

Semantic Domains of Fire and Ice in Latvian
20th-Century Poetry:
Comparison of Manual Annotation of
Metaphors and Word2Vec Word Associations

Anda Baklāne, National Library of Latvia

Valdis Saulespurēns, National Library of Latvia

Shifts in language and culture: computational approaches to variation and change

October 5-7, 2022, Tallinn University, Estonia





Context

- Digital research services at the National library of Latvia: exploration with the aim to develop new digital services for the National Digital Library of Latvia
- Creating ideas, "sketches" for research projects; collaborative research projects
- Word2Vec explored mostly as a tool to identify various categories of entities in literary works (e.g. vehicles, occupation etc.)

* Data and support for text analysis is provided within collaborative projects: contact us dh@lnb.lv

<https://lndb.lv/>

<https://gramatas.lndb.lv/>

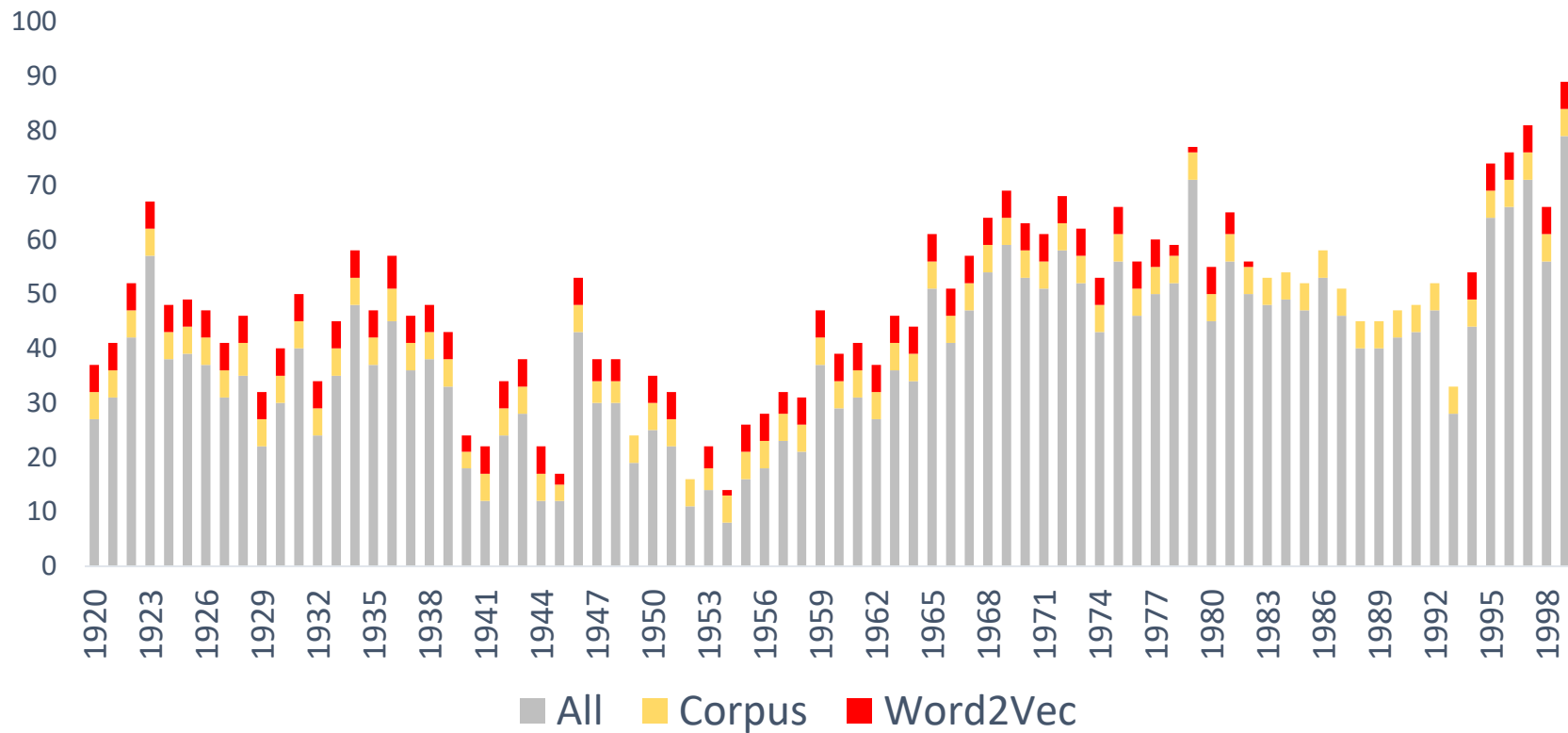
<https://periodika.lndb.lv/>

<https://nosketch.lnb.lv/>

<https://proza.lnb.lv/>

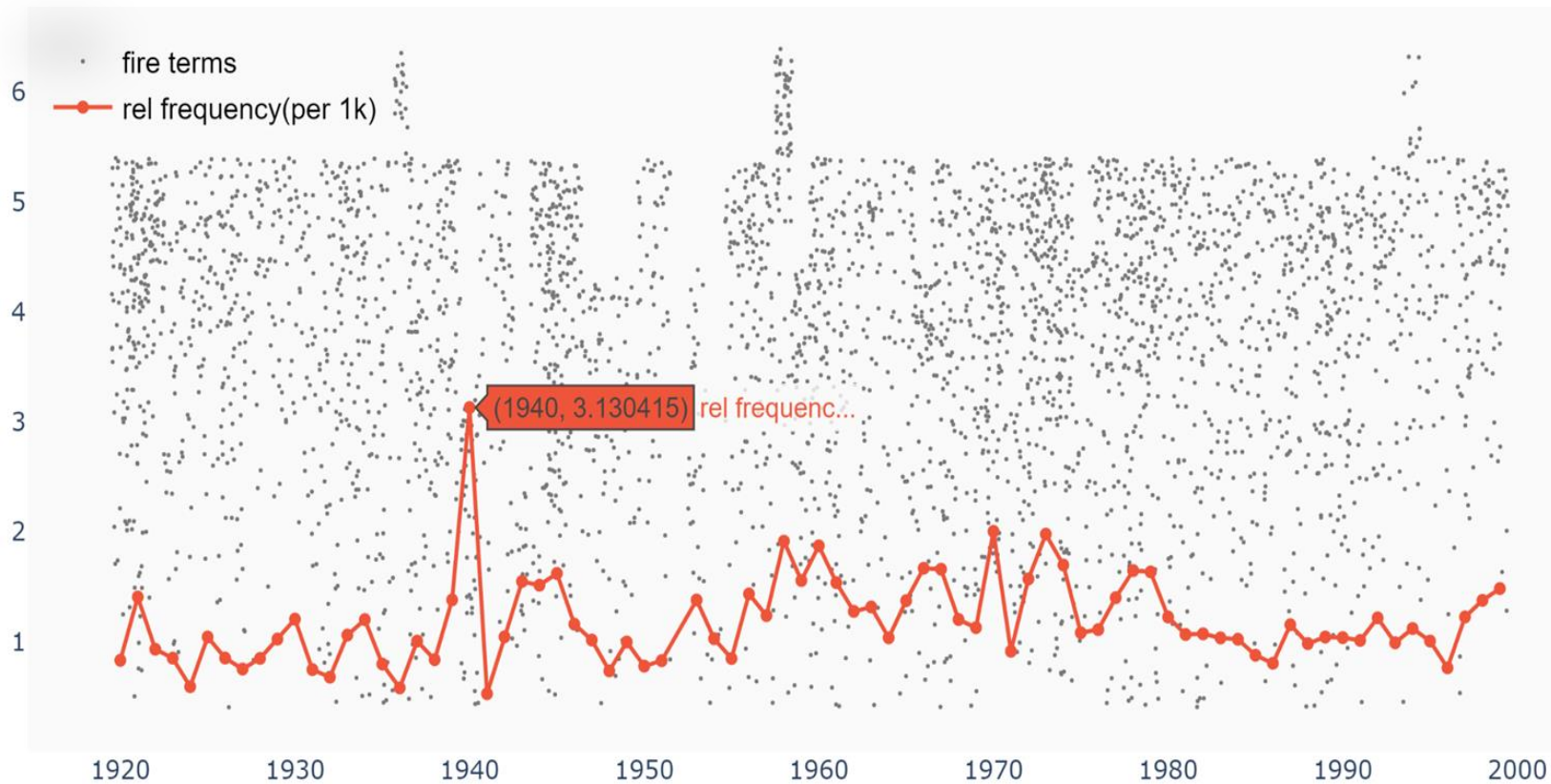


Data: number of documents, Latvian poetry



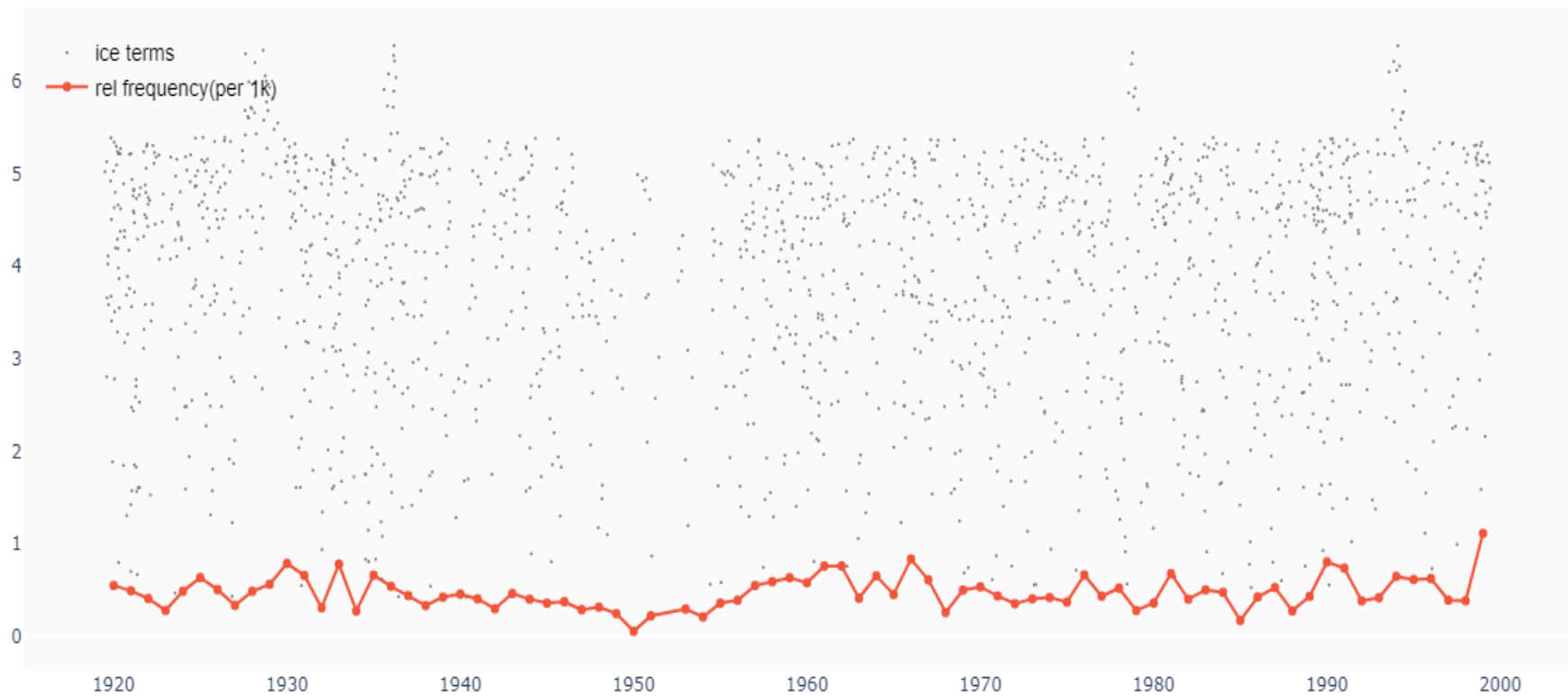


Term frequency: character string “fire”





Term frequency: character string “ice”





