



ON THE IMPACT OF DIFFERENT WORD EMBEDDINGS ON METAPHOR DETECTION

Giorgio Ottolina - 838017
Laurea Magistrale in Data Science

She devoured his **sandwiches**.
She devoured his **novels**.

} Anomalies

“devoured” means “**enjoyed avidly**”.

“devoured” and “enjoyed” are **different concepts** literally.

“devoured” is **metaphorical**.

Common sense

Interpretation

Identification

Conceptual Metaphor Theory (CMT) -Lakoff and Johnson - Metaphor:

not just property of language, but cognitive mechanism; knowledge projection process

Link between metaphors and language evolution

Metaphor Detection

1. In single sentence
2. In discourse

</>

‘The wheels of Stalin’s regime
were well-oiled and
already turning,’

– Lakoff and
Johnson, 1980:
“Metaphors we live by”

Metaphor Classification

Goal → identify all
metaphorical words
inside a sentence

Procedures

1. contrast between a word’s contextual and literal meanings
2. contrast between a target word and its context

WORD EMBEDDINGS

INTUITION



Vector representations computed
with different models on different
corpora

CORE MEANINGS



GloVe & Word2Vec

CONTEXTUAL



BERT & ELMO

Words' core meanings change over time → depend on corpus and time
Specific words' meanings can be learned using specific temporal slices

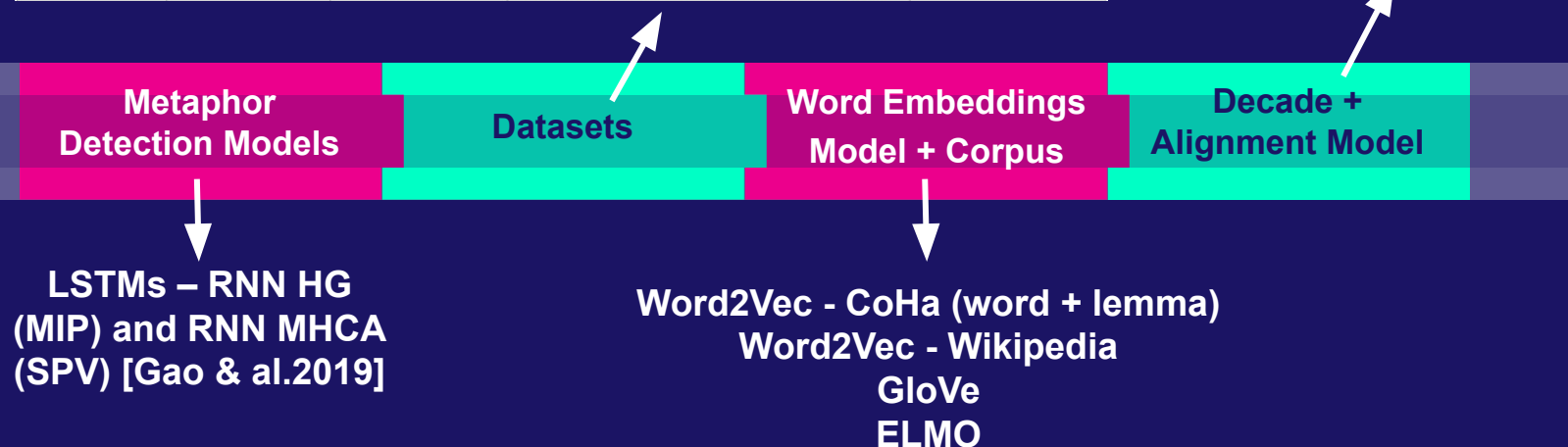
OVERVIEW

THESIS OBJECTIVES

1. **Explorative analysis** based on **word embeddings**
2. **Quantitative analysis** → **impact of embeddings on metaphor detection**
3. **Qualitative analysis** → **model's predictions**
4. **New hypotheses** to improve metaphor detection

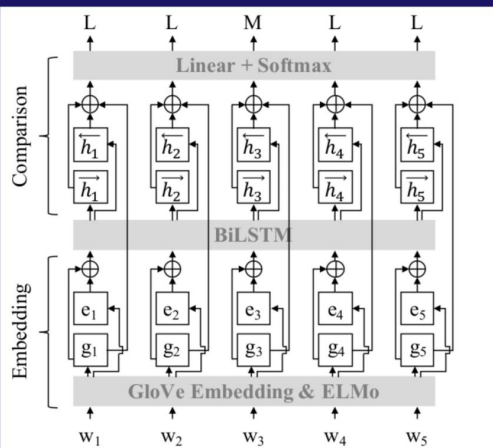
METHODOLOGY – INTRO

Dataset	N. of sentences	Train/test/split	How it was built	Clear temporal connotation
MOH-X	646	Cross-validation required	Derived from the subset of MOH dataset used by Ekaterina et. al. The verbs are annotated for metaphoricity and they come from WordNet.	No
VUAsquence	6323	Yes	117 fragments sampled across 4 genres from the British National Corpus (academic, news, conversation and fiction).	Yes 1985-1994
TroFi	3737	Cross-validation required	The sentences (each one with a single annotated target verb) are taken from the '87-'89 Wall Street Journal corpus (WSJ).	Yes 1987-1989



