EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Philosophische Fakultät

Hizniye Isabella Boga

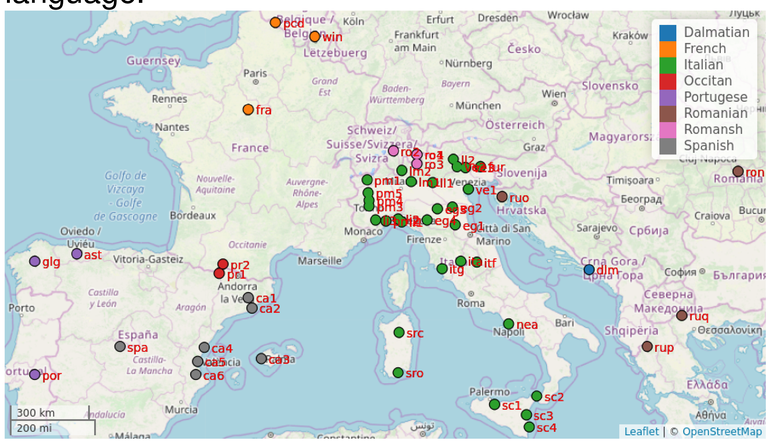
Distinguishing Between Close and Distant Romance Varieties

1 Introduction

How can we distinguish dialect pairs from language pairs? Some legitimate approaches include intelligibility measurements, perceptive distance, linguistic affiliation or the combination of some of the above.
Here, I apply computational distance and similarity measurements in order to objectively determine the affiliation, disregarding other linguistic parameters. I claim that there are not only very close and very distant groups but also groups with gradual differences and no clear affiliation.

2 Data

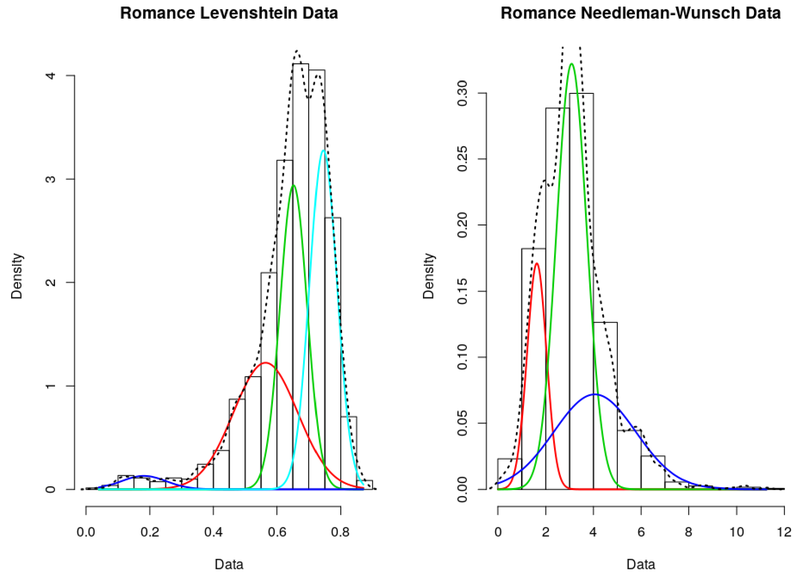
58 Romance Varieties, with 110 concepts for each language.



Universität Tübingen · Philosophische Fakultät · Seminar für Sprachwissenschaft · Wilhelmstraße 19 · 72074/Tübingen · Germany · Telefon +49 7071 29-75665 · http://sfs.uni-tuebingen.de/~iboga/ · isabella.boga@uni-tuebingen.de

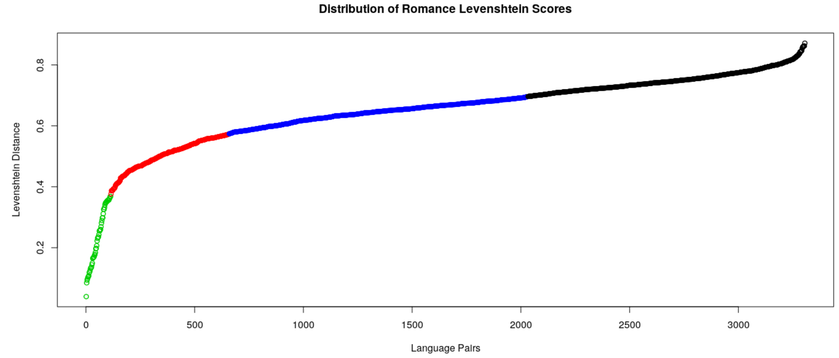
3 Methods

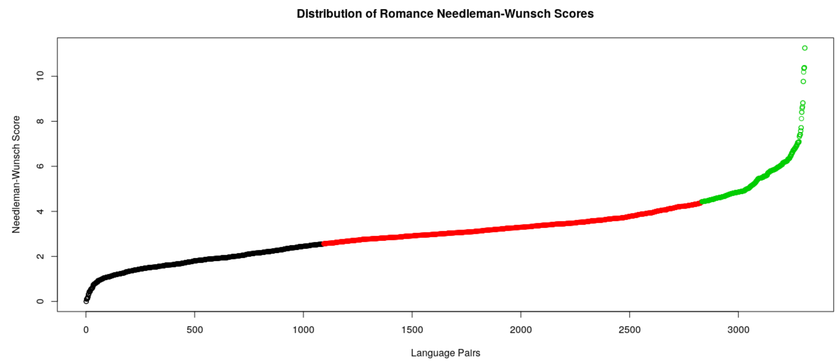
In order to determine the distances, I used the Levenshtein and the Needleman-Wunsch algorithms and compared the results of these two methods.
In addition to merely determining the distance/similarity, I added a weighted Scorer System based on Pointwise Mutual Information (PMI) scores to the Needleman-Wunsch measurement. The PMI distances served as further information on the relatedness of sound classes.



After the scores were calculated, I determined the number of subcomponents with mixture models. The resulting number of components can hence be used as the number of clusters for *k*-means clustering.

4 Results





Both methods yield a different number of clusters which can be analysed according to different underlying assumptions.