Metaphor

- Metaphor is understanding one thing in terms of another
- A conceptual metaphor is a frequently used, very productive domain of metaphors that can be manifested as a network of semantically interrelated terms
- E.g., LIFE IS A JOURNEY, UNDERSTANDING IS SEEING, SAD IS DOWN, **LOVE IS FIRE**
- **Source domain** is a vocabulary that is used to express metaphoric meaning
- **Target domain** is the meaning a speaker aims to convey



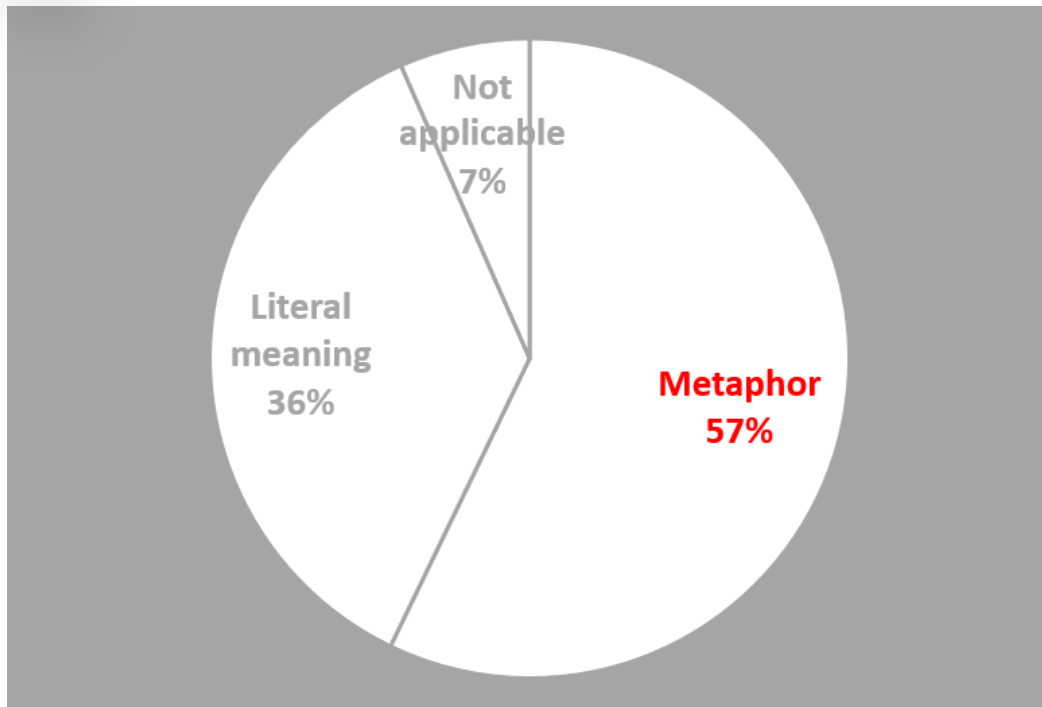
Poster: *In the fire*, opera by J. Kalniņš
Fire = source domain; *feelings, love, tribulation* = target domain



Book cover: *Burned lives*, a novel by Elita
Fire = source domain; *ruined lives* = target domain