METHODOLOGY

RNN HG & RNN MHCA models - Rui Mao et al. (2019)
for Sequential Metaphor Identification:
RNN HG \rightarrow better with Temporal Embeddings

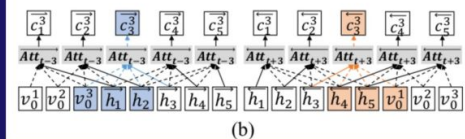
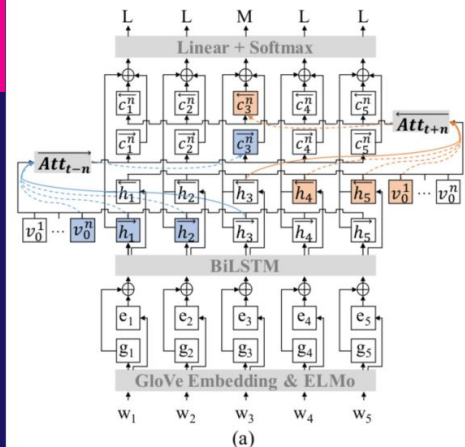


RNN Hidden Glove

MIP procedure:
metaphor is classified
by contrast between a
word's contextual
(hidden state) and literal
(GloVe) meanings

RNN Multi-Head Context Attention

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



GloVe

MOH-X

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
Wikipedia, Gigaword, Common Crawl*	NA	All	0.77	0.81	0.78	0.79
Wikipedia	NA	All	0.81	0.79	0.80	0.81
Full CoHa SGNS	Procrustes	All	0.79	0.81	0.79	0.80
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
Eng-all SGNS	Procrustes	1900	0.78	0.81	0.79	0.80
Eng-all SGNS	Procrustes	1950	0.80	0.78	0.79	0.80
Eng-all SGNS	Procrustes	1990	0.76	0.84	0.80	0.79
Eng-fiction	Procrustes	1900	0.77	0.82	0.79	0.79
Eng-fiction	Procrustes	1950	0.77	0.82	0.79	0.79
Eng-fiction	Procrustes	1990	0.80	0.81	0.80	0.81
Full CoHa Word	CADE – Compass	All	0.81	0.79	0.79	0.80
CoHa Corpus Slices	CADE – Compass	1900	0.77	0.80	0.78	0.79
CoHa Corpus Slices	CADE – Compass	1910	0.78	0.79	0.79	0.79
CoHa Corpus Slices	CADE – Compass	1920	0.78	0.80	0.78	0.79
CoHa Corpus Slices	CADE – Compass	1930	0.79	0.80	0.79	0.80
CoHa Corpus Slices	CADE – Compass	1940	0.80	0.78	0.79	0.80
CoHa Corpus Slices	CADE – Compass	1950	0.79	0.81	0.80	0.80
CoHa Corpus Slices	CADE – Compass	1960	0.77	0.81	0.79	0.79
CoHa Corpus Slices	CADE – Compass	1970	0.80	0.80	0.80	0.80
CoHa Corpus Slices	CADE – Compass	1980	0.79	0.79	0.79	0.79
CoHa Corpus Slices	CADE – Compass	1990	0.80	0.79	0.79	0.80
CoHa Corpus Slices	CADE – Compass	2000	0.78	0.82	0.79	0.80

GloVe

VUA

Wikipedia < Temporal

			Metrics and Scores			
Main Corpus	Alignment	Slice	Precision	Recall	F1 Score	Accuracy
Wikipedia, Gigaword, Common Crawl*	NA	All	0.72	0.76	0.74	0.93
Wikipedia	NA	All	0.75	0.69	0.72	0.93
Full CoHa SGNS	Procrustes	All	0.76	0.71	0.73	0.94
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	0.76	0.71	0.73	0.94
CoHa lemma SGNS	Procrustes	2000	0.76	0.70	0.73	0.94
Eng-all SGNS	Procrustes	1900	0.77	0.69	0.73	0.94
Eng-all SGNS	Procrustes	1950	0.74	0.74	0.74	0.94
Eng-all SGNS	Procrustes	1990	0.77	0.70	0.73	0.94
Eng-fiction	Procrustes	1900	0.75	0.71	0.73	0.94
Eng-fiction	Procrustes	1950	0.75	0.73	0.74	0.94
Eng-fiction	Procrustes	1990	0.76	0.70	0.73	0.94
Full CoHa Word	CADE – Compass	All	0.67	0.72	0.70	0.92
CoHa Corpus Slices	CADE – Compass	1900	0.76	0.71	0.74	0.94
CoHa Corpus Slices	CADE – Compass	1910	0.75	0.73	0.74	0.94
CoHa Corpus Slices	CADE – Compass	1920	0.72	0.74	0.73	0.93
CoHa Corpus Slices	CADE – Compass	1930	0.76	0.70	0.73	0.94
CoHa Corpus Slices	CADE – Compass	1940	0.74	0.74	0.74	0.94
CoHa Corpus Slices	CADE – Compass	1950	0.72	0.74	0.73	0.93
CoHa Corpus Slices	CADE – Compass	1960	0.75	0.72	0.73	0.94
CoHa Corpus Slices	CADE – Compass	1970	0.75	0.71	0.73	0.93
CoHa Corpus Slices	CADE – Compass	1980	0.73	0.75	0.74	0.93
CoHa Corpus Slices	CADE – Compass	1990	0.75	0.71	0.73	0.93
CoHa Corpus Slices	CADE – Compass	2000	0.76	0.71	0.73	0.94

