# AUTOMATIC INFORMATION SEARCH FOR COUNTERING COVID-19 MISINFORMATION THROUGH SEMANTIC SIMILARITY

SCAN ME

*MSc project*


**Author: ÁLVARO HUERTAS GARCÍA**

**Directors: Manuel Sánchez-Montañés Isla, Alejandro Martín García**

MSc Project - MSc in Bioinformatics and Computational **Biology**

UNIVERSIDAD AUTÓNOMA DE MADRID – UNIVERSIDAD POLITÉCNICA DE MADRID


## INTRODUCTION

The technology we rely on to stay connected and informed is enabling and amplifying an infodemic that continues to undermine the global response and jeopardizes measures to control the COVID-19 pandemic. Social media platforms have been identified as the best sources for monitoring misinformation and dispelling rumors among the general people. The use of Natural Language Processing (NLP) can help to detect and assess them in an automatic approach.

## OBJECTIVES

**Literature review of the background and state-of-the-art in Natural Language Processing**

- To apply Deep Learning models for countering misinformation through semantic similarity
- To develop a collection of tweets from fact-checkers Twitter accounts
- To develop a dashboard to visualize the results

## METHODOLOGY

**SEMANTIC SEARCH** – Compute the cosine similarity between two sentences, unkwown source vs verified, using the embeddings from Transformer-based models

**TOPIC MODELING** – Unsupervised clustering of the verified claims dataset extracted from Twitter

**DASHBOARD** – Interactive web page using Dash to analyze unkwown claims

## KEYWORDS

Natural Language Processing, Deep Learning, Transformers, Topic Modeling. Sentiment Analysis, COVID-19, Infodemic, Misinformation, Factchecking

## RESULTS AND DISCUSSION

To overcome the massive computational overhead produced by original Transformer architecture, deriving semantically meaningful sentence embeddings can be achieved by using siamese architectures for training. Siamese architectures consist of two pre-trained Transformer-based models with tied weights that can be fine-tuned on a specific task like compute similarity scores (see Figure 1). This approach is also called dual-encoders o bi-encoders.

In Semantic Textual Similarity (STS) tasks, the systems need to compute how similar two sentences are, returning a similarity score between 0 and 5. STS Benchmark comprises a selection of the English datasets used in the STS tasks between 2012 and 2017.

We extend this dataset to 15 languages and use it to fine-tuned multilingual bi-encoders in STS tasks. We focused on applying these semantic-aware models to measure how similar two sentences are, being one an unknown claim and the other one a verified tweet from Fact-checkers. We achieved a 0.82 average score within the different multilingual STS tasks using the official Spearman correlation coefficient as the evaluation metric.