“Fire”: proportion of metaphor and literal meaning*



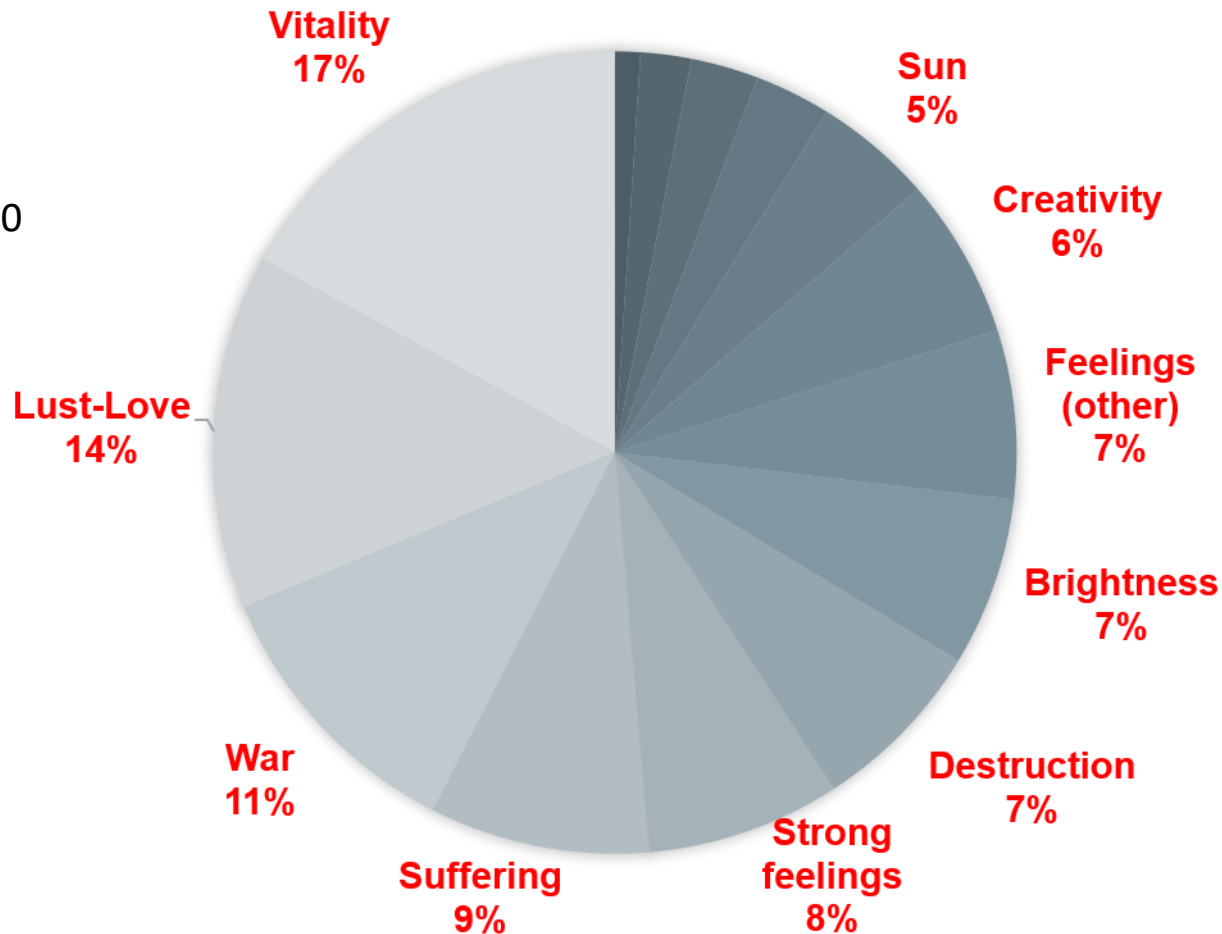
**Subjective annotation of 520 random examples in the corpus of Latvian poetry.*



Categories of the target domain of the term "fire" (subjective annotation of 520 examples)

LUST IS FIRE
LOVE IS FIRE
LIFE IS FIRE
VITALITY IS FIRE
WAR IS FIRE
SUFFERING IS FIRE
STRONG FEELINGS ARE FIRE
DESTRUCTION IS FIRE
CREATIVITY IS FIRE

...



Categories of the target domain of the term "ice"

- DEATH IS FRIGIDNESS / ICE
- LACK OF FREEDOM IS / ICE
- REJECTION IS COLDNESS / ICE
- CALLOUSNESS IS COLDNESS / ICE
- ...



- Often part of weather phenomena and natural landscape
- Rather part of an allegory, not an individual metaphor
- Difficult to evaluate with certainty if a natural scene is an allegory or just an artistic depiction of nature



Metaphor detection

- Identifying the source domain
is a given term a metaphor?
- Identifying the target domain
what the term signifies?
is the signal of the target domain present in the text?

Soft approach: digital analysis as supplementary method for qualitative analysis. The aim is exploration, not obtaining hard evidence (risks of unjustified usage, pattern seeking, confirmation bias, etc.)

