ON THE IMPACT OF TEMPORAL REPRESENTATIONS ON METAPHOR DETECTION

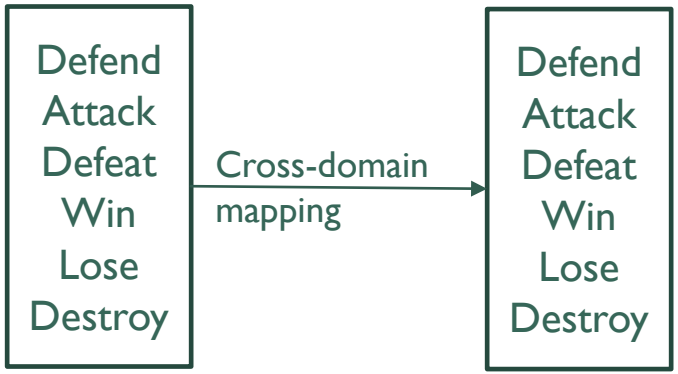
Giorgio Ottolina, Matteo Palmonari,
Manuel Vimercati, Mehwish Alam
Language Resources and Evaluation Conference
(LREC 2022)

ROADMAP

- Why Metaphors?
- Related Studies
- Methodology
- Experimentation
- Conclusion and Outlook

WHY METAPHORS ARE DIFFICULT?

ARGUMENT IS WAR



Sweet Person

Crime is a disease

Love is a journey

He Hulked out!!

SOURCE: WAR

TARGET: ARGUMENT

In my opinion this is a cold answer

LAKOFF AND JOHNSON, METAPHORS WE LIVE BY, 1980.

EXISTING STUDIES

- [Mao & al.2019]: *Different approaches*: Word's contextual vs literal meanings / Target word vs its context
- [Dankers & al.2020]: fine-tuned BERT model, fed with a discourse fragment as input. Hierarchical attention computes both token and sentence level attention after the encoded layers
- [Rai & al.2020]**:** dedicated survey on methods for metaphor detection
- Research works on time, language evolution and their impact on metaphors, but **no systematic studies yet**

HOW DOES TIME AFFECT METAPHORS?

- **Time —> new expressions —> used metaphorically —> stay metaphors: ``*Hulk out!*”**, meaning become enraged. It came into existence with the comic “The incredible Hulk” in 1962
- **Time —> new expressions —> used metaphorically —> becomes literal meaning: ``*Front runner*"**, also frontrunner, of political candidates, 1908, American English, a metaphor from horse racing (where it is used by 1901 of a horse that runs best while in the lead)
- ``The **virus attacked** Argonne National Laboratory outside Chicago starting at 11.54 pm EST Wednesday and throughout the night": ``**virus"** in this sentence is a computer one, used metaphorically along with ``**attacked"**

OTTOLINA ET AL. "ON THE IMPACT OF TEMPORAL REPRESENTATIONS ON METAPHORS"

LITERAL VS. METAPHORICAL MEANINGS CONTRAST CAN REVEAL METAPHORS' USAGE

- The contrast between words' literal and contextual meanings can highlight the presence of metaphors
- Temporal word embeddings may provide representations of words' core meaning too close to their metaphorical meaning, thus confusing the classifiers

OTTOLINA ET AL. "ON THE IMPACT OF TEMPORAL REPRESENTATIONS ON METAPHORS"

THE GOAL OF THIS EXPERIMENTAL STUDY

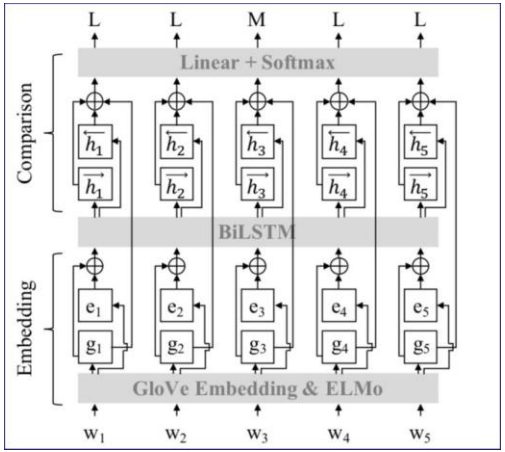
- **RQ1:** Does the use of temporal representations of literal meaning have an impact on metaphor detection, if measured on well-known benchmarks for the task? **YES**
- **RQ2:** Are certain temporal representation methods better than others at supporting metaphor detection? **NO**

