HEBREW UNIVERSITY

REPORT

Language families from a computational perspective

Author:

Leshem CHOSHEN

For:

Dr. Eithan Grossman

Dmitry Nikolaev

A thesis submitted in fulfillment of the requirements

for the degree of Report in the

July 14, 2017

Hebrew University

Abstract

Faculty Of Humanities

Report

Language families from a computational perspective

by Leshem CHOSHEN

Grammatical Error Correction systems (henceforth, *correctors*) aim to correct ungrammatical text, while changing it as little as possible. However, whereas such conservatism is a virtue for correctors, we find that state-of-the-art systems make substantially less changes to the source sentences than needed. Analyzing the distribution of possible corrections for a given sentence, we show that this over-conservatism likely stems from the inability of a handful of reference corrections to account for the full variation of valid corrections for a given sentence. This results in undue penalization of valid corrections, thus disincentivizing correctors to make changes. We also show that simply increasing the number of references is unlikely to resolve this problem, and conclude by presenting an alternative reference-less approach based on semantic similarity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction