

# Clustering of Analogies for Inter-Language Similarities

Software project - 3<sup>rd</sup> presentation

Justine Diliberto, Cindy Pereira, Anna Nikiforovskaja

Université de Lorraine, IDMC

04.11.2021



UNIVERSITÉ  
DE LORRAINE



Institut des  
sciences du Digital  
Management & Cognition

# Summary of the project

Subject: Analogies between morphological rules  
to stay -> stayed PST  
to play -> ? PST  
= PLAYED

Main goal: continuation of work done by Safa et al. to find out more about the closeness of languages and if they have common rules

Final product: Predict if two languages will transfer well, based on the rules they share

# What was done before

- read some articles about classification of languages
  - tried running the baseline
  - experimented with some visualisation methods
- 
- adapted new dataset to baseline
  - ran a full transfer on new languages
  - read articles about language similarities
  - listed many possibly close pairs of languages
  - started studying some rules in these pairs

# Adapt the dataset

## Problems:

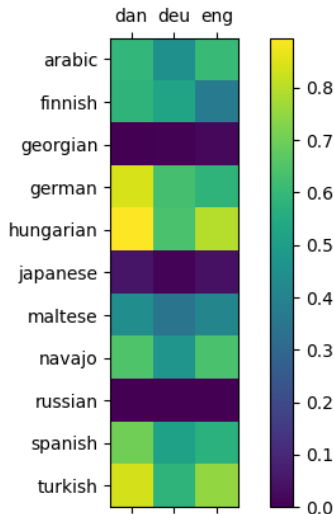
- SIGMORPHON 2020 is much bigger
- A little bit different format of data

## Overall solution:

- Dictionary of sets to store classes of analogies for each tag
- Restriction on number of analogy pairs taken from each class (1 000 000)

# Transfer on new dataset

Full transfer, 1000 negative analogies on new Danish, German and English

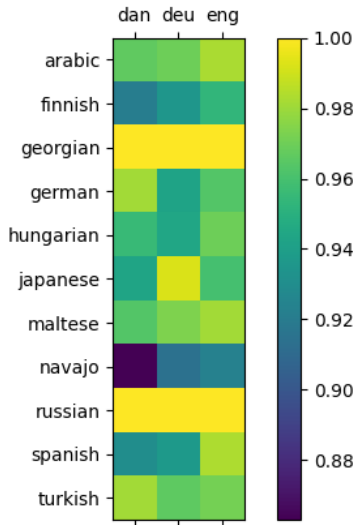


## Interesting things

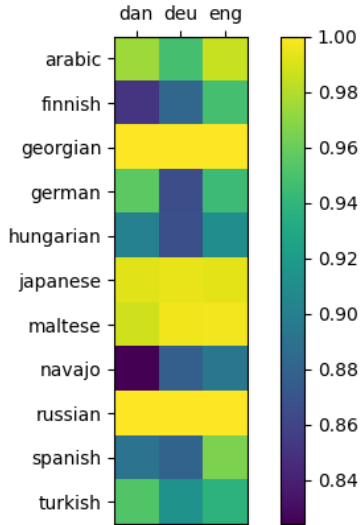
- German did not transfer well
  - New German dataset is much bigger and has a better representation of different morphological rules
- From Hungarian to Danish the best transfer

# Transfer on new dataset

## Positive, 1000



## Raw, 1000



# Close language pairs

Papers read :

- SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection
- Automatic Language Classification by means of Syntactic Dependency Networks
- Finding families: Quantitative methods in language classification

Hints:

- same genus or family
- similarities in morphological system
- subject of specific studies

# Similar morphological rules - Example

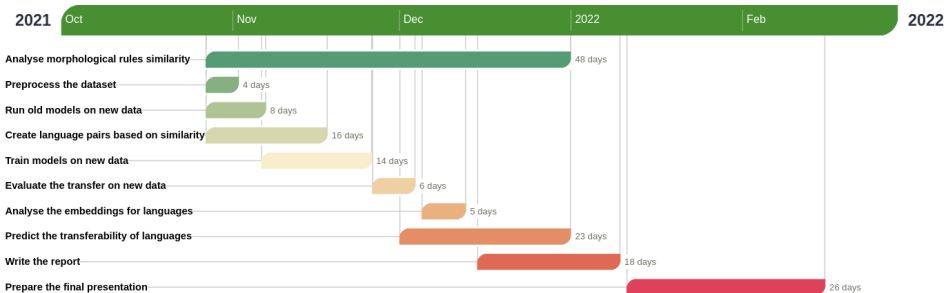
Comparison between verbs in Cebuano and Hiligaynon:

- Hiligaynon:  $\emptyset$ obra -> *nagobra* V;PST
- Cebuano: *mogunit* -> *nagunit* V;PRF;PST
- Hiligaynon:  $\emptyset$  -> *nag*
- Cebuano: *mog* -> *nag*

30 to 50% words in common between the two languages



# Timeline



Thank you for your attention.