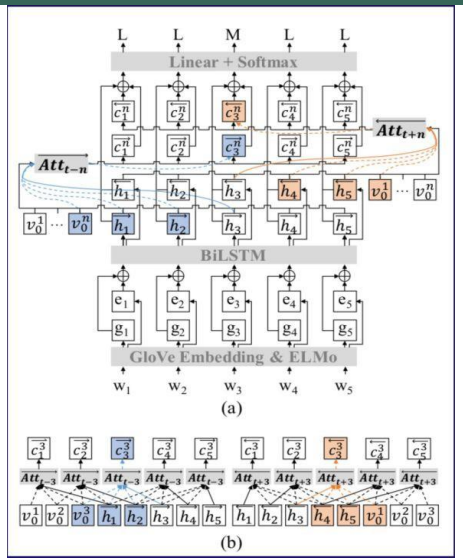
METHODOLOGY

- Explorative analysis based on word embeddings
- Quantitative analysis → impact of embeddings on metaphor detection
- Qualitative analysis → model's predictions
- New hypotheses to improve metaphor detection

MODEL SELECTION

MIP Procedure:
Classification by contrast between a contextual and literal meaning of a word





SPV procedure:
Metaphorical label prediction conditioned on a hidden state of a target word and its attentive context representation

[Mao & al.2019] - *End-to-End Sequential Metaphor Identification Inspired by Linguistic Theories* - Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics

RNN Hidden Glove

RNN Multi-Head Context Attention

SELECTING TEMPORAL WORD EMBEDDINGS

- Temporal word embeddings are trained with different approaches and corpora to account for several variables: corpus, word embedding algorithm, alignment method
- **HistWords – SGNS (*https://stanford.io/3txN0Hd*):** set of pre-trained temporal word embeddings generated using the Skip-gram variant of Word2vec and trained with negative sampling on different sliced diachronic corpora: 42 models based on **Procrustes alignment method** [Grave & al.2018]
- **CADE - Compass Aligned Embeddings**: temporal word embeddings trained with Word2vec and aligned with the **Compass alignment method (**[Di Carlo & al.2019] - [Bianchi & al.2020])

EXPERIMENTATION - DATASETS

Dataset	#sentences	Avg. length of sentence	%Metaphors	#unique verbs	Clear Temporal Connotation
MOH-X	646	8.0	49%	214	No
VUAsequence	6323	24.5	28%	2047	Yes (1985-1994)
TroFi	3737	28.3	43%	50	Yes (1987-1989)

RESULTS ON MOH-X DATASET

GloVe

Wikipedia > GloVe;
Precision > Recall

CoHa Word → best **SGNS** slices;
generally **better than GloVe**;
Recall > Precision

No clear temporal connotation, from wordnet (from 199x and by lexicographers);
Recent slices a bit better

CADE = other embeddings;
MOH-X → best dataset

			Metrics and Scores			
Main Corpus	Alignment	Slice	Precision	Recall	F1 Score	Accuracy
Common Crawl GloVe	NA	All	0.77	0.81	0.78	0.79
Wikipedia CBOW	NA	All	**0.81**	0.79	0.80	**0.81**
CoHa Word SGNS	Procrustes	1900	0.79	0.81	0.80	0.80
CoHa Word SGNS	Procrustes	1910	0.77	0.83	0.80	0.79
CoHa Word SGNS	Procrustes	1920	0.79	0.80	0.79	0.80
CoHa Word SGNS	Procrustes	1930	0.78	0.81	0.79	0.79
CoHa Word SGNS	Procrustes	1940	0.79	0.81	0.80	0.80
CoHa Word SGNS	Procrustes	1950	**0.81**	0.80	0.80	**0.81**
CoHa Word SGNS	Procrustes	1960	0.79	0.80	0.80	0.80
CoHa Word SGNS	Procrustes	1970	0.80	0.80	0.80	0.80
CoHa Word SGNS	Procrustes	1980	0.78	0.81	0.79	0.80
CoHa Word SGNS	Procrustes	1990	0.78	0.80	0.79	0.79
CoHa Word SGNS	Procrustes	2000	0.80	0.82	**0.81**	**0.81**
CoHa Lemma SGNS	Procrustes	1900	0.79	0.81	0.80	0.80
CoHa Lemma SGNS	Procrustes	1950	0.77	0.81	0.79	0.79
CoHa Lemma SGNS	Procrustes	1990	0.76	0.83	0.79	0.79
NGrams English All	Procrustes	1900	0.78	0.81	0.79	0.80
NGrams English All	Procrustes	1950	0.80	0.78	0.79	0.80
NGrams English All	Procrustes	1990	0.76	**0.84**	0.80	0.79
NGrams English Fiction	Procrustes	1900	0.77	0.82	0.79	0.79
NGrams English Fiction	Procrustes	1950	0.77	0.82	0.79	0.79
NGrams English Fiction	Procrustes	1990	0.80	0.81	0.80	**0.81**
Full CoHa CBOW	Compass	All	**0.81**	0.80	0.79	0.78
CoHa Word CBOW	Compass	1900	0.77	0.81	0.79	0.79
CoHa Word CBOW	Compass	1910	0.80	0.78	0.78	0.79
CoHa Word CBOW	Compass	1920	0.78	0.79	0.78	0.79
CoHa Word CBOW	Compass	1930	0.78	0.81	0.80	0.80
CoHa Word CBOW	Compass	1940	**0.81**	0.78	0.79	0.80
CoHa Word CBOW	Compass	1950	0.77	0.80	0.78	0.78
CoHa Word CBOW	Compass	1960	0.79	0.80	0.79	0.80
CoHa Word CBOW	Compass	1970	0.78	0.81	0.79	0.79
CoHa Word CBOW	Compass	1980	0.79	0.81	0.80	0.80
CoHa Word CBOW	Compass	1990	0.80	0.79	0.79	0.80
CoHa Word CBOW	Compass	2000	0.79	0.79	0.78	0.79

