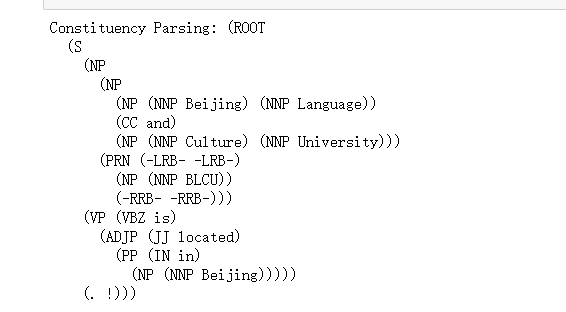
print ('Named Entities1:', nlp.ner(sentence))

print ('Named Entities2:', nlp.ner(sentence2))

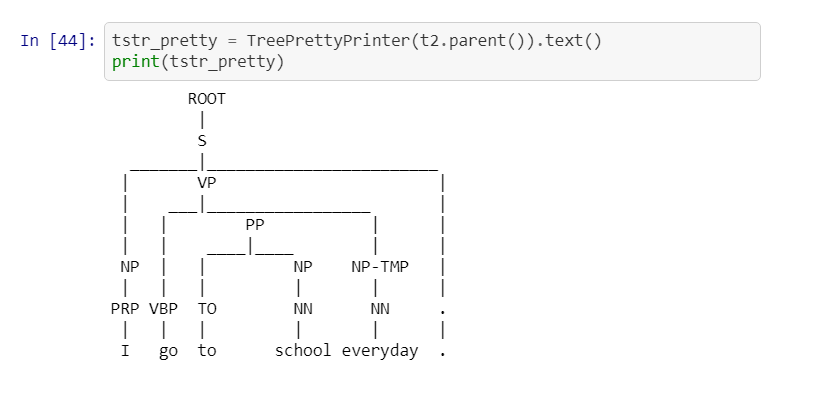


**功能四：句法分析**

print ('Constituency Parsing:', nlp.parse(sentence))



使用nltk进行可视化



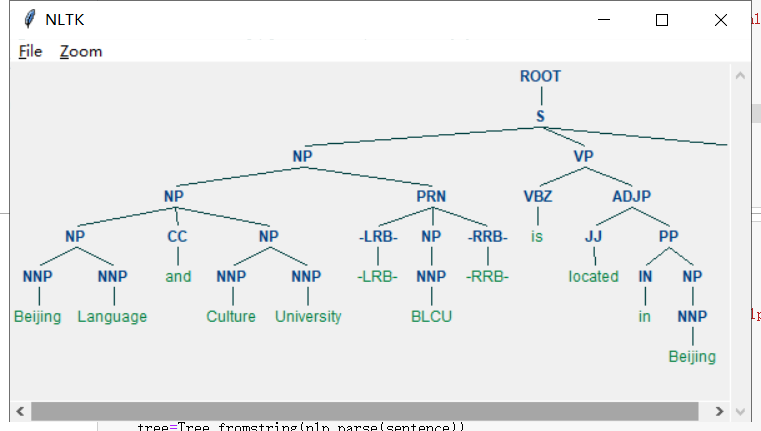
上图为老师上课分析的可视化方式

也可以：👇

from nltk.tree import Tree

tree=Tree.fromstring(nlp.parse(sentence))

tree.draw()

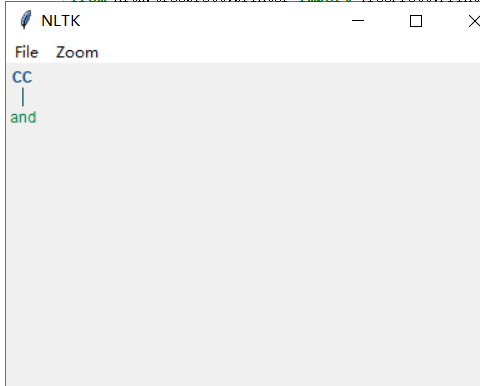


trees = [t for t in pt.subtrees()]

trees[4].draw()



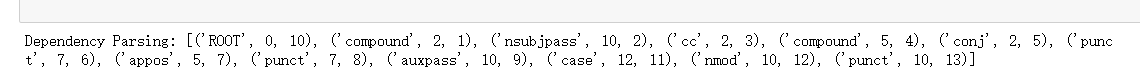
trees[7].draw()



**功能五：依存句法分析**

sentence = 'Beijing Language and Culture University(BLCU) is located in Beijing!'

print ('Dependency Parsing:', nlp.dependency\_parse(sentence))



ROOT：要处理文本的语句

IP：简单从句

NP：名词短语

VP：动词短语

PU：断句符，通常是句号、问号、感叹号等标点符号

LCP：方位词短语

PP：介词短语

CP：由‘的’构成的表示修饰性关系的短语

DNP：由‘的’构成的表示所属关系的短语

ADVP：副词短语

ADJP：形容词短语

DP：限定词短语

QP：量词短语

NN：常用名词

NR：固有名词

NT：时间名词

PN：代词

VV：动词

VC：是

CC：表示连词

VE：有

VA：表语形容词

AS：内容标记（如：了）

VRD：动补复合词

CD: 表示基数词

DT: determiner 表示限定词

EX: existential there 存在句

FW: foreign word 外来词

IN: preposition or conjunction, subordinating 介词或从属连词

JJ: adjective or numeral, ordinal 形容词或序数词

JJR: adjective, comparative 形容词比较级

JJS: adjective, superlative 形容词最高级

LS: list item marker 列表标识

MD: modal auxiliary 情态助动词

PDT: pre-determiner 前位限定词

POS: genitive marker 所有格标记

PRP: pronoun, personal 人称代词

RB: adverb 副词

RBR: adverb, comparative 副词比较级

……

<https://blog.csdn.net/l919898756/article/details/81670228>

再试试中文版：

sentence = '欢迎来到宇宙中心五道口旁的北京语言大学！'

with StanfordCoreNLP(r'D:\Advanced-Programming\stanford-corenlp-

master\stanfordnlp', lang='zh') as nlp:

print('一：',nlp.word\_tokenize(sentence))

print('二：',nlp.pos\_tag(sentence))

print('三：',nlp.ner(sentence))

print('四：',nlp.parse(sentence))

print('五：',nlp.dependency\_parse(sentence))

tree=Tree.fromstring(nlp.parse(sentence))

tree.draw()

