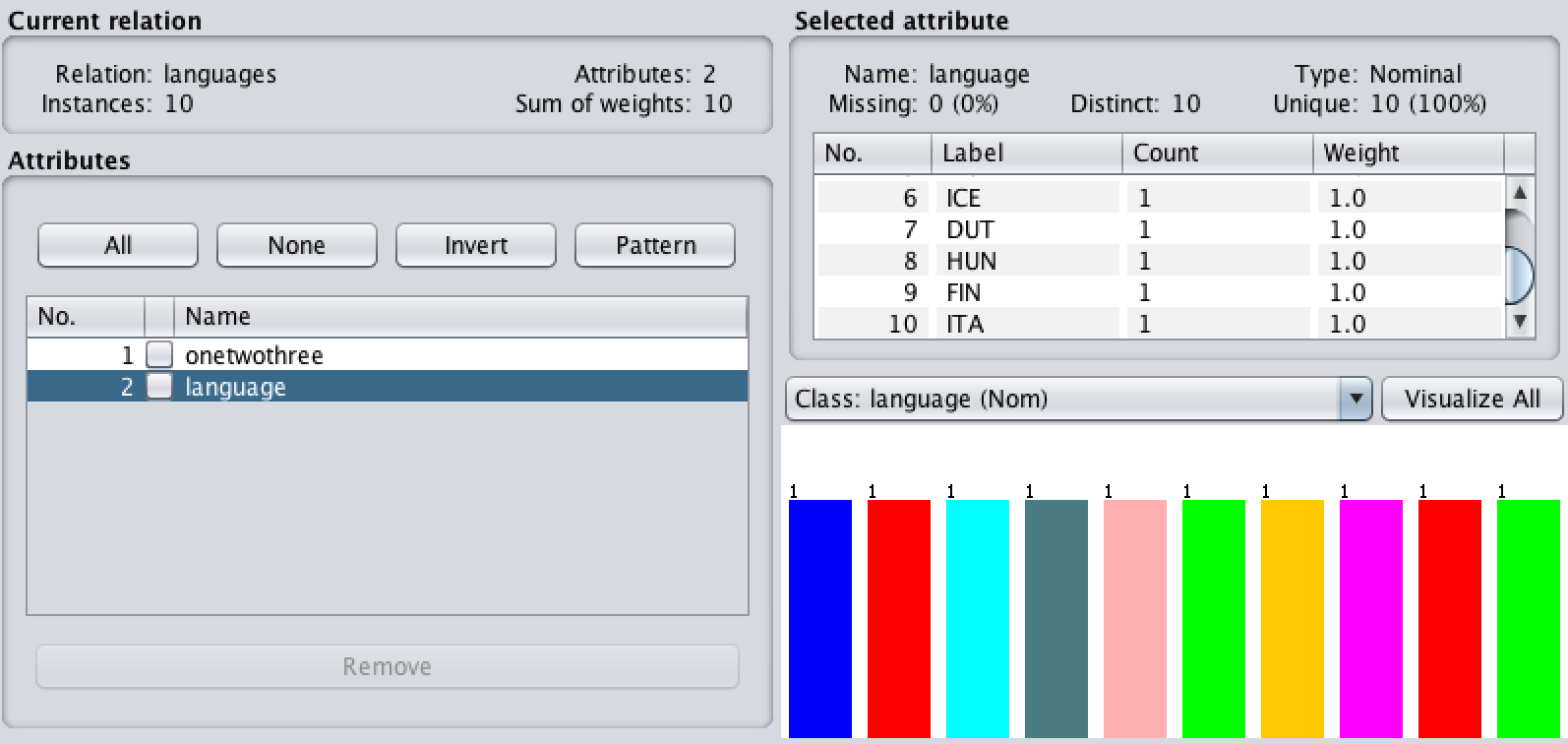
IS590DT assignment2

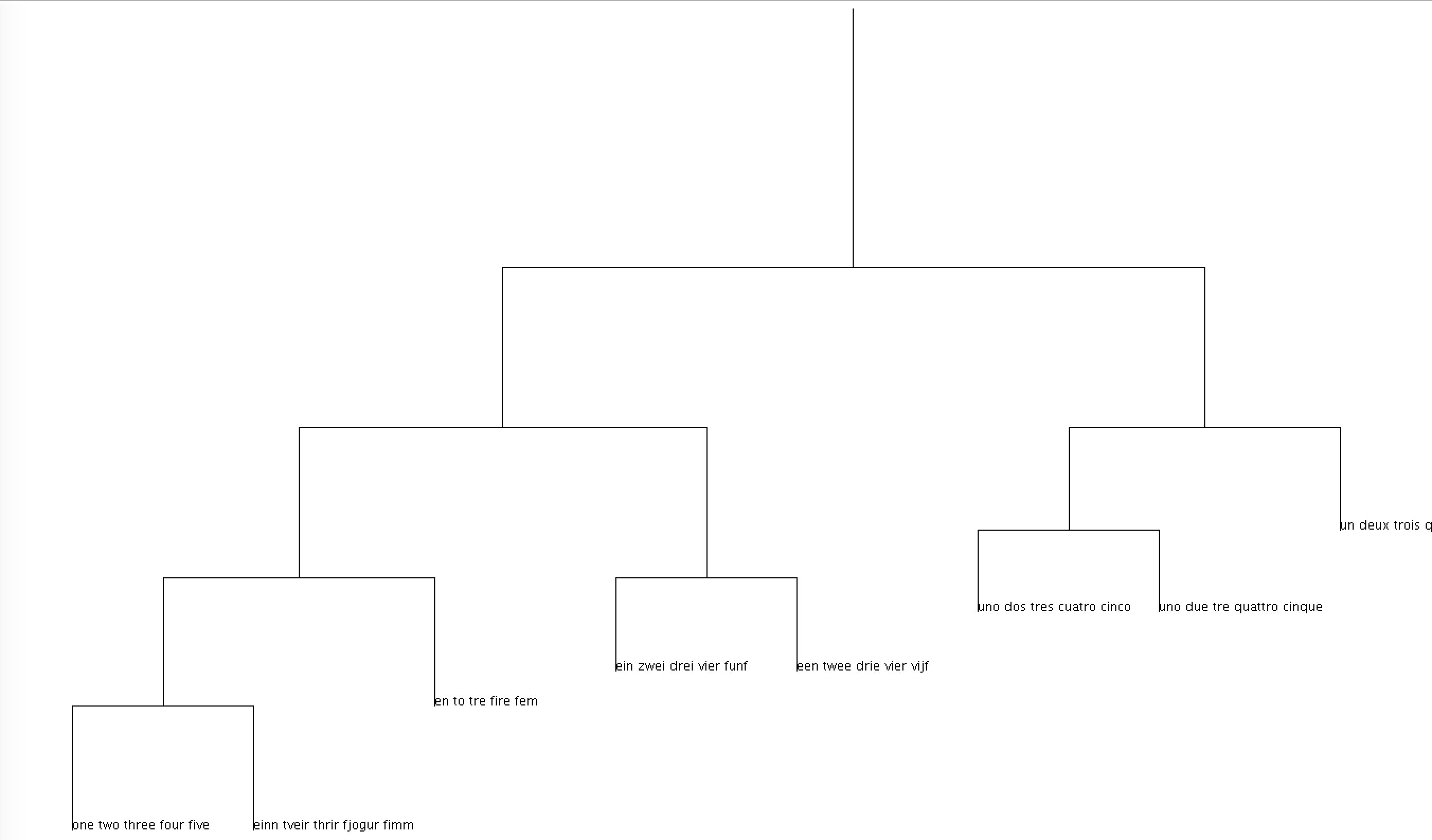
yuweic3

2017/9/12

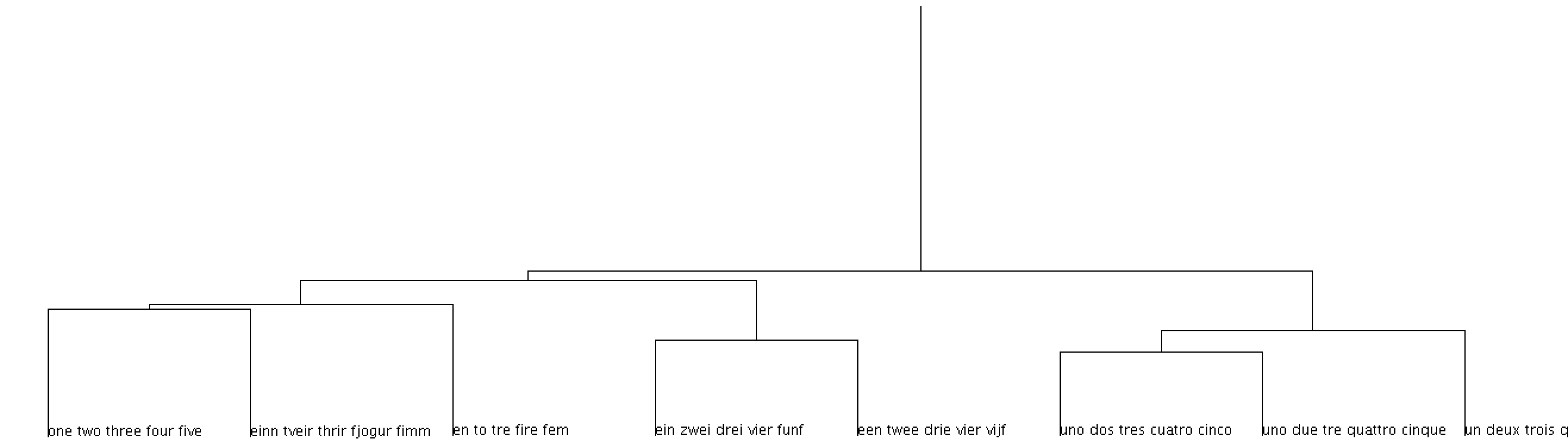
I add Italian in this dataset. “uno due tre quattro cinque”



Using EditDistance clustering with 2 clusters, we get a new tree



(distanceBranchLength = True)



(distanceBranchLength = False)

From the tree graph, we find that there are two clusters.

For cluster 0, we have 8 languages: ENG, SPA, GER, FRE, NOR, ICE, DUT and ITA.

For cluster 1, we have 2 languages: HUN and FIN

The new language(ITA) we add is assigned to cluster 0.

Besides, among the languages in the cluster 0, ITA has more similar features with SPA than other language.