11-712: NLP Lab Report

[Kartik Goyal –NAS]

April 26, 2013 [due date –NAS]

Abstract

[one paragraph here summarizing what the paper is about –NAS]

[brief introduction –NAS]

1 Basic Information about [Spanish –NAS]

Spanish is a Romance(Ibero romance group) language that originated in Castile region of Spain. Apart from the other Romance languages, it enriched it's vocabulary from Basque and Arabic. Spanish is closely related to other Iberian Languages like Italian and Portuguese with which, it has 82% and 89% lexical similarity respectively. Spanish is written in Latin script. The interrogative and exclamatory clauses are introduces with inverted question and exclamation marks.

It is the second most spoken language by number of native speakers and is one of the 6 official languages of the united nations. It is spoken by around 329 million people all over the world. Spanish is the primary language of 20 countries worldwide. Apart from Spain, it is official language of many Latin American regions(Argentina, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Uruguay, Venezuela, Peru,Puerto Rico). It is also, the official language of Equatorial Guinea in Africa.

There are some grammatical and lexical differences between the Spanish spoken in regions of Spain and the Spanish spoken in Latin America. The main grammatical variations between dialects of Spanish involve differing uses of second person pronouns. For eg. In most of Spain, 'ustedes' and 'vosotros' are used according to the degree of formality, but only 'ustedes' is used in Latin America. There are some important vocabulary differences too between the dialects.

Spanish has an S-V-O grammar and is a heavily inflected language. It has a two gender noun system and the inflections of nouns, adjectives and determiners are caused by the number and gender. There are about 50 conjugated forms per verb which are caused by tense, number, person, T-V distinctions(formal) for second person, mood, aspect and voice. Spanish has some irregular verbs with respect to the conjugation rules like 'sentir', 'requerir' etc. The modifying constituents tend to be placed after their head words. Also, usually the adjectives are placed after nouns.

Spanish is a morphologically rich language and this tool aims to carry out a resonably accurate morphological analysis of the language.

2 Past Work on the Morphology of [Spanish –NAS]

Tzoukerman and Liberman discus about a transducer which is essentially is a finite directed graph with edges labeleed by relations between surface and lexical forms. Santiago Rodriguez and Jesus Carretero describe the formal approach behind their spanish morphological analyzer, COES. Atserias et al. discus about the morphosyntactic analyzer which can be executed in GATE environment. Carmona et al. discus about MACO+, a tool for morphological analysis of spanish.

3   Available Resources

[include discussion of your corpora –NAS]

4   Survey of Phenomena in [Your Language –NAS]

5   Initial Design

6   System Analysis on Corpus A

7   Lessons Learned and Revised Design

8   System Analysis on Corpus B

9   Final Revisions

10   Future Work