

TOWARDS AN EMPIRICALLY GROUNDED FRAMEWORK FOR EMOTION ANALYSIS

Background

Dutch emotion detection system for identifying emotions in texts by using machine learning

First step = manually labeling data

Als antwoord op @DordtsePolitiek en @ADnl waardeloos is deze 🤬 treinreizen word een stuk minder plezierig 😞 hoef je niet eens in dord te wonen omdat te ervaren



Heftig, je eigen opa met de dag achteruit zien gaan 😞 hij kan bijna niks meer... 😞



Problem

- No consensus on standard emotion framework (Ekman, Plutchik, Izard, Roseman, ...)
- Label set often chosen arbitrary, not empirically grounded

Goal

Establish empirically grounded label set for emotion detection on Dutch Tweets
 → Start from psychological theory
 → Perform experiments on real-life data

Approach

- Collect corpus of real-life data in domain of the task (Dutch Tweets)
- Annotate corpus with initial extensive label set
- Cluster analysis to merge labels

Data

- 229 emotional Dutch tweets (emoji's as queries in Twiqs.nl)
- Multi-label annotation by 3 trained annotators
- IAA – mean Kappa score: 0.498

Label set

Requirements initial label set:

- sufficiently large
- capture enough nuances
- even distribution regarding sentiment polarity
- Not biased towards certain basic emotion set

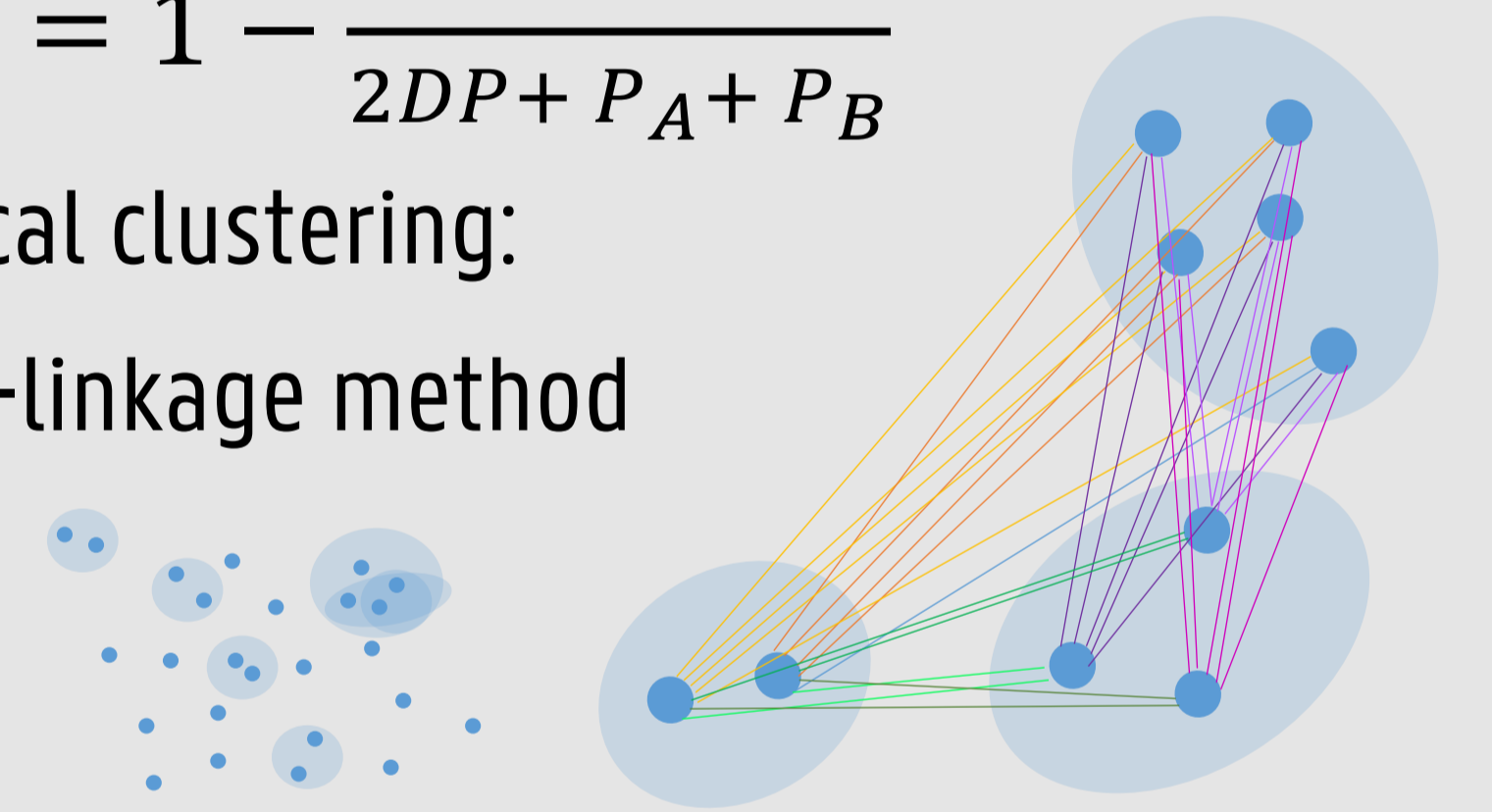
→ 25 emotion categories from Shaver et al. (1987): *anger, contentment, disappointment, disgust, enthralment, enthusiasm, envy, fear, frustration, irritation, joy, longing, love, lust, nervousness, optimism, pity, pride, rejection, relief, remorse, sadness, suffering, surprise, torment*

Clustering

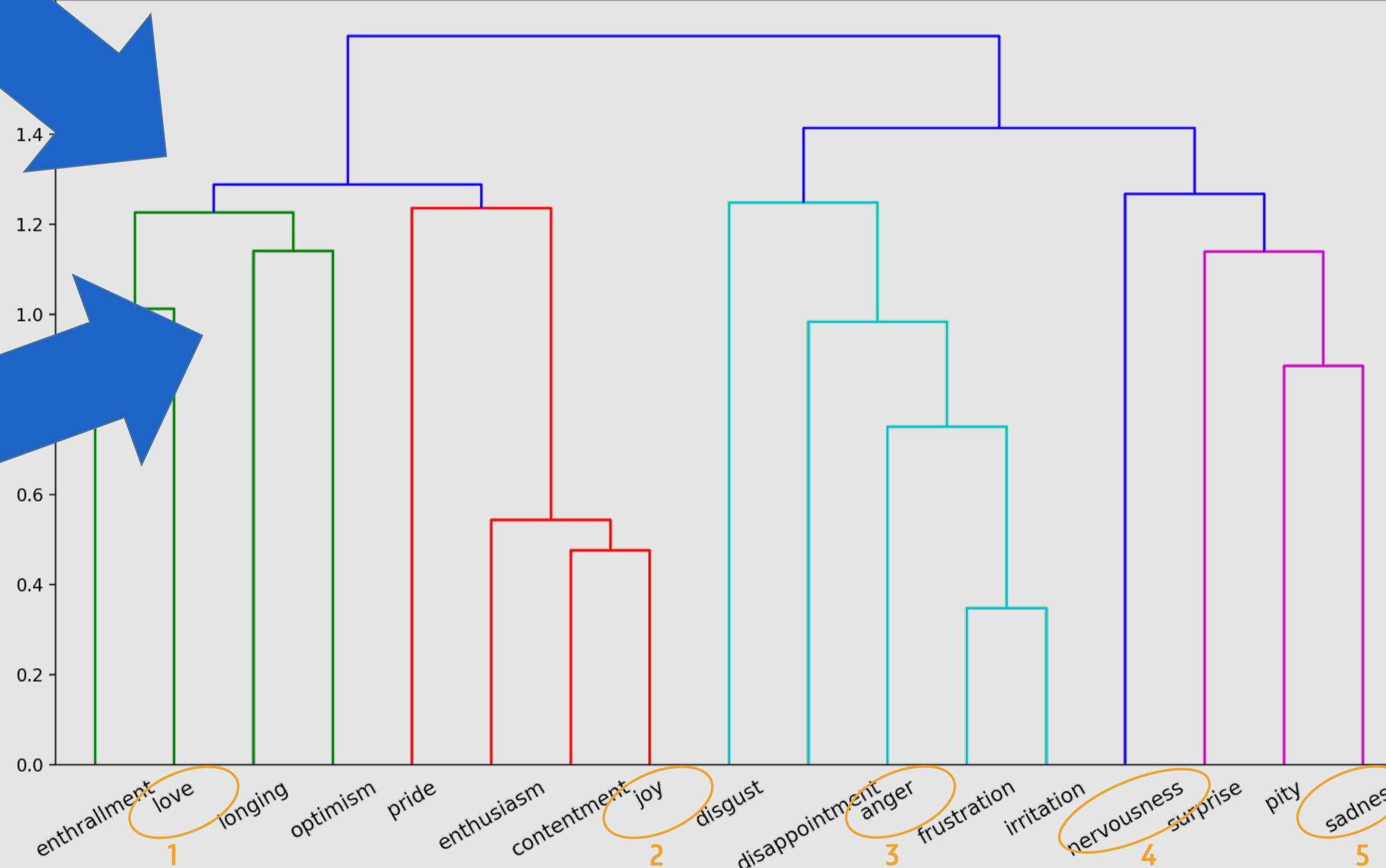
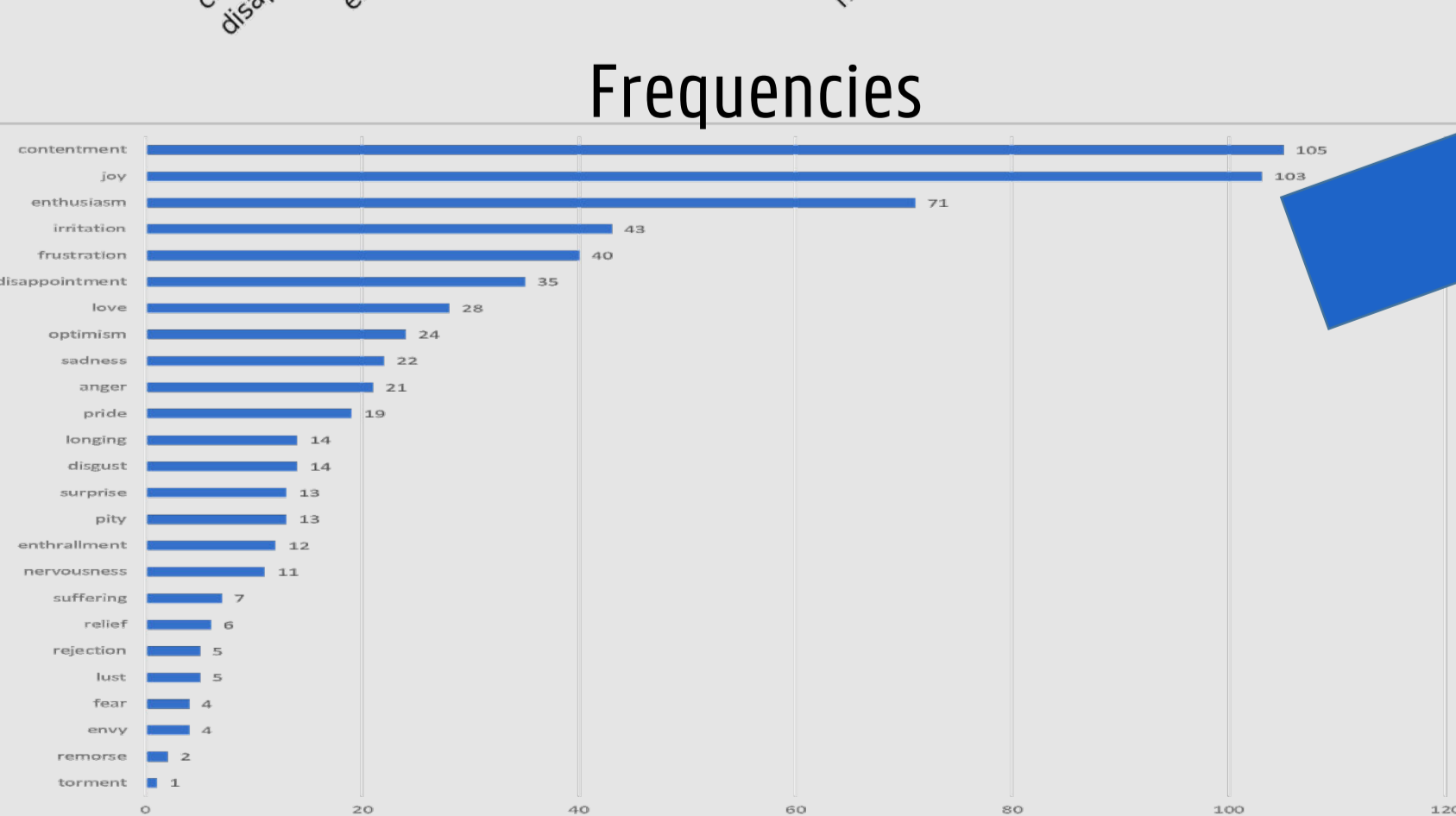
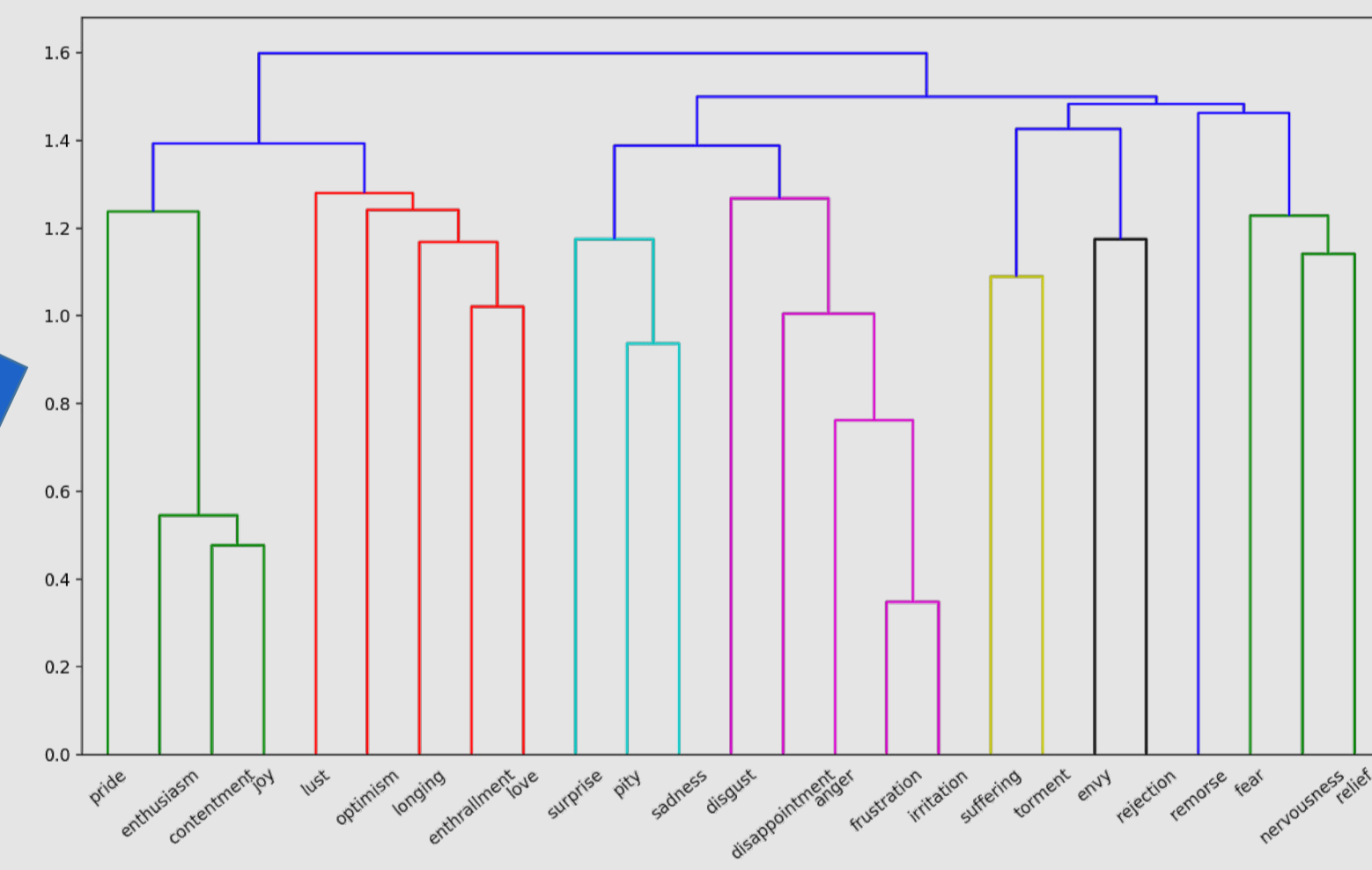
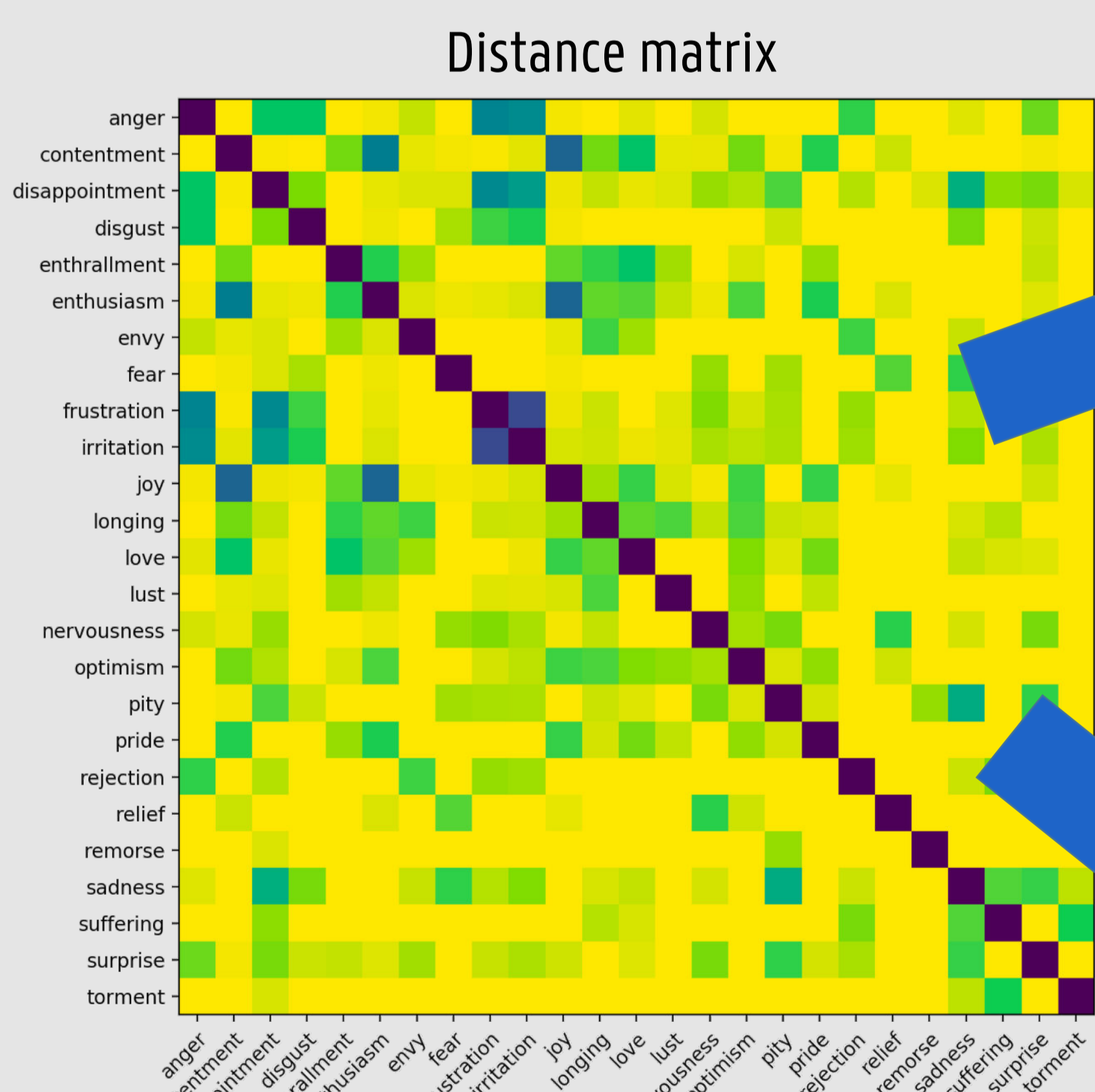
- Vector per emotion category: → 229-dimensional vector → 25 vectors in total
- Dice dissimilarity between each emotion vector pair:

$$d(A, B) = 1 - \frac{2DP}{2DP + P_A + P_B}$$

- Hierarchical clustering: weighted-linkage method



Results



Conclusion



- Frequency & cluster analysis resulted in label set with 5 basic emotions
- Fair resemblance with Ekman's basic emotions
- First study that gives empirical motivation for chosen labels
- Labels more grounded in the task & domain

Contact

luna.debruyne@ugent.be
<https://www.lt3.ugent.be/people/luna-de-bruyne>

LT³ Research Group

@ugentnlp

Ghent University