

# Quantitative and Qualitative Evaluation of Human and Machine-Translated EU Economic Texts in the English-Greek Language Pair

Dimitris Boumparis\* and Christos Giannoutsos\*\*

\*Master's student, University of Antwerp, Belgium

\*\*Freelance Translator, Rhodes, Greece

## Corpus

**650+ texts**

from the website of the ECB

**English-Greek**

originals vs translations

**Google NMT**

neural translations

## Preprocessing

- Remove noise (images, links)
- Lowercasing
- Remove stopwords
- Tokenization with spaCy

## Human vs Machine Translations

**85% similar**

Cosine similarity (avg)

**10.5% longer**

Total word length (avg)

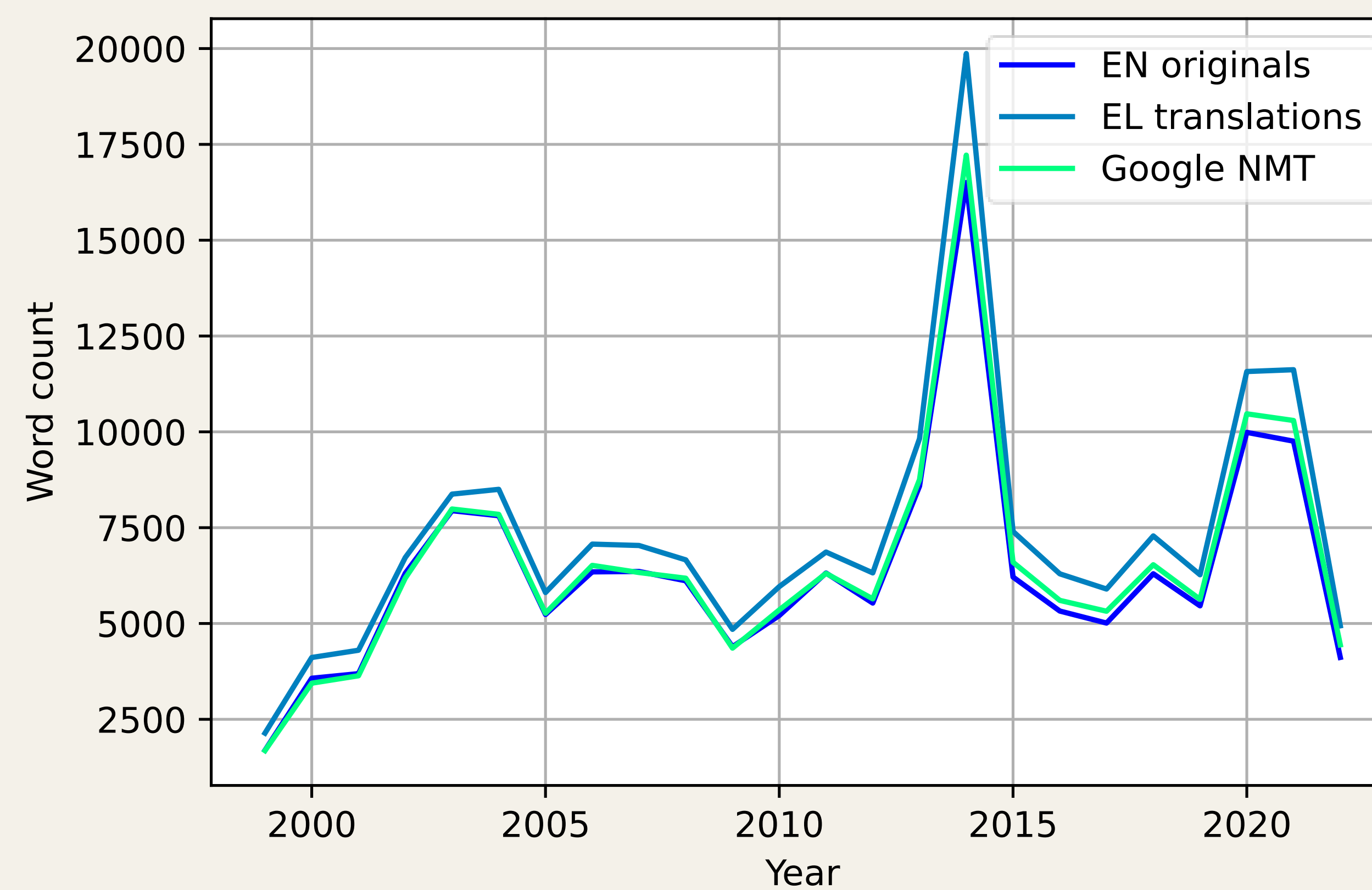
**40 points**

BLEU score for NMT (understandable to good translations)

