Middle East Technical University
Department of Computer Engineering

CENG 463

Introduction to Natural Language Processing

Spring 2010

Homework 2

Due: 30 May 2010 23:55:50

Introduction

In this assignment you are expected to construct a feature-based phrase structure grammar for Turkish language and implement a parser based on that grammar. Your parser is expected to handle

- The basic word order variation in Turkish.
- Subject verb agreement.
- Basic subject relativisation such as "kitap okuyan çocuk".

You will use the following set as part of speech tag set example and write your grammar rules according to that.

Name	POS Tag	Examples
Nouns	N	sincap, köpek
Determiners	Det	bu, şu, bir
Nominal Relativisers	NomRel	minderdeki
Adjectives	Adj	mavi, küçük
Subject Relativisers	VerbRel	okuyan, uyuyan
Verbs	V	uyudu, kaçtı

You will use inflected words as lexical entities, however, remember that you will implement plural and person agreement, thus you must include the agreement feature values of the words in their feature structures.

Specifications

- You will use Python programming language and NLTK.
- You will implement your parser as a class that inherits *nltk.parse.api.ParserI* class [1] and you will define *nbest_parse()* method.
- You will write your grammar in a file, and use it in your parser by using *load_parser()* method of NLTK [2].
- You will choose 10 nouns, 5 determiners, 10 nominal relativisers, 10 adjectives, 10 subject relativisers and 10 verbs as your lexical entities. Furthermore, you must make sure that your lexicon could generate sentences that your grammar would reject.
- Your parser will return several parses for ambiguous constructions, however it should not return more than one parse for non-ambiguous sentences, in other words it should not *overgenerate*.
- Your grammar should be able to handle sentences of the form:

  Köpeği kovalayan sincap kaçtı.

  Sincabın sevdiği peyniri yiyen fare beni kızdırdı.

  Minderdeki köpek avludaki kediyi kovaladı.

Minderdeki sincabı kovalayan köpek kaçtı.

Çocuk bana baktı.

Bana çocuk baktı.

and reject the sentences of the form:

* Koltuktaki köpek kaçtım.

* Kediyi kovalayan adam uyudular.

* Kitabı uyuyan adam delirdi.

Please note that the last negative example does not parse because uyumak is intransitive.

Submission

- Submit a ***.tar.gz*** file, through COW, containing:
  - Your source code as a single ***.py*** file,
  - Your grammar file as a ***.fcfg*** file,
  - Your lexicon in a ***.txt*** file in the following format.

N:	noun1 ... noun10
Det:	det1 ... det5
NomRel:	nomrel1 ... nomrel10
Adj:	adj1 ... adj10
VerbRel:	verbrel1 ... verbrel10
V:	verb1 ... verb10

References

[1] ParserI class API documentation

[2] load_parser method documentation