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

GloVe

TROFI

Wikipedia < Temporal

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

Eng-All → best SGNS slices;
Precision increases with time
and recall drops, except for
Eng-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
Wikipedia, Gigaword, Common Crawl*	NA	All	0.68	0.76	0.71	0.74
Wikipedia	NA	All	0.70	0.71	0.71	0.74
Full CoHa SGNS	Procrustes	All	0.69	0.73	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
Eng-all SGNS	Procrustes	1900	0.71	0.71	0.71	0.75
Eng-all SGNS	Procrustes	1910	0.72	0.70	0.71	0.75
Eng-all SGNS	Procrustes	1920	0.70	0.72	0.71	0.74
Eng-all SGNS	Procrustes	1930	0.70	0.71	0.71	0.74
Eng-all SGNS	Procrustes	1940	0.71	0.71	0.71	0.75
Eng-all SGNS	Procrustes	1950	0.68	0.75	0.71	0.74
Eng-all SGNS	Procrustes	1960	0.69	0.73	0.71	0.74
Eng-all SGNS	Procrustes	1970	0.70	0.72	0.71	0.74
Eng-all SGNS	Procrustes	1980	0.71	0.72	0.71	0.74
Eng-all SGNS	Procrustes	1990	0.70	0.73	0.71	0.74
Eng-fiction	Procrustes	1900	0.69	0.73	0.71	0.74
Eng-fiction	Procrustes	1950	0.68	0.75	0.71	0.73
Eng-fiction	Procrustes	1990	0.70	0.73	0.71	0.74
Full CoHa Word	CADE – Compass	All	0.68	0.77	0.72	0.74
CoHa Corpus Slices	CADE – Compass	1900	0.70	0.74	0.72	0.74
CoHa Corpus Slices	CADE – Compass	1910	0.69	0.76	0.72	0.75
CoHa Corpus Slices	CADE – Compass	1920	0.69	0.76	0.72	0.75
CoHa Corpus Slices	CADE – Compass	1930	0.70	0.74	0.72	0.75
CoHa Corpus Slices	CADE – Compass	1940	0.69	0.75	0.72	0.74
CoHa Corpus Slices	CADE – Compass	1950	0.69	0.76	0.72	0.75
CoHa Corpus Slices	CADE – Compass	1960	0.69	0.75	0.72	0.74
CoHa Corpus Slices	CADE – Compass	1970	0.70	0.74	0.72	0.75
CoHa Corpus Slices	CADE – Compass	1980	0.69	0.76	0.72	0.74
CoHa Corpus Slices	CADE – Compass	1990	0.70	0.74	0.72	0.75
CoHa Corpus Slices	CADE – Compass	2000	0.69	0.75	0.72	0.75

OVERALL PERFORMANCES

RESULTS OVERVIEW

1. Word2Vec > GloVe
2. SGNS > GloVe
3. SGNS > Wikipedia (TroFi and VUA)
4. MOH-X → generally the best dataset
5. CADE alignment → better with TroFi dataset
6. Procrustes alignment → better with VUA dataset

QUALITATIVE ANALYSIS – 1

</>

`She wanted to buy his love with her dedication to him and his work.'

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

</>

'We're not attacking 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. ('attacking' - TroFi)

Nearest Neighbors Analysis:
Words most similar to predictions

QUALITATIVE ANALYSIS – 2

RESULTS OVERVIEW

1. Core meanings used in economical, political and emotional contexts → most frequently identified
2. Physical (core meanings) verbs → metaphorical if used in previous contexts
3. CoHa (SGNS) 1990 slice → no `news' genre correct predictions. Metaphorical representations bias?

CONCLUSIONS

**Contextual word embeddings (BERT) → fine-tuned model trained on state of the art datasets:
Metaphorical sentences classification on new test set (LCC dataset).
Core meanings → Economics, politics and emotions contexts**

1. **Word2Vec temporal word embeddings > GloVe with specific temporal ranges and slices**
2. **Core meanings/Contexts qualitative analysis patterns**
3. **Future research: new ad hoc dataset with words with known semantic changes & more BERT experiments**





**THANK YOU
FOR YOUR
ATTENTION!**