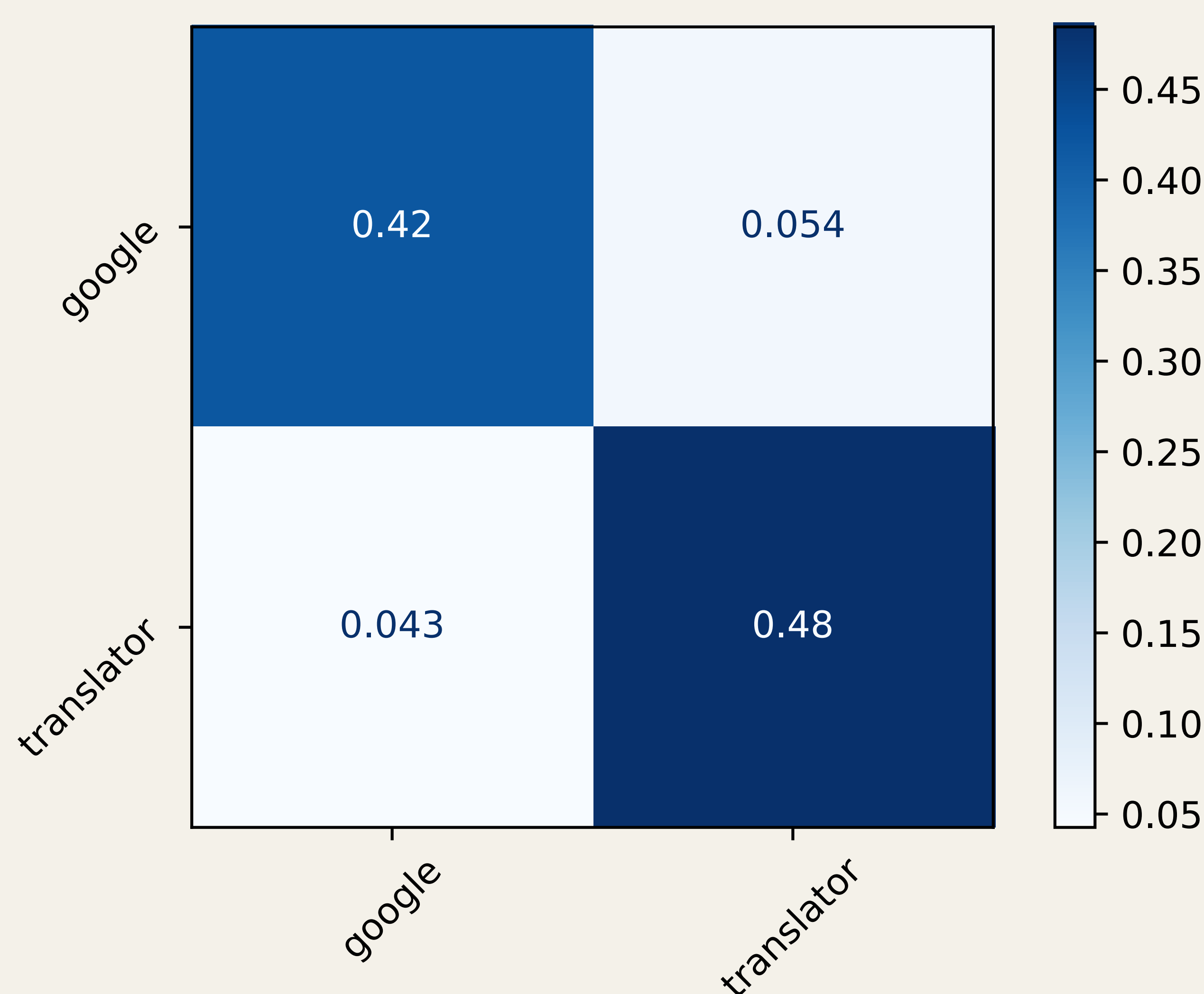
## Methodology

- **Supervised ML (SVM):**
  - Features:
    - 1000 most common **bigrams**
    - 1000 most common **trigrams**
  - **10-fold** Cross-Validation with **Grid Search**
  - **.90** f1-macro score
- **Unsupervised ML:**
  - **Principal Component Analysis (PCA)**
  - **Hierarchical Cluster Analysis (HCA)**
  - **t-distributed Stochastic Neighbour Embedding (tSNE)**

