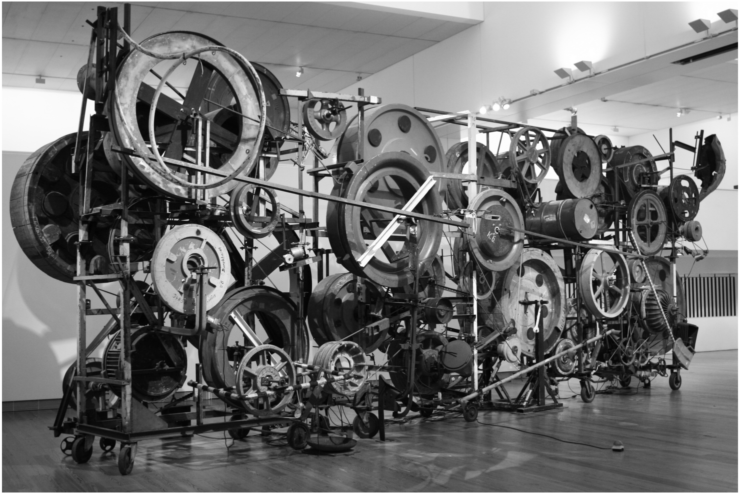
# Grasping the finer point: Metaphor identification in text and brain imaging data

**Ekaterina Shutova**

ILLC
University of Amsterdam

Stanford University, 23 August 2018

# What is metaphor?

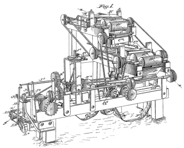


# How does metaphor work?

Association between two concepts

(Gentner, 1983; Lakoff and Johnson, 1980)

$\underbrace{\textit{POLITICALSYSTEM}}_{\textit{target}}$ is a $\underbrace{\textit{MECHANISM}}_{\textit{source}}$



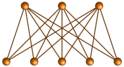
"*rebuilding* *the campaign* *machinery*"

"*Time to* *mend* *our foreign policy*"

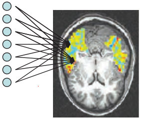
"*20 Steps towards a* *working* *democracy*"

# Today's talk

1. Metaphor identification method
   (Rei, Bulat, Kiela & Shutova, EMNLP 2017)



2. Using NLP techniques to study metaphor processing in the brain
   (Gamez-Gjokic, Maillard, Bulat & Shutova, forthcoming)



## Metaphor identification: Existing approaches

Linguistic resources:

- Semantic roles
  (Gedigian et al., 2006)
- Concreteness
  (Turney et al., 2011)
- Imageability
  (Strzalkowski et al., 2013)
- WordNet supersenses
  (Tsvetkov et al., 2014)

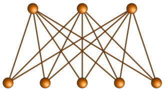
Data-driven methods & cognitive features:

- Clustering with sparse distributional features
  (Shutova et al., 2010)
- Visual vectors
  (Shutova et al., 2016)
- Attribute-based vectors
  (Bulat et al., 2017)

# A neural architecture for metaphor processing

*Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection*.

Rei, Bulat, Kiela & Shutova, EMNLP 2017.



- Supervised classification setting
- Identifying metaphorical uses of verbs and adjectives

**Mohammad et al. (2016)**

| Verb noun | Class |
|---|---|
| *boost* economy | met. |
| boost voltage | lit. |

**Tsvetkov et al. (2014)**

| Adj. noun | Class |
|---|---|
| *cloudy* future | met. |
| cloudy sky | lit. |

# Approach

INPUT: skip-gram word embeddings

- 100-dimensional
- trained on Wikipedia

OUTPUT: a metaphoricity score between 0 and 1

# Approach

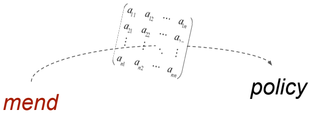
INPUT: skip-gram word embeddings

- 100-dimensional
- trained on Wikipedia

OUTPUT: a metaphoricity score between 0 and 1

Key intuitions:

1. model domain interaction via gating
2. specialise word representations
3. quantify metaphoricity via a weighted similarity function

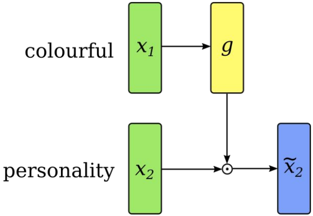


# Word representation gating

$$g = \sigma(W_g x_1)$$

$$\widetilde{x}_2 = x_2 \odot g$$

$W_g$ — a weight matrix
$\sigma$ — sigmoid activation function
$\odot$ — element-wise multiplication.

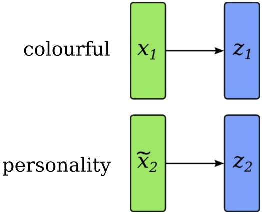



Some properties of the source domain are projected onto the target

# Specialisation

$$z_1 = tanh(W_{z_1} x_1)$$

$$z_2 = tanh(W_{z