

# Simple discovery of COVID IS WAR metaphors using word embeddings

Mojca Brglez, Senja Pollak, Špela Vintar



# INTRODUCTION

- COVID-19: global problem and topic of discourse
- CMT: “metaphors [...] are among our principal vehicles for understanding” and “play a central role in the construction of social and political reality” (Lakoff & Johnson, 2003, p. 151)

## **DISEASE/COVID is WAR**

(TARGET DOMAIN is SOURCE DOMAIN)

# AIM & PREVIOUS WORKS

- Linguistic manifestations of COVID IS WAR
  - Limited vocabulary sets, manual lookup
  - Somewhat extended vocabulary sets obtained through a lot of manual work using several linguistic resources (dictionaries, thesauri, previous studies, samples from corpora)
  - Manual reading and annotation of smaller corpora
  - Complex NLP methods and specialized resources

# RESOURCES AND TOOLS

- Slovene corpus of metaphors [KOMET](#) (Antloga, 2020)
- [CLARIN token embeddings](#) (Ljubešič & Erjavec, 2018) trained with fastText (Bojanowski et al., 2017)
- Gensim library (Rehurek & Sojka, 2016)
- LemmaGen lemmatization module (Juršič et al., 2010)
- NLTK library (Bird et al., 2009)

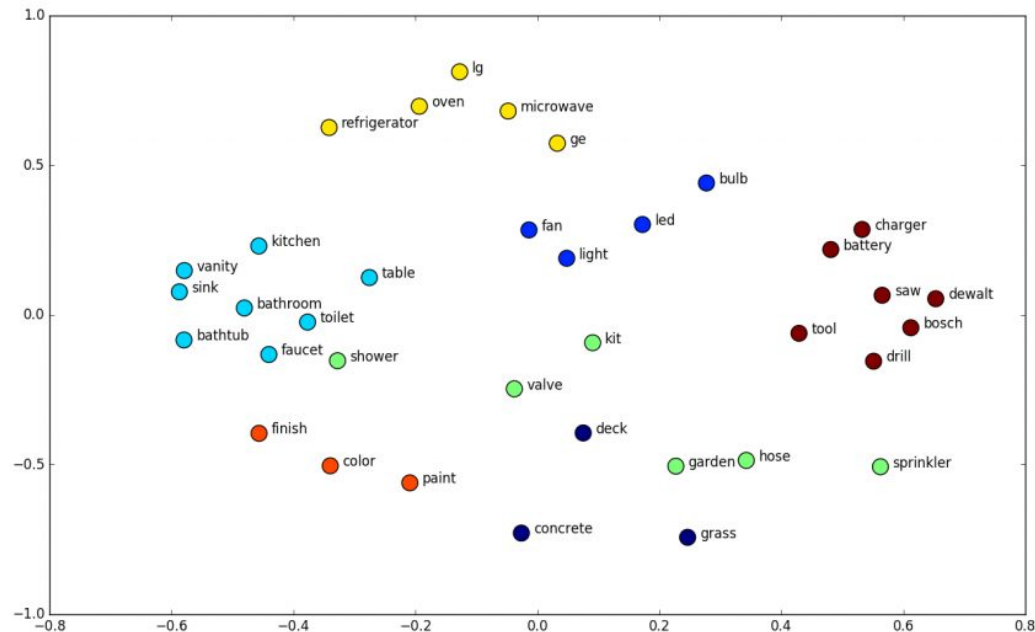
## DOMAIN CORPUS:

- COVID-19 news articles (233 articles crawled from the web (Feb – Dec 2020), 151,947 tokens)

# METHOD

- extending the vocabulary set of potentially metaphoric war-related lexis using word embeddings:

linguistic regularities, including syntactic and **semantic relationships** between words



# METHOD

## 1. Extract battle-related lexis from KOMET

```
<seg type="metaphor" subtype="frame" ana="#met.battle">  
  <seg type="metaphor" subtype="MRW1">  
    <w ana="mte:Ncmsn" msd="UposTag=NOUN|Case=Nom|Gender=Masc|Number=Sing" lemma="meč" xml:id="komet111.10.3.t15">meča</w>  
  </seg>  
</seg>
```

## 2. Extend vocabulary with embeddings: $N = 50$ nearest neighbours

## 3. Intersection filter (Vintar et al. 2020): intersection of $k = 3$ seed words

Initial seed set	Nearest neighbours (tokens)	Nearest neighbours (lemmas)	After intersection	Final vocabulary
51	2078	1539	<b>184</b>	<b>191</b>

# CORPUS EXPERIMENTS

Potentially metaphoric instances	Potentially metaphoric lemmas
335	49



## LITERAL USES

*Poveljstvo ameriške vojske za transport*

*US Transportation **Command***

## LEMMATIZATION ERRORS

*strela > **strel***

*lightning > **shot***

## OTHER TARGET DOMAIN

*kibernetski **napadi***

*cyber **attacks***



Targeted metaphoric instances	Targeted metaphoric lemmas
123	25

# RESULTS

- Mostly highly conventional metaphors (*boj* [fight] with 46 occurrences = 37,7 % of all metaphorical expressions!)

*Čeprav v **boju** z virusom to nikakor ni hitro.*

[Although this is by no means fast in the **fight** against the virus.]

*Kako bo jeseni, ko bodo »**udarili**« še drugi virusi?*

[What will happen in autumn, when other viruses also “**strike**”?] ]

*Prvi organski sistem v organizmu, ki ga virus **napade**, povzroči pljučnico, ...*

[The first system in the organism that the virus **attacks** causes pneumonia ...]