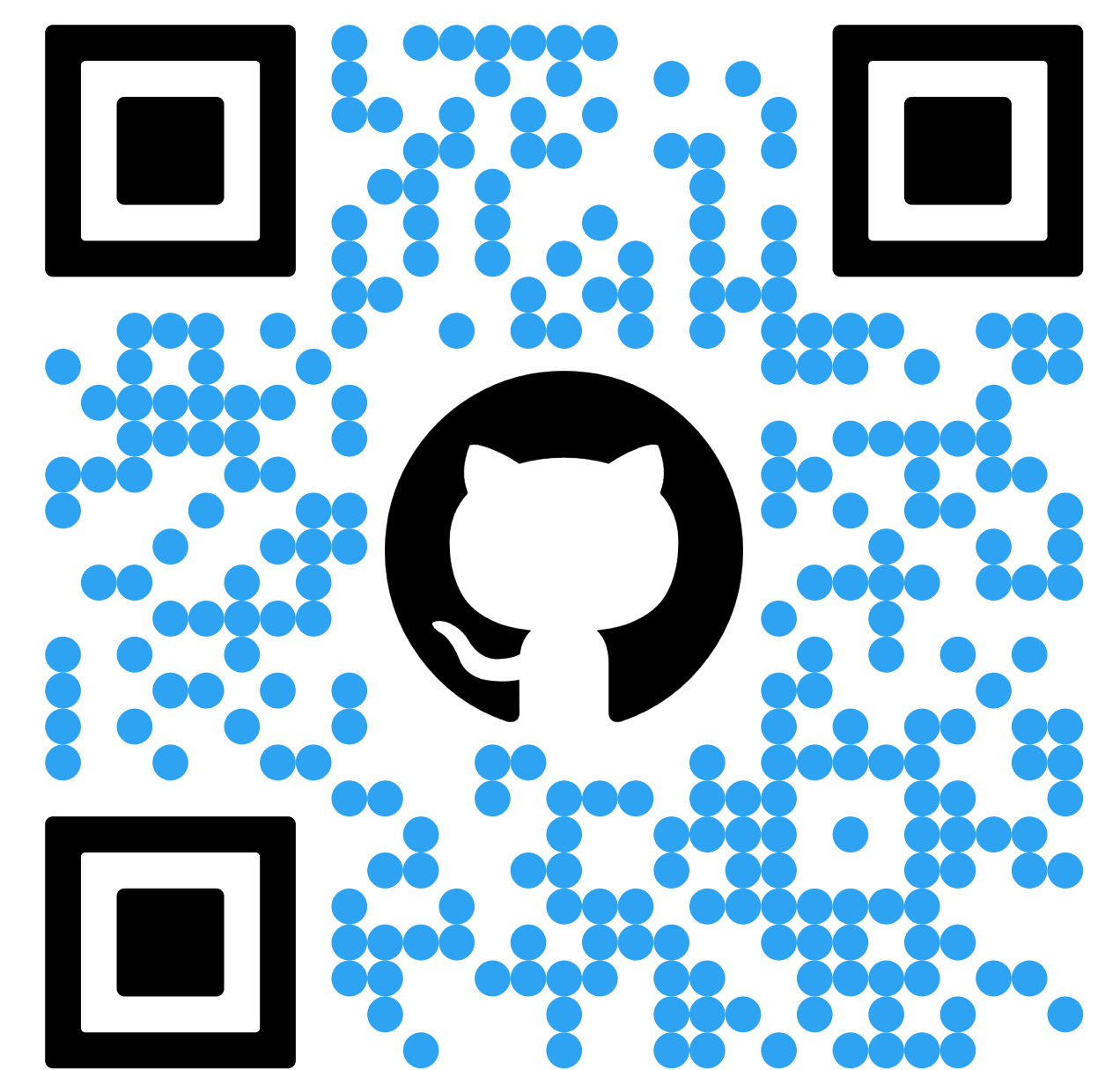
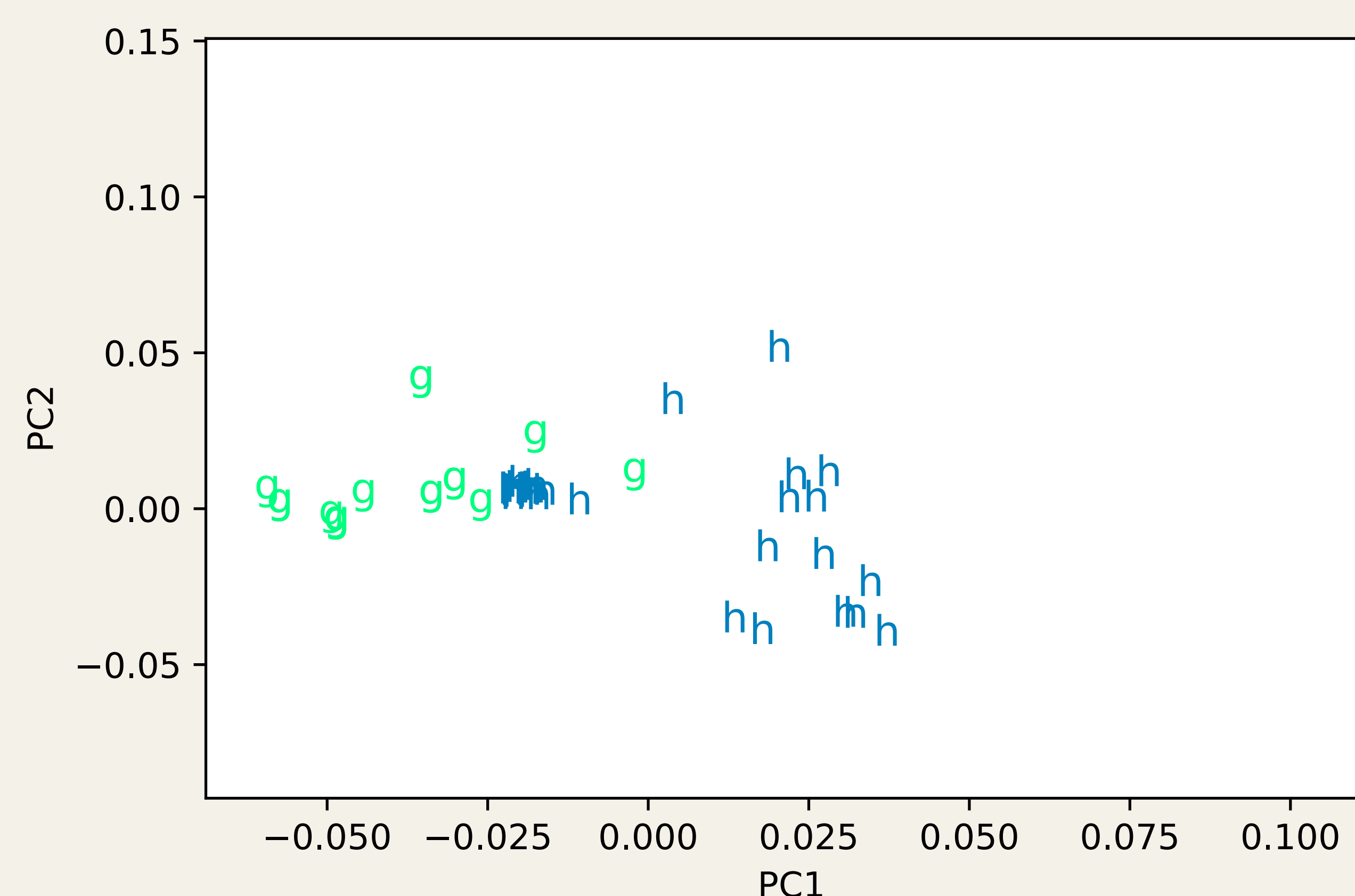
## Descriptive Statistics



## Support Vector Machines Classifier



## Principal Component Analysis



**Open source (eventually)**

Check out the code on GitHub!

<https://github.com/dimboump/nettt-2022>

## Conclusion

- Human and machine translations are still **easy to tell apart**.
- Human translators use **different collocations**.
- **Post-editing** is gaining **more attention** over 100% human translation.
- Humans tend to **translate more similarly**.
- **Repetitiveness** of data corresponds to **even more similar texts** (cf. press releases).

## Food for thought

- Should we **distinguish human from machine translations**?
- Is the 'cosine' of **each person's idiolect** becoming all the **more narrow**?
- Will data eventually make us **speak, write** and **translate more similarly**?

## Contact Information

Dimitris Boumparis

[dimitrios.boumparis@student.uantwerpen.be](mailto:dimitrios.boumparis@student.uantwerpen.be)

Christos Giannoutsos

[christosgiannoutsos@gmail.com](mailto:christosgiannoutsos@gmail.com)