Optimistic outlook: development of the method, gradual improvement of the procedure, improvement of the results ("from the few light sensitive cells to the fully developed eye")

Supervised machine learning methods can be used for training models based on preliminary manual annotation, however,

subjective annotation of the target domain is problematic:

- interpretation is subjective;
- **arbitrary names of categories** (the name of the category is not present in the text);
- the source domain can signify several target domains simultaneously;
- the results of the annotation often not suitable for mapping to computationally obtained models.



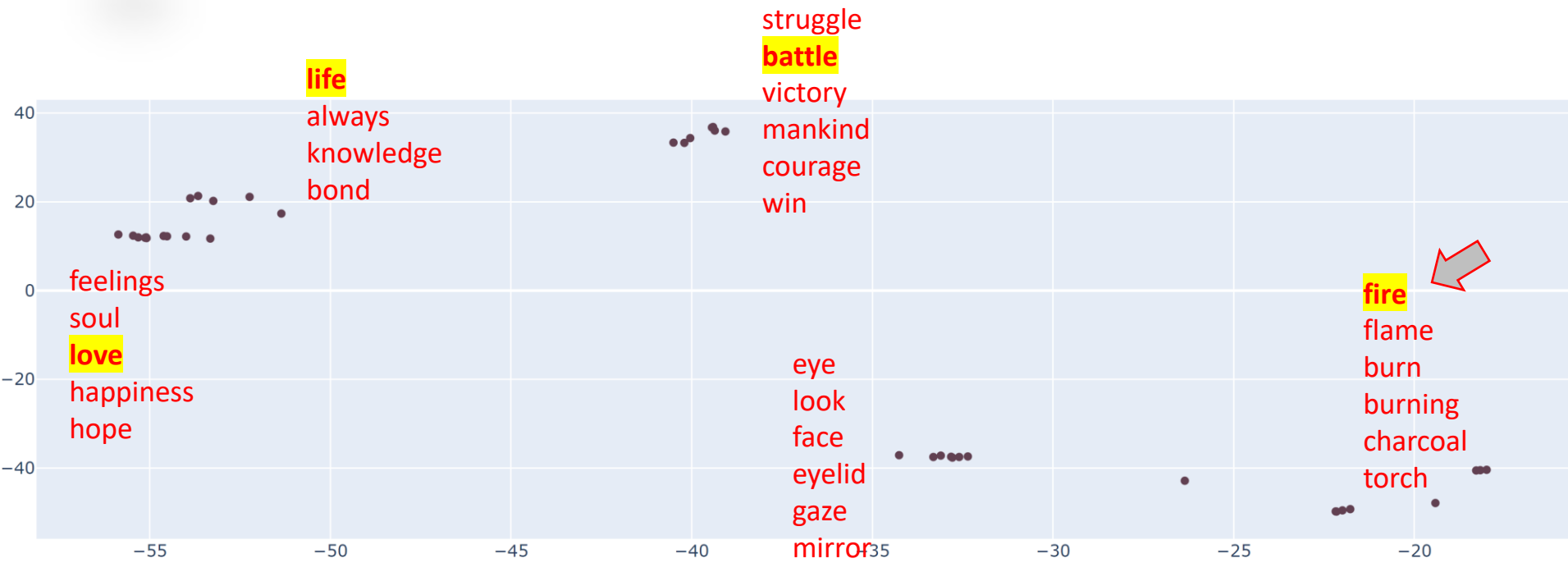
Word2Vec –

algorithm uses a neural network model to learn word associations from a large corpus of text.

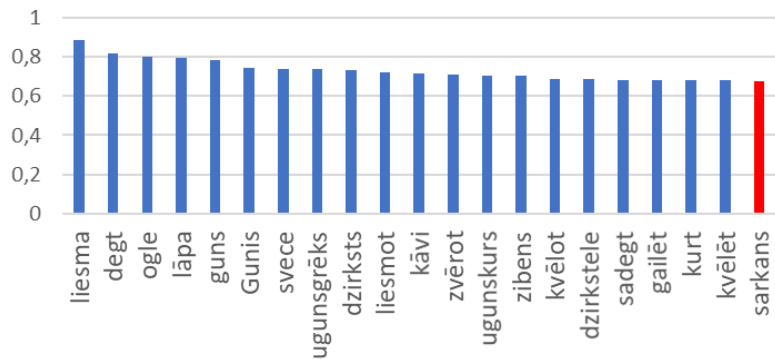
First published	2013
Embedding Results	word vectors
Resulting context	context free
Neural Network	Shallow
Training	Fast(CPU is fine)
Training type	Unsupervised
Popular Python library	Gensim



