9th International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Lost in Translation: Investigating Multilingual Misinformation & Its Evolution

Keywords: misinformation, fact-checking, multilingual NLP, information diffusion, social media

Cross-lingual misinformation, or misinformation that crosses language barriers, is a critical area of study due to the global nature of information dissemination. Despite numerous studies investigating the spread of misinformation, there remains a gap in our understanding of its evolution and the extent to which it crosses language barriers. Understanding the prevalence and characteristics of cross-lingual misinformation is essential for developing effective strategies to combat its spread and minimize its impact, particularly in the current era where misinformation can have far-reaching consequences, such as influencing political outcomes, shaping public opinion, and impacting public health. By investigating the extent of cross-lingual misinformation, we can gain valuable insights into the nature of this phenomenon and develop more effective strategies for combating its spread.

Previous research indicates that misinformation can be more problematic in some languages and countries than others' [2]. However, few studies have empirically investigated cross-language misinformation diffusion and how it contributes to a global misinformation network. This paper discusses potential methods to understand the characteristics that distinguish misinformation in mono- and multi-language settings. We empirically evaluated the cross-language misinformation diffusion with the most comprehensive fact-checking dataset to date. Based on the clustered misinformation claims, we applied rumor mutation theoretical framework [1] to show early results of a qualitative assessment of the differences between multi- mono-lingual claims, and the mutation of multilingual misinformation when traveling across languages.

We used Google Fact-check and fact-checks scraped by an NGO. The combination of these datasets results in 244,196 unique fact-checks from 2019 to 2023. Each observation represents a fact check with an associated claim, verdict, date, and author. The dataset is based on the ClaimReview tagging system, an HTML markup for fact-checkers to standardize their work. To compare misinformation spread across languages, we embedded all fact-checks with LaBSE embeddings [3]. The 109 languages covered account for 99.30% of our data.

We employed Locality Sensitive Hashing (LSH) to reduce the number of computations. LSH retrieves the most-similar observations for each fact-check without explicitly computing each similarity. We used 100 hyperplanes and retrieved the 500 nearest neighbours.[1] We then calculated the cosine similarity of each claim with its 500 nearest neighbours. The resulting edge list can be modeled as an extremely sparse graph, with clusters of densely connected nodes. Each connected component is a cluster of fact-checks where each element has at least one neighbour that exceeds the preset distance. To determine the optimal minimal distance we evaluated the average intra-cluster and the inter-cluster inertia. After manual validation the final cosine-similarity used was 0.875. Which is in line with previous research [4]. Figure 1 displays an exemplary cluster of eight fact-checks authored in three different languages.

The clustering results revealed the following preliminary findings. Above all, most misinformation is not fact-checked twice, and therefore seems to be unique to its location, language, or cultural bubble. Only 19.1% of fact-checks occur more than once in our data set, accounting for 9.17% of claims. Therefore, the proportion of multilingual misinformation diffusion represents

[1] In an initial test, LSH returned the correct 500 nearest neighbours in over 99% of cases.

only a small proportion of the global misinformation network at around 5.88% of fact-checks and 2.37% of claims, which shows that most misinformation is location-, culture- or language-sensitive on the Internet. Of the claims, consisting of more than one fact-check, 24.9% of claims were multilingual. We demonstrate the propensity to cluster within languages is significantly elevated compared to a Null-model (see Figure 2). English is the most significant bridging language in all multilingual misinformation claims (Figure 3). And cross-language transmission is more likely to happen in countries or regions with multilingual contexts, such as English, Hindi, Bengali, and Telugu; or geologically or culturally close, for example, Portuguese and Spanish.

Last, we note two potential factors that affect cross-language misinformation diffusion via computer-mediated discourse analysis. One is the locality of topics. Cross-language misinformation often focuses on shared concerns between communities, like vaccine rumors, COVID severity, and man-made virus theories. It may also relate to regional public or political figures. The other is the valence of description. For example, misinformation claiming that vaccines kill ”thousands” of people in English, changed to ”700,000 people” in German, Spanish, and Croatian, and ”millions” in Spanish and English again after some time. Compared to monolingual claims, multilingual claims are more likely to have altered numbers ($\beta = 0.074$, $t = 2.05$, $p < 0.05$). To test what type of misinformation is more likely to cross borders we fitted a BERTopic topic model. We extracted all topics referring to COVID, Climate Change, or local issues[2] and tested whether claims associated with topics relating to these issues were more likely to cross-languages. Initial results confirm that ”global” issues are fact-checked in more languages, and repeated more often. These differences in propensity to be multilingual could not be explained by the number of associated claims nor the variance of the underlying clusters. [3]

In conclusion, we find that while most misinformation is local, a substantial proportion of repeated misinformation crosses borders and languages. Our research demonstrates the prevalence of re-occurring fact-checks with 19.71% of fact-checks being repeated. We further show that a quarter of repeated misinformation crosses languages. By identifying the factors that contribute to the spread of misinformation across languages, our findings can inform more effective content moderation strategies and improve the current literature on fact-checking.

