11-712: NLP Lab Report

Mario Piergallini

April 26, 2013

Abstract

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

[brief introduction –NAS]

1 Basic Information about Caribbean Spanish

Spanish is a language belonging to the western branch of the Romance language family. While it only inflects nouns and adjectives for gender and number, it has a rich verbal inflectional system. Verbal suffixes can potentially account for person, number, tense, aspect and mood and has about fifty conjugated forms per verb.

Caribbean Spanish is a family of dialects primarily spoken in the Hispanophone islands of the Caribbean. Some parts of mainland Latin America can also be included in the definition, such as southern Florida, Panama, parts of Colombia and Venezuela. I will, however, primarily be concentrating on Puerto Rican (PR Spanish) and Dominican Spanish (Dom Spanish), which are fairly similar to each other and more divergent than some of the other varieties.

There are commonly thought to be three major areas in Spanish dialects: the Castillian area, covering central and northern Spain; the textitzona alteña, based in highland Latin America and covering most of Mexico, Central America and non-coastal South America; and the *zona bajeña* covering Andalusia in Spain, the Canary Islands and lowland Latin America, including Caribbean, Argentinian and Chilean Spanish.

As Latin American dialects, Puerto Rican and Dominican Spanish have certain features. Firstly, they lack a distinction between /s/ and /z/ (represented by ¡z¿ in Spanish orthography), a pattern commonly referred to as *seseo*. Likewise, they feature *yeísmo*, the lack of distinction between /j/ and /ll/ (represented by ¡y¿ and ¡ll¿ in Spanish orthography, respectively). More specifically, as *bajeña* dialects, they also have a tendency to reduce /s/ to an [h] sound, or delete it altogether when it appears in the coda of a syllable, as well as to velarize /n/ when it appears in a syllable coda.

Features that are more particular to these two dialects are frequent deletion of word final and intervocalic /d/ (so a word such as *todo* may be produced as [to:] or [to]) as well as neutralization of the distinction between /r/ and /l/ in syllable codas. In PR Spanish, it is generally that /r/ merges to /l/, while in Dominican Spanish they may both reduce to [j]. They both also have a tendency to shorten words, so that *¿Para adónde vas?* "Where are you going?" may be rendered in Dominican Spanish as *¿P'ónde va?* These phonological features combine to make much Spanish morphology more difficult to distinguish.

I will specifically be working with highly informal online written Spanish. This has a number of orthographical features that would need to be adjusted for as well. Silent ¡h¿ will often be left off, while graphemes like ¡y¿ and ¡ll¿ or ¡b¿ and ¡v¿ that represent no distinction in these dialects can be freely substituted. The end result is Spanish that can look like this:

manito no boi a desi k paso pero ya eso se ta resolviendo y a eso 3 lo tan ayudando lo mismo king [...] si eso no fuera verda ase mucho uviesen tao partio eso tiguere

Which corresponds to the following in standard Spanish orthography:

Hermanito no voy a decir que pasó pero ya eso se está resolviendo y a eso 3 lo están ayudando lo mismo [Latin] King. [...] si eso no fuera verdad hace mucho hubiesen estado partido eso tiguere.

2   Past Work on the Morphology of [Your Language –NAS]

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

References