

Clustering of Analogies for Inter-Language Similarities

Software project

Justine Diliberto, Cindy Pereira, Anna Nikiforovskaja

Université de Lorraine, IDMC

28.09.2021



UNIVERSITÉ
DE LORRAINE

IDMC Institut des
sciences du Digital
Management & Cognition

Project Description

- Inflection analogies
 - *cat is to cats as dog is to dogs*
- Multilingual
 - SIGMORPHON 2020 - task 0¹
- Predict the analogies
 - Deep-learning models
- Clusterisation of languages
 - Distance between languages

¹<https://sigmorphon.github.io/sharedtasks/2020/task0/>

Expected Final Product

Final product:

- Dendrograms
- Clustering visualisation (2D space)
- Python library
- Website?

SIGMORPHON 2020 shared task

- **Task 0: Typologically Diverse Morphological Inflection**
- Task 1: Multilingual Grapheme-to-Phoneme
- Task 2: Unsupervised Morphological Paradigm Completion
- 90 languages:
 - `<lemma inflected_form features>`
 - `<respect respected V.PTCP;PST>`

Plan

For next time:

- Study previous researches about language similarities and analogies
- Run the CNN models previously created²
- Visualise the results in different ways

Later:

- Try new approaches and visualise them
- Create library (website?)

²Safa Alsaïdi et al., *On the Transferability of Neural Models of Morphological Analogies*, 2021.