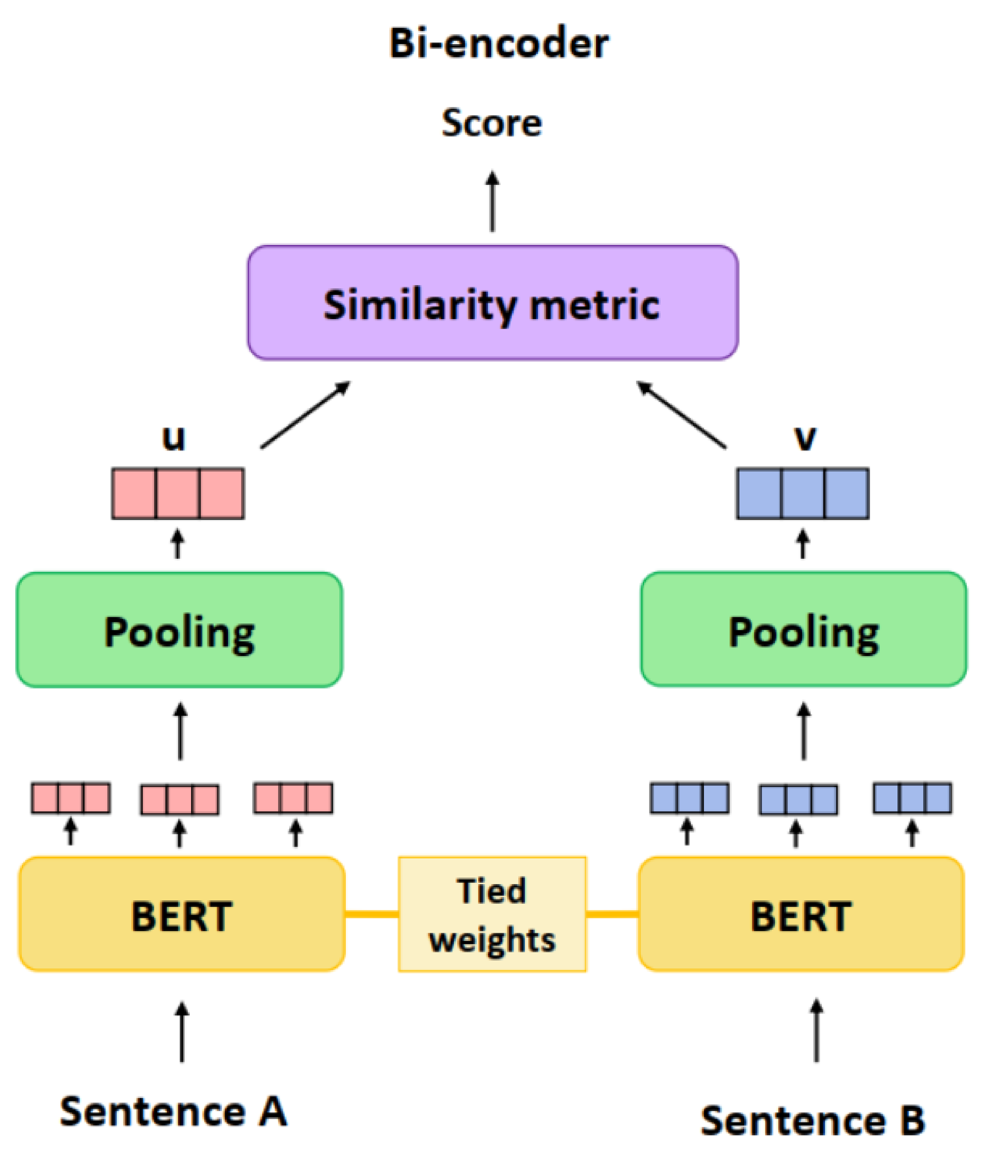


*Figure 1 – Illustration of the Transformer-based model architecture applied to compute semantic similairty*