# RESULTS

- 9 new lemmas not in the initial seed set

*soočiti, izbojevati, zmagati, obraniti, uiti, soočanje, spopadati, zoperstaviti, podleči*  
[confront, fight, win, defend, escape, confrontation, combat, oppose, succumb]

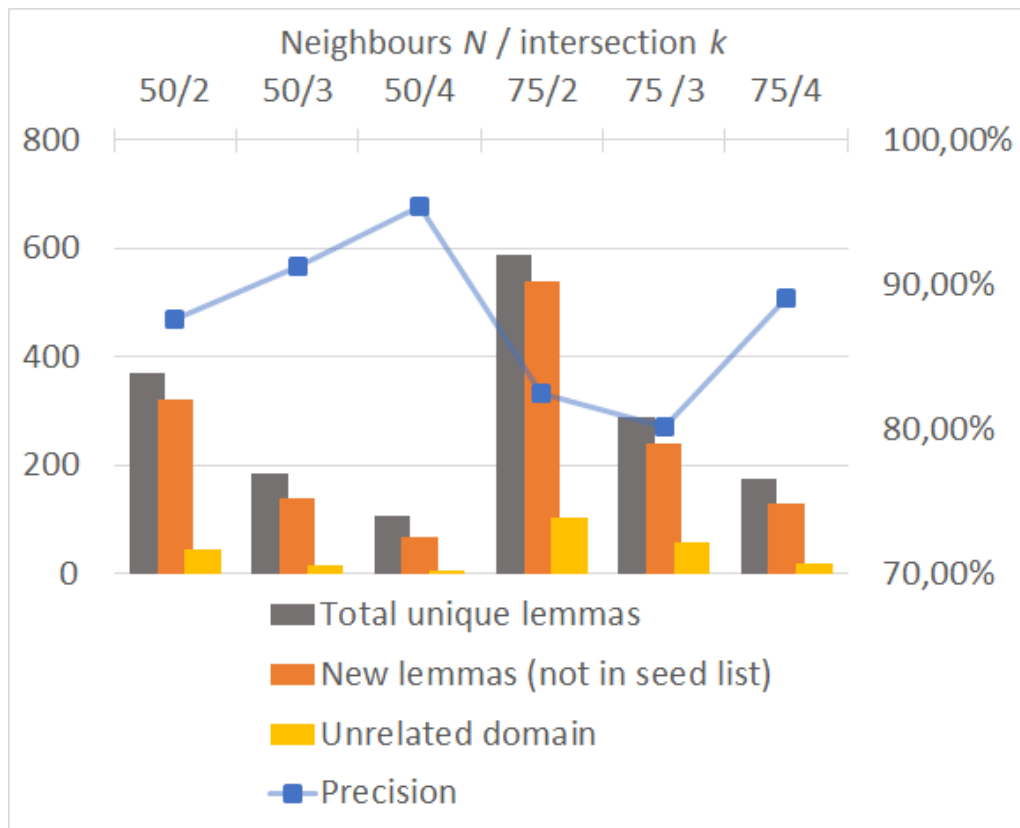
- Additional expressions forming “metaphor clusters” (Koller, 2003)

*fronta, strategija, preboj, akcijski načrt, vojna mentaliteta, sovražnik*  
[front, strategy, breakthrough, action plan, war mentality, enemy]

***Bitka** proti virusu na več **frontah***

[**Battle** against the virus on multiple **fronts**]

# DIFFERENT PARAMETER SETTINGS



$N=50, k=2$  OR  $N=75, k=3$

*vojen, sovražnik*

*[war [ADJ], enemy]*

Some WAR expressions nowhere in the first 200  $N$ :

*fronta, strategija, preboj, akcijski*

*[front, strategy, breakthrough, action [ADJ]]*

Increasing  $N$  ? > more words unrelated to WAR

# CONCLUSION

- † Simple method for vocabulary extension and top-down search for a known conceptual metaphor
  - some metaphors not found
  - still requires some manual analysis
  
- Further research to compare different word embeddings, larger expansion and better filtering

# Acknowledgements

- This work is supported by the Slovenian Research Agency by the research core funding P6-0215 and P2-0103, as well as by the research project CANDAS (J6-2581). The work has also been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media).



# REFERENCES

- Koller, V. (2003). Metaphor Clusters, Metaphor Chains: Analyzing the Multifunctionality of Metaphor in Text. *Metaphorik.de*, 5, pp. 115-134.
- Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.
- Juršič, M., Mozetič, I., Erjavec, T., and Lavrač, N. (2010). LemmaGen: Multilingual Lemmatisation with Induced Ripple-Down Rules. *Journal of Universal Computer Science*, 16(9), pp. 1190-1214. <https://doi.org/10.3217/jucs-016-09-1190>.
- Antloga, Š. (2020). Metaphor corpus KOMET 1.0, Slovenian language resource repository CLARIN.SI. Retrieved from <http://hdl.handle.net/11356/1293>.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, pp. 135-146.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. Chicago/London: University of Chicago Press.
- Vintar, Š., Grčić Simeunović, L., Martinc, M., Pollak, & S., Stepišnik, U. (2020). Mining Semantic Relations from Comparable Corpora through Intersections of Word Embeddings. *Proceedings of the 13th Workshop on Building and Using Comparable Corpora*, pp. 29-34.
- Rehurek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45-50. <https://doi.org/10.13140/2.1.2393.1847>.
- Ljubešič, N., & Erjavec, T. (2018). Word embeddings CLARIN.SI-embed.sl 1.0, Slovenian language resource repository CLARIN.SI. Retrieved from <http://hdl.handle.net/11356/1204>.