

# Clustering of Analogies for Inter-Language Similarities

Software project - 6<sup>th</sup> presentation

Justine Diliberto, Cindy Pereira, Anna Nikiforovskaja

Université de Lorraine, IDMC

14.01.2022



UNIVERSITÉ  
DE LORRAINE



Institut des  
sciences du Digital  
Management & Cognition

# Summary of the project

**Subject:** Analogies between morphological rules

**Main goal:** find out 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

- Listed possibly close and far pairs of languages
  - Looked for new ways to simplify rule extraction
  - Trained the multilingual model on old data and transferred to the new data
- 
- Chose 6 languages to focus on
  - Trained the model on these languages

# Change of plan

Our corpus: 90 languages = too much to study

We chose 6 languages instead:

- they belong to 3 main language families
- some are expected to be very close, and others to be far
- we know someone who can speak these languages (except for one)

# Meet our 6 languages

- English, German, Swedish (Germanic family)
- Finnish, Karelian (Uralic family)
- Mezquital Otomi (Oto-Manguean family)

In corpus:

- only verbs
- 5 types of inflections:
  - infinitive
  - singular 3rd person
  - past
  - past participle

In corpus:

- verbs and nouns
- 8 types of noun inflections:
  - 4 cases (nominative, accusative, dative, genitive)
  - 2 forms (singular, plural)
- 29 types of verb inflections:
  - 4 modes (indicative, subjunctive, imper, inf)
  - 2 forms (singular, plural)
  - 3 persons

In corpus:

- Verbs, nouns and adjectives
- Types of noun inflections:
  - Genitive, nominative, plural, singular, definite, indefinite
- Types of verb inflections:
  - Present, past, participle, converb, active, passive, imperative, indicative, infinitive
- Types of adjective inflections:
  - Feminine, masculine, neuter, plural, singular, comparative, superlative, definite, indefinite



In corpus:

- Verbs, nouns and adjectives
- Types of verb inflections:
  - Present, past, imperative, indicative, infinitive, active, passive, positive, conditional, potential, plural, singular
  - 3 persons
- Types of noun and adjective inflections:
  - 10 cases
  - 3 cases (genitive, nominative, accusative)
  - 2 forms (plural, singular)

In corpus:

- Verbs, nouns and adjectives
- Types of verb inflections:
  - Present, past, imperative, indicative, perfect, infinitive, positive, negative, imperfective, conditional, potential, plural, singular
  - 3 persons
- Types of noun or adjective inflections:
  - 10 cases
  - 3 cases (genitive, nominative, accusative)
  - 2 forms (plural, singular)

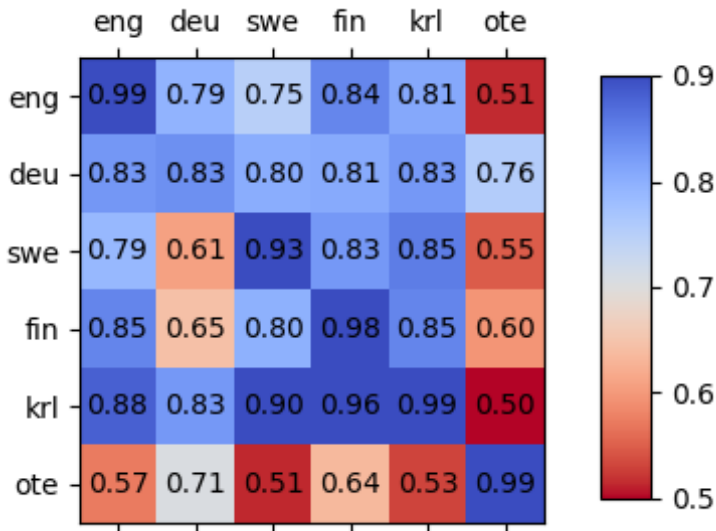
In corpus:

- only verbs
- only singular
- 18 types of inflections:
  - perfect, perfective, imperfective, irregular
  - past, present
  - 3 persons

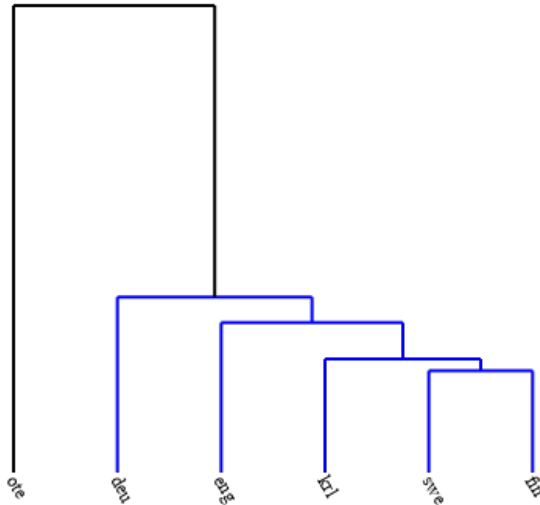
# Transfer between the 6 languages

- Balanced training on:
  - English, German, Swedish - Germanic group
  - Finnish, Karelian - Uralic group
  - Mezquital Otomi - Oto-Manguean group
- Partial transfer evaluation with  $F_1 = 2 \cdot \frac{p \cdot n}{p+n}$ , where:
  - $p$  — accuracy on positive analogies
  - $n$  — accuracy on negative analogies

# Partial transfer: results



# Partial transfer: dendrogram



# Observations of the transfer

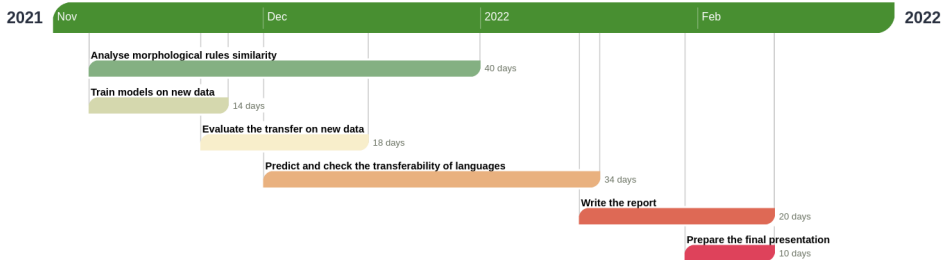
Non-surprising:

- Otomi is the furthest language.
- Finnish and Karelian are close.

Surprising:

- Swedish does not transfer well to German.
- German transfers well to everything.
- Karelian transfers well to everything except for Otomi.

# Timeline





Thank you for your attention!