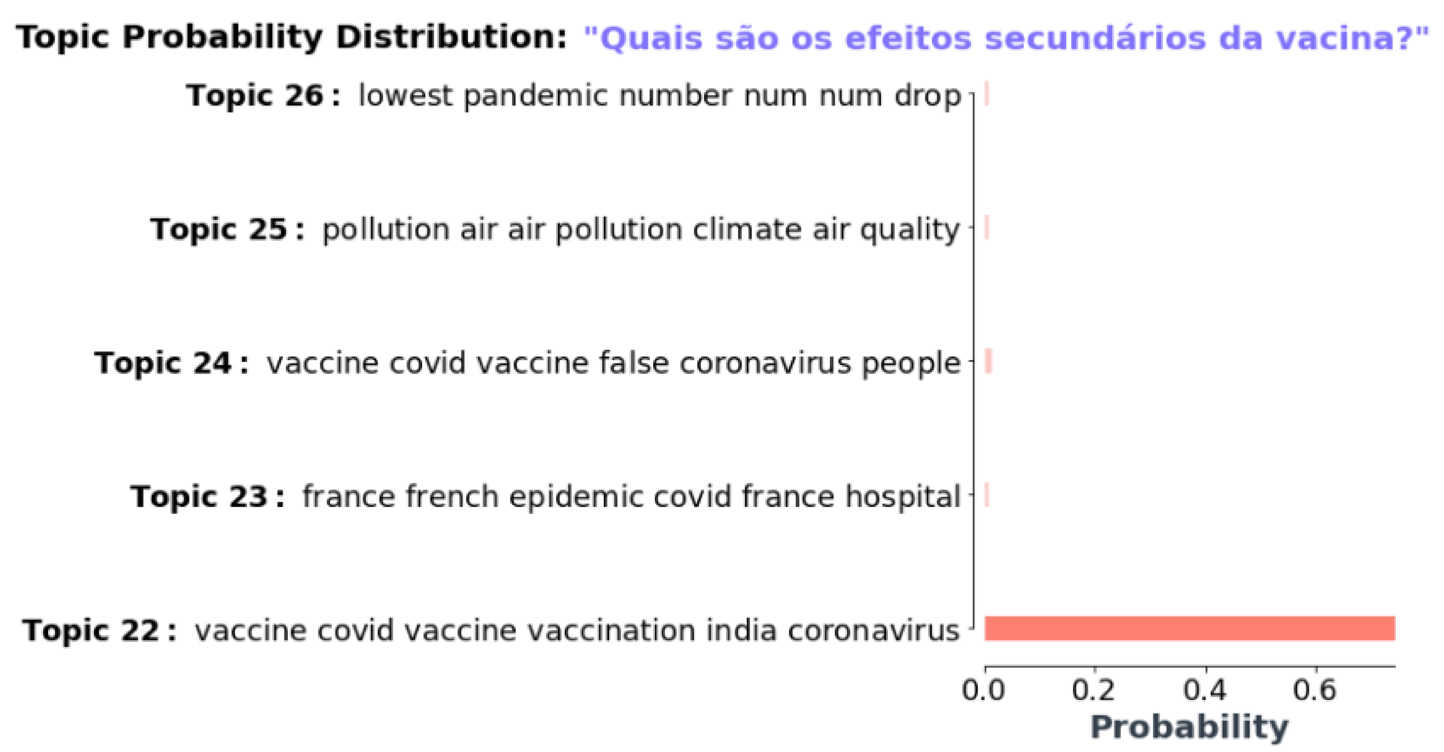


*Figure 2 – Example of topic detections from an unkwon claim according to the Twitter database*

We have also developed a topic NLP model that gathers documents into a set of interpretable topics in an unsupervised way, where each topic embodies a group of words associated under a single theme.

This topic model uses the embeddings computed from the Fact-checker tweets to organise the database into topics to fulfil data interpretability. Furthermore, the unsupervised topics detected are applied to predict the related topic of an unknown claim (see Figure 2 for an example where the unknown claim is in blue).

## CONCLUSIONS

It has been developed a set of 5 multilingual Transformer-based models fine-tuned on STSb to study the semantic similarity between an unknown text and a verified text.

The STS Benchmark dataset has been extended to 15 languages, available to the scientific community and used to fine-tune the models.

Our fact-checked tweets database combined with the multilingual topic model developed denotes that we can capture a diversity of topics gaining insights into how our database is structured. Furthermore, its use to predict a text's topic has been proved.

Taking advantage of Dash tool, we have developed a dashboard where the data and models mentioned above can be accessible.

**Multilingual STSb dataset:**
https://github.com/Huertas97/Multilingual-STSB

**Figures Repository:**
https://sites.google.com/view/tfm-alvaro-huertas-garcia/home?authuser=0

**Dashboard:** https://youtu.be/04CnFbcJWn8

**References:**


1. Steven Abreu. Automated Architecture Design for Deep Neural Networks. 2019. arXiv: 1908.10714 [cs.LG]..
2. Hunt Allcott and Matthew Gentzkow. "Social media and fake news in the 2016 election". In: Journal of economic perspectives 31.2 (2017), pp. 211–36.
3. Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019. arXiv: 1810.04805 [cs.CL]
4. Daniel Cer et al. "SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation". In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (2017). doi: 10.18653/v1/s17-2001.