

# Extracting and Analysing Metaphors in Migration Media Discourse: towards a Metaphor Annotation Scheme

Ana Zwitter Vitez<sup>1</sup>, Mojca Brglez<sup>1</sup>, Marko Robnik-Šikonja<sup>2</sup>, Tadej Škvorc<sup>2</sup>, Andreja Vezovnik<sup>3</sup>, Senja Pollak<sup>4</sup>

<sup>1</sup> University of Ljubljana, Faculty of Arts, <sup>2</sup> University of Ljubljana, Faculty of Computer and Information Science, <sup>3</sup> University of Ljubljana, Faculty of Social Sciences, <sup>4</sup> Jožef Stefan Institute

## INTRODUCTION

The key feature of metaphors is the establishment of parallels between two domains which are not literally connected. The metaphor *wave of migrants*, which represents *migrants as flood* (where *wave* is the source domain and *migrant* the target domain), may act as the basis for how we perceive migrants as a danger. This can lead to different social and political measures such as militarization of borders to assure safety for the majoritarian population.

Fairclough (1995) claims that different metaphors may correspond to different interests and perspectives and may therefore have different ideological loadings. Analysis of metaphors in media is therefore an important tool to unmask various types of ideologies, for instance racism, ethnocentrism or dehumanization.

## OBJECTIVES

The aim of the study is threefold. First, we present a neural metaphor extraction method for Slovene, where we use the method by Škvorc et al. (2022) for idiom detection and adapt it to detect metaphors. Next, we evaluate the metaphor extraction on a domain specific corpus of migration news to extract metaphorical sentences. Finally, we perform a fine-grained linguistic analysis of 100 examples, and introduce a novel metaphor annotation scheme, which can serve for future metaphor annotation initiatives.

## METHODOLOGY

To automatically detect metaphors, we use a neural-network classifier based on the SloBERTa model. Since large Slovene datasets of metaphors do not exist, we make use of transfer learning from a related problem: idiom detection. We first use a large dataset of Slovene idiomatic expressions (SloIE). This dataset focuses only on idioms and therefore does not cover the full scope of metaphors.

However, as both idioms and metaphors deal with highly context-dependent figurative language, we believe it is possible to transfer knowledge from one task to the other. We do this by first training the network on the SloIE corpus and then fine-tuning it on a smaller dataset of Slovene metaphors (KOMET).

The analysis of metaphors consists of multiple layers and steps, from their identification to a more fine-grained interpretation. For the identification step, we apply the principles of well-established MIPVU (Steen et al. 2010) procedure considering words as metaphors if their meaning in context differs from a more basic, physical and embodied sense. At the conceptual level, we annotate metaphors with relevant source and target domains.

|                              |  |                     |
|------------------------------|--|---------------------|
| <b>Example</b>               | <i>Police officers are directing all their efforts into handling the security conditions and ensuring order and peace with the arrival of such a large number of foreigners.</i> |                     |
| <b>Topic of migrations</b>   | Yes  |                     |
| <b>Linguistic annotation</b> | <b>Expression</b>  | Metaphor            |
|                              | <b>Source Frame</b>  | Handling            |
|                              | <b>Target Frame</b>  | Security conditions |
| <b>Conceptual annotation</b> | <b>Source Domain</b>   | Beast/Opponent      |
|                              | <b>Target Domain</b>   | Migrant situation   |
| <b>Stance toward</b>         | <b>Target Frame</b>  | Negative            |
|                              | <b>Migrations</b>  | Negative            |

Table 1. Linguistic annotation scheme.

## RESULTS

We evaluated the automatic detection using two different approaches: 1) training the network on metaphors only and 2) training the network on idioms and metaphors.

Training the network only on metaphors gives a small improvement over the default classifier (classification accuracy of 0.609, compared to the default accuracy of 0.575). Transfer learning from idioms improves the result with a classification accuracy of 0.725.

We attempted to further improve the result by extending the KOMET corpus with edge-case and unclear metaphors by using semi-supervised learning: we first trained the model on the KOMET corpus and then extended the training data with edge-case metaphors where the prediction certainty of the model was either above 95% or below 5% (i.e. sentences where the model was confident in its predictions). This approach performed worse than the default classifier.

The fine-grained linguistic analysis of the automatically extracted sentences revealed 40 sentences out of 100 contained metaphors related to migration. We found several different source domains that were used to frame the topic of migration, among which the most common were LIQUID, CONTAINER, NATURAL PHENOMENON, CONTAINER, OBJECT, and BURDEN. The most frequent target concepts framed by the metaphors were MIGRANTS, COUNTRIES, MIGRANT CRISIS, CRISIS WORKERS, and BORDERS. With regard to the sentiment and stance being conveyed by the examples, the stance towards migrants in general was predominantly negative (78%), while the sentiment toward the given target concept was a bit more balanced (70% negative).

| Source domain      | Freq. | Target domain     | Freq. |
|--------------------|-------|-------------------|-------|
| Liquid             | 8     | Migrants          | 12    |
| Container          | 6     | Country           | 9     |
| Natural phenomenon | 6     | Migration         | 7     |
| Object             | 6     | Migrant admission | 5     |
| Burden             | 6     | Borders           | 3     |
| Beast              | 4     | Regulation        | 3     |
| Journey            | 4     | Europe            | 2     |
| Living being       | 4     | Crisis workers    | 2     |
| Opponent           | 3     | Society           | 1     |
| Defender           | 3     |                   |       |
| Contestant         | 3     |                   |       |
| Weapon             | 2     |                   |       |
| Building           | 2     |                   |       |

Table 2. List of detected source and target domains.

## CONCLUSION

In the process of extracting and annotating metaphors in the selected examples from the Migrations corpus, we found various types of metaphors (such as migrants as liquid, host countries as containers), which matches the most frequently identified metaphorical uses related to migration in qualitative research. The results show that the proposed method can significantly shorten the analytical procedure, allows a much bigger set of data to be analysed, and works well on topical collections. In addition, automated extraction leads towards a more objective sample selection.

Some limitations of the proposed methodology are as follows. First, due to the automatic extraction we extracted many metaphoric sentences, however, the recall of the method should be computed on a representative sample directly from the corpus and not only from the sentences selected by the system. Next, the linguistic annotation procedure in which source and target domains are identified turned out to be challenging mostly because of lack of coherent identification of metaphors across annotators. In our case, the annotation process was done by two linguists. For this proof of concept study, the two annotators reached consensus on all the examples. When moving to a larger annotation campaign, we will release clear guidelines and use an overlap of annotated examples to assess the inter-annotator agreement.

## REFERENCES

- Fairclough, N. (1995). *Media Discourse*. London: Edward Arnold.
- Steen, J. G., Dorst, G. A., Berenike Herrman, J., Kaal, A. A., Krennmayr, T., Pasma, T. (2010). *A method for linguistic metaphor identification: from MIP to MIPVU*. Amsterdam: John Benjamins Publishing Company.
- Škvorc, T., Gantar, P. and Robnik-Šikonja, M., 2022. MICE: Mining Idioms with Contextual Embeddings. *Knowledge-Based Systems*, 235, p.107606.

## CONTACT

Ana.ZwitterVitez@ff.uni-lj.si