RESULTS ON VUA DATASET

GloVe

Wikipedia < Temporal

CoHa Lemma → best **SGNS** slices; Precision > SOTA; Recall drops

From BNC (1985 - 1994) - four genres - algorithm-annotated statements; middle slices are a bit better

Almost on par with Wikipedia and SGNS; Procrustes > **CADE**

			Metrics and Scores			
Main Corpus	**Alignment**	**Slice**	**Precision**	**Recall**	**F1 Score**	**Accuracy**
Common Crawl GloVe	NA	All	0.72	0.76	**0.74**	0.93
Wikipedia CBOW	NA	All	0.75	0.69	0.72	0.93
CoHa Word SGNS	Procrustes	1900	0.76	0.72	**0.74**	**0.94**
CoHa Word SGNS	Procrustes	1950	0.76	0.71	0.73	**0.94**
CoHa Word SGNS	Procrustes	1990	0.76	0.71	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1900	**0.77**	0.70	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1910	0.76	0.71	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1920	0.76	0.70	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1930	**0.77**	0.70	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1940	**0.77**	0.68	0.72	**0.94**
CoHa Lemma SGNS	Procrustes	1950	**0.77**	0.69	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1960	0.75	0.73	**0.74**	**0.94**
CoHa Lemma SGNS	Procrustes	1970	0.76	0.70	0.73	0.93
CoHa Lemma SGNS	Procrustes	1980	0.76	0.71	0.73	**0.94**
CoHa Lemma SGNS	Procrustes	1990	**0.77**	0.71	**0.74**	**0.94**
CoHa Lemma SGNS	Procrustes	2000	0.76	0.70	0.73	**0.94**
NGrams English All	Procrustes	1900	**0.77**	0.69	0.73	**0.94**
NGrams English All	Procrustes	1950	0.74	0.74	**0.74**	**0.94**
NGrams English All	Procrustes	1990	**0.77**	0.70	0.73	**0.94**
NGrams English Fiction	Procrustes	1900	0.75	0.71	0.73	**0.94**
NGrams English Fiction	Procrustes	1950	0.75	0.73	**0.74**	**0.94**
NGrams English Fiction	Procrustes	1990	0.76	0.70	0.73	**0.94**
Full CoHa CBOW	Compass	All	0.70	**0.80**	0.74	**0.94**
CoHa Word CBOW	Compass	1900	0.73	0.67	0.70	0.93
CoHa Word CBOW	Compass	1910	0.70	0.69	0.69	0.92
CoHa Word CBOW	Compass	1920	0.72	0.69	0.70	0.93
CoHa Word CBOW	Compass	1930	0.72	0.68	0.70	0.93
CoHa Word CBOW	Compass	1940	0.51	**0.80**	0.63	0.88
CoHa Word CBOW	Compass	1950	0.74	0.67	0.70	0.93
CoHa Word CBOW	Compass	1960	0.73	0.67	0.70	0.93
CoHa Word CBOW	Compass	1970	0.72	0.68	0.70	0.93
CoHa Word CBOW	Compass	1980	0.72	0.66	0.69	0.93
CoHa Word CBOW	Compass	1990	0.68	0.76	0.72	0.91
CoHa Word CBOW	Compass	2000	0.69	0.72	0.70	0.92

RESULTS ON TROFI DATASET

GloVe

Wikipedia < Temporal

From 87-'89 Wall Street Journal Corpus (WSJ); old middle seem to have a better balance

English All → best **SGNS** slices; **Precision increases with time and recall drops, except for English All SGNS 1910**