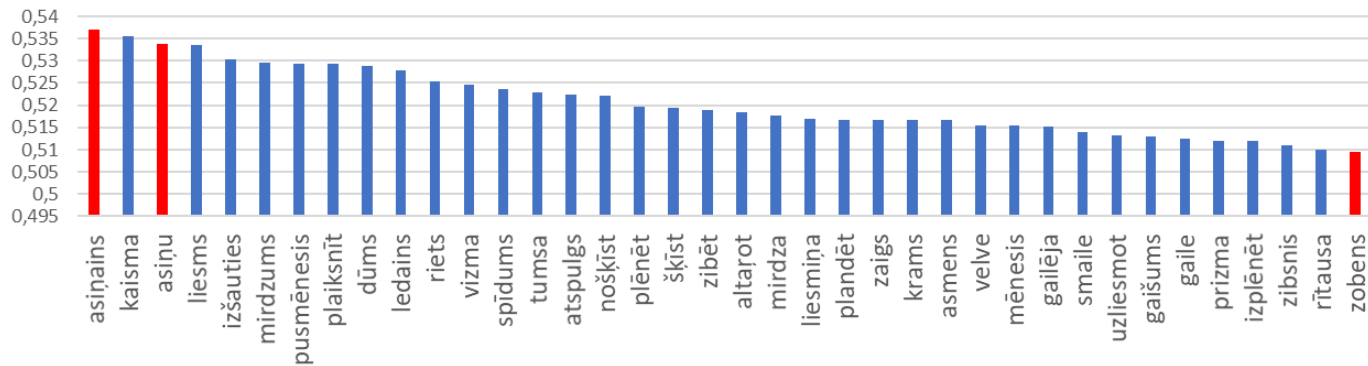
How close are the target domains to the source domain in the vector space: Word2Vec



Words closest to the term "fire" according to Word2Vec



No.10 – *red*; other close terms directly refer to burning



No.93 – *bloody*;

No.95 – *blood*;

No.130 – *sword*;

...

...

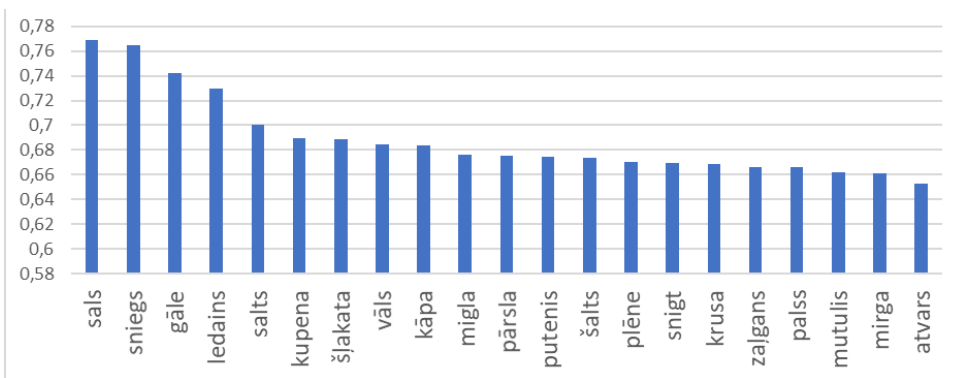
No.191 – *passion*

No.214 – *hate*

No.406 – *lust*

No.409 – *bliss*

Words closest to the term "ice" according to Word2Vec



No.18 –
whirlpool;

No.20 –
maelstrom;

other close
terms refer to
cold weather
events

* There are no terms in
the Top 500 directly
referring to death,
getting free, or
loneliness.



Conclusion

- In the current form, Word2Vec-processed poetry corpus suitable for exploratory analysis
- Not suitable for obtaining "hard evidence" on metaphor unless more robust methodology is developed
- The signal of the target domain of the metaphor may not be detectable as lexical entity since it is not explicitly formulated in the text, rather implied.

