1. Introduction

2. Literature Review

3. Introduction to Corpora

    3.1. List corpora, size, organization, etc.

    3.2. Give statistics for native and non-native instances

4. Overview of Tests and Classification

    4.1. Use of the Stanford Parser for parsing and dependencies

        4.1.1. Stanford Parser

        4.1.2. Stanford Typed Dependencies

    4.2. The Tests

        4.2.1. The project consists of a suite of tests

        4.2.2. Each test concerns a particular grammatical point

        4.2.3. Each test operates on the output from the Stanford parser and generates training or testing cases for the Weka classifier.

        4.2.4. The cases consist of one or (usually) more attributes with continuous values

        4.2.5. The values are generally relative frequencies (number of occurrences of that particular grammatical feature divided by the number of occurrences of all grammatical features considered in that test).

    4.3. Use of Weka for classification

        4.3.1. C4.5 Classifier

        4.3.2. RandomForest Classifier

    4.4. Use of Clojure

5. Classification Method 1 - Grammatical Relations

    5.1. Intro to method 1

    5.2. Detailed explanation of CS side

    5.3. Show results

    5.4. Attempt to identify the most informative features and explain these linguistically

6. Classification Method 2

7. Classification Method 3

8. ...

9. Comparative Discussion of classification methods

10. Future avenues of research, applications

11. Conclusion