Generally better than SGNS; F1 scores from 71% to 72%; **CADE** > Procrustes

			Metrics and Scores			
Main Corpus	**Alignment**	**Slice**	**Precision**	**Recall**	**F1 Score**	**Accuracy**
Common Crawl GloVe	NA	All	0.68	0.76	0.71	0.74
Wikipedia CBOW	NA	All	0.70	0.71	0.71	0.74
CoHa Word SGNS	Procrustes	1900	0.69	0.73	0.71	0.74
CoHa Word SGNS	Procrustes	1950	0.69	0.74	0.71	0.74
CoHa Word SGNS	Procrustes	1990	0.70	0.72	0.71	0.74
CoHa Lemma SGNS	Procrustes	1900	0.69	0.73	0.71	0.74
CoHa Lemma SGNS	Procrustes	1950	0.69	0.73	0.71	0.74
CoHa Lemma SGNS	Procrustes	1990	0.70	0.72	0.71	0.74
NGrams English All	Procrustes	1900	0.71	0.71	0.71	**0.75**
NGrams English All	Procrustes	1910	**0.72**	0.70	0.71	**0.75**
NGrams English All	Procrustes	1920	0.70	0.72	0.71	0.74
NGrams English All	Procrustes	1930	0.70	0.71	0.71	0.74
NGrams English All	Procrustes	1940	0.71	0.71	0.71	**0.75**
NGrams English All	Procrustes	1950	0.68	0.75	0.71	0.74
NGrams English All	Procrustes	1960	0.69	0.73	0.71	0.74
NGrams English All	Procrustes	1970	0.70	0.72	0.71	0.74
NGrams English All	Procrustes	1980	0.71	0.72	0.71	0.74
NGrams English All	Procrustes	1990	0.70	0.73	0.71	0.74
NGrams English Fiction	Procrustes	1900	0.69	0.73	0.71	0.74
NGrams English Fiction	Procrustes	1950	0.68	0.75	0.71	0.73
NGrams English Fiction	Procrustes	1990	0.70	0.73	0.71	0.74
Full CoHa CBOW	Compass	All	**0.72**	0.76	**0.74**	0.74
CoHa Word CBOW	Compass	1900	0.69	0.77	0.72	**0.75**
CoHa Word CBOW	Compass	1910	0.69	0.76	0.72	0.74
CoHa Word CBOW	Compass	1920	0.70	0.75	0.72	**0.75**
CoHa Word CBOW	Compass	1930	0.68	0.77	0.72	0.74
CoHa Word CBOW	Compass	1940	0.69	0.76	0.72	0.74
CoHa Word CBOW	Compass	1950	0.68	0.77	0.72	0.74
CoHa Word CBOW	Compass	1960	0.68	0.78	0.72	0.74
CoHa Word CBOW	Compass	1970	0.68	0.77	0.72	0.74
CoHa Word CBOW	Compass	1980	0.68	0.78	0.73	**0.75**
CoHa Word CBOW	Compass	1990	0.70	**0.80**	0.73	0.74
CoHa Word CBOW	Compass	2000	0.69	0.76	0.72	**0.75**

ANALYSIS OF THE RESULTS

- Temporal representations (i.e., more specific) seem to achieve better results
- Representations related to different time intervals and trained with different corpora and alignment methods perform slightly better / worse in different datasets (no clear superiority)

QUALITATIVE ANALYSIS

"She wanted to buy his love with her dedication to him and his work"

Core meanings used in economical/political and emotional/feelings contexts. (MOH-X)

"We're not attaching the core assets; we're looking at what we consider to be our less – profitable assets," the spokesman said

Physical (core meaning) verbs → metaphorical in economical and emotional contexts. (TroFi)

*The **virus attacked** Argonne National Laboratory outside Chicago starting at 11:54 pm EST Wednesday and throughout the night*

Nearest Neighbors Analysis:
Comparison of target words' meanings in different embeddings
(e.g. non-temporal vs temporal)

Temporal word embeddings → models classify ***virus attacked*** as **metaphorical**

QUALITATIVE ANALYSIS

- Temporal representations sometimes are even too close to metaphorical meanings
- Metaphor detection datasets do not provide representative enough metaphors
- Economical, political and emotional core meanings → most frequently identified
- CoHa (SGNS) 1990: no `news' genre correct predictions → bias towards metaphorical meaning

CONCLUSION & OUTLOOK

- First attempt to investigate the **interaction between metaphorical word usage and semantic change** using computational metaphor detection methods and corpus-specific word embeddings
- If temporal representations are too close to metaphorical meanings, less metaphors are identified
- There are some weak patterns of some approaches performing slightly better than others on a specific dataset, but no strong conclusion can be drawn from the experiments
- Future work:
  - collect more data - metaphorical word usage and temporal semantic change

THANK YOU FOR YOUR ATTENTION!

CONTACT US

GIORGIO OTTOLINA: (✉: g.ottolina l @campus.unimib.it, giorgio.ottolina@gmail.com)

MATTEO PALMONARI: (✉: matteo.palmonari@unimib.it, : @MatteoPalmonari)

MANUEL VIMERCATI: (✉: manuel.vimercati@unimib.it, : @ManuelVimercati)

MEHWISH ALAM: (✉ : mehwish.alam@kit.edu, : @em_alam)