References

[1] Gordon W. Allport and Leo Postman. *The psychology of rumor*. The psychology of rumor. Pages: xiv, 247. Oxford, England: Henry Holt, 1947. xiv, 247.

[2] Meeyoung Cha et al. “Prevalence of Misinformation and Factchecks on the COVID-19 Pandemic in 35 Countries: Observational Infodemiology Study”. In: *JMIR Human Factors* 8.1 (Feb. 13, 2021), e23279. ISSN: 2292-9495. DOI: 10.2196/23279.

[3] Fangxiaoyu Feng et al. *Language-agnostic BERT Sentence Embedding*. Mar. 8, 2022. DOI: 10.48550/arXiv.2007.01852. arXiv: 2007.01852[cs].

[4] Ashkan Kazemi et al. “Claim Matching Beyond English to Scale Global Fact-Checking”. In: ACL-IJCNLP 2021. Aug. 2021, pp. 4504–4517. DOI: 10.18653/v1/2021.acl-long.347.

[2]i.e. containing references to local officials such as councilor, mayor, governor, representative, etc.

[3]It should be noted that this is the preliminary result. We also noticed other linguistic features, such as adjectives. We will continue to quantify the factors and model them in the next step.

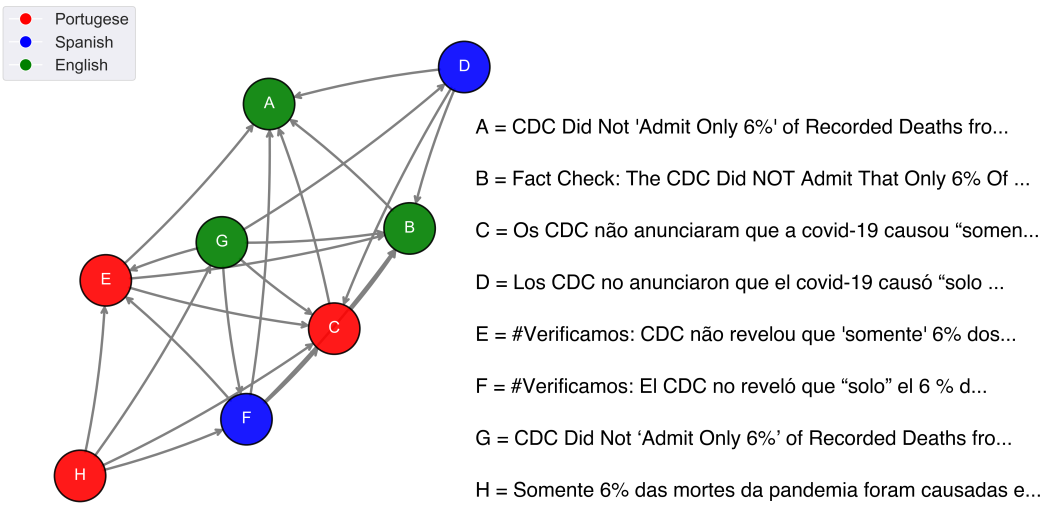

Figure 1: Example of a Densely Connected Cluster

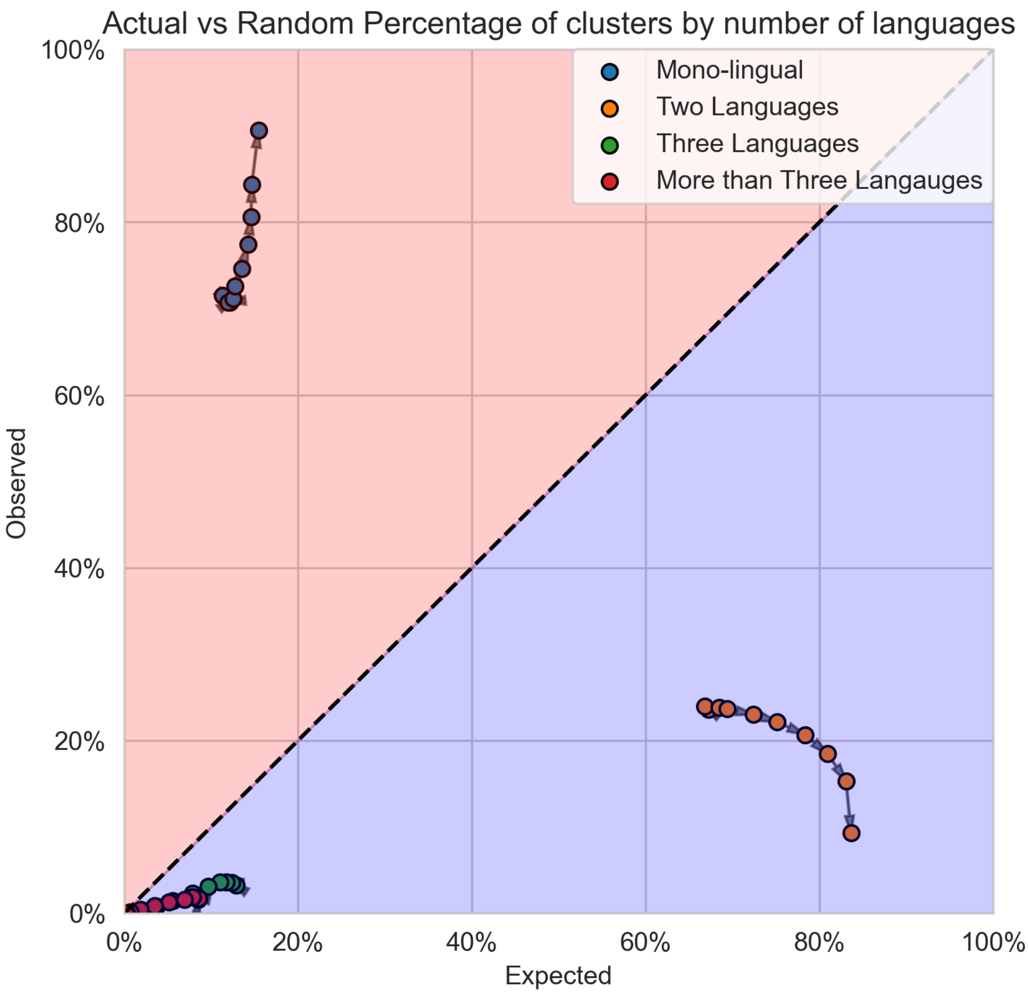

Figure 2: Experimental Evaluation of Language Homophily

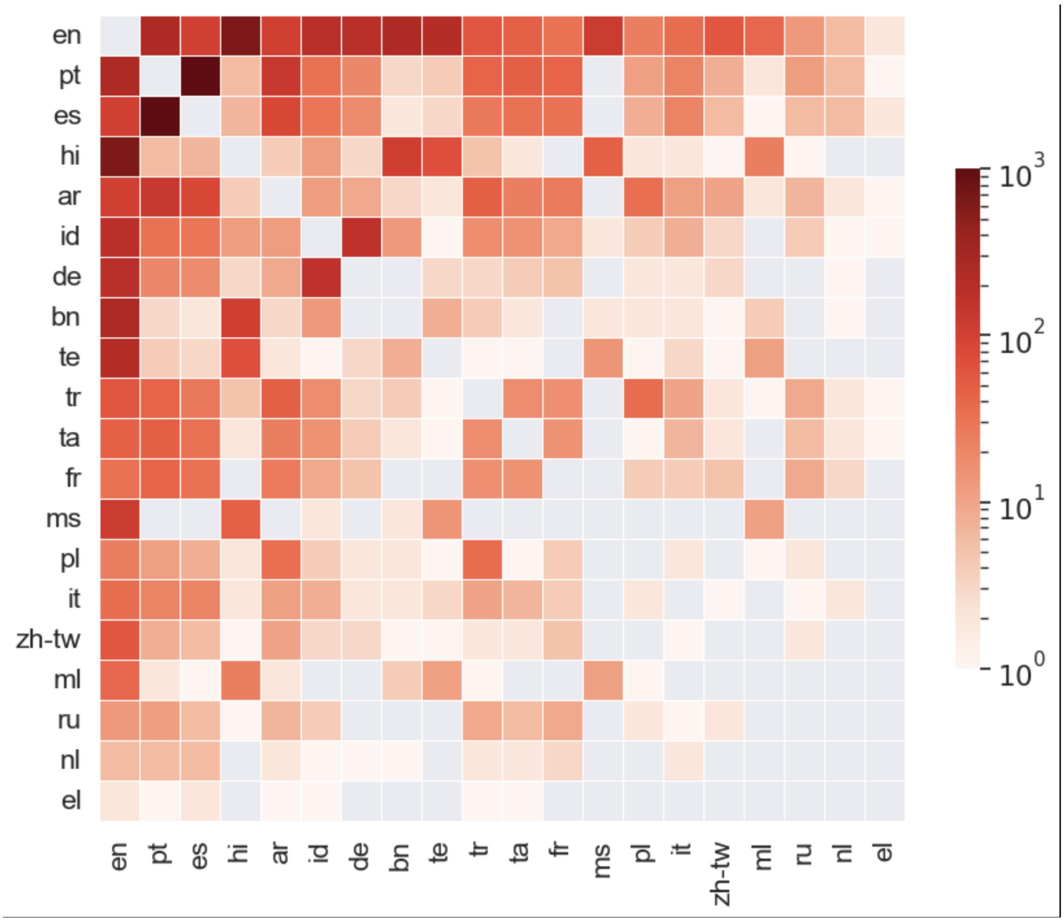

Figure 3: Co-occurrence of Top 20 Languages in Multilingual Misinformation Claims. Colours indicate the number of co-occureces of unique languages in claims.

Notes. Languages on the x-axis are English (en), Portuguese (pt), Spanish (es), Hindi (hi), Arabic (ar), Indonesian (id), German (de), Bengali (bn), Telugu (te), Turkish (tr), Tamil (ta), French (fr), Malay (ms), Polish (pl), Italian (it), Traditional Chinese (zh-tw), Malayalam (ml), Russian (ru), Dutch (nl), Greek (el).