The Use of the Stanford Lexicalized Parser

Zhao Jiefu

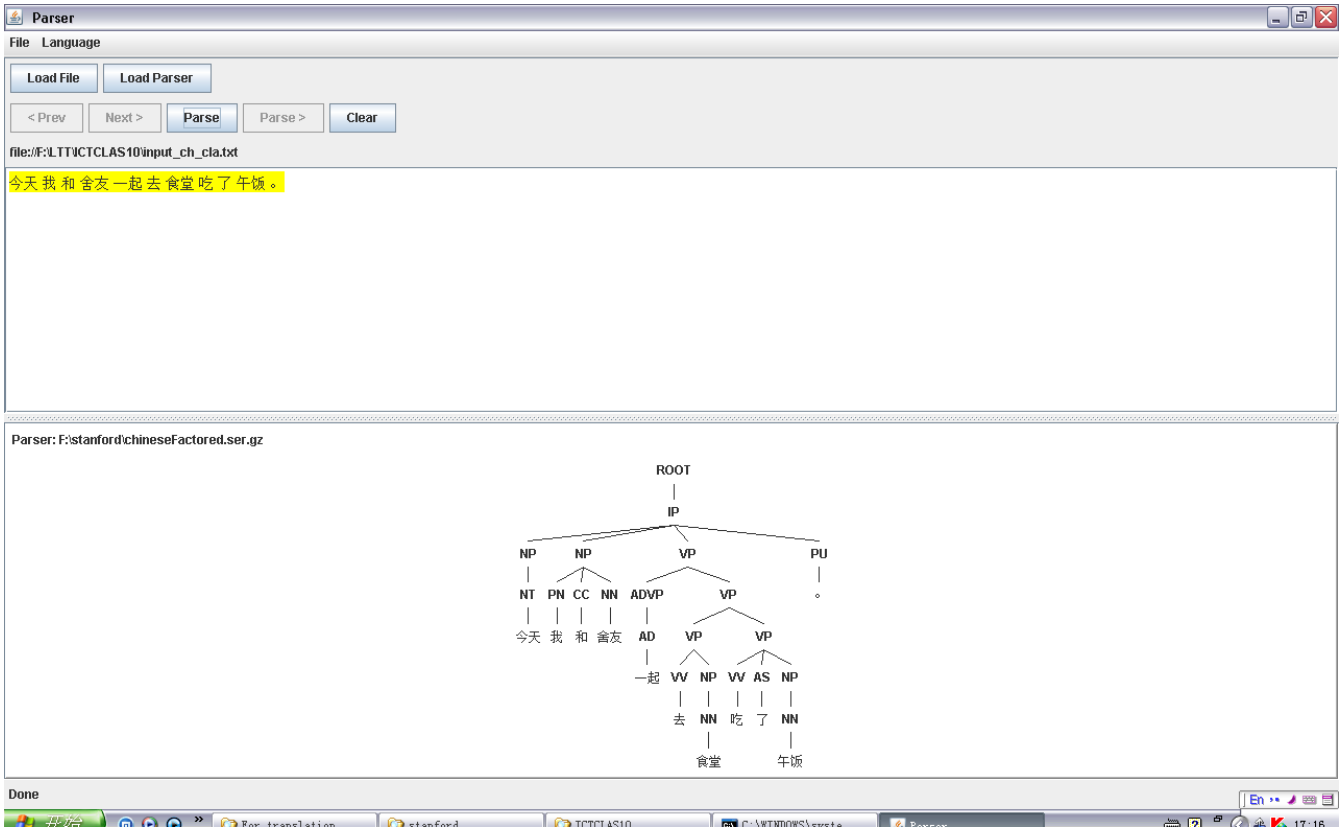
The Stanford Parser can only process plane texts. It can parse English, Chinese and German.

Parsers for English: englishPCFG.ser.gz, englishfactored.ser.gz;
Parsers for Chinese: chinesefactored.ser.gz, Xinhuafactored.ser.gz and XinhuaPCFG.ser.gz.

For the differences between the PCFG and factored parsers, the readme document says: "This package contains 3 parsers: a high accuracy unlexicalized PCFG, a lexical dependency parser, and a factored model, where the estimates of dependencies and an unlexicalized PCFG are jointly optimized to give a lexicalized PCFG treebank parser. Also included are grammars for various languages for use with these parsers." "The parser is supplied with several grammars. There are English grammars both based on the standard LDC Penn Treebank WSJ secs 2-21, and ones based on a slightly augmented data set. There are Chinese grammars trained just on mainland material from Xinhua and more mixed material from the LDC Chinese Treebank, and there is a German parser trained from the Negra corpus. We also provide several test sentences."

My experiment:

For parsing the Chinese sentence "今天我和舍友一起去食堂吃了午饭", you should first segment the sentence manually or by ICTCLAS and get something like "今天 我 和 舍友 一起 去 食堂 吃了 午饭":. (Remember: ICTCLAS can only process texts in ANSI format while Stanford Parser can only deal with UTP-8 formatted Chinese, so you need to re-encode the texts when you have finished segmentation). Then double-click lexparser-gui.bat to open the user interface. Load the Chinese parser (chinesefactored.ser.gz), choose "tokenized simplified Chinese" in Language, and then load the text you want to parse. The process of loading the Chinese parser and parsing this Chinese sentence is much, much slower than an English sentence. Here is the result:


Parser
File Language
Load File
Load Parser
< Prev
Next >
Parse
Parse >
Clear
file://F:\LTT\CTCLAS10\input_ch_cla.txt
今天 我 和 舍友 一起 去 食堂 吃 了 午饭 。
Parser: F:\stanford\chineseFactored.ser.gz
ROOT
IP
NP
NP
VP
PU
NT
PN
CC
NN
ADVP
VP
。
今天
我
和
舍友
AD
VP
VP
一起
VV
NP
VV
AS
NP
去
NN
吃
了
NN
食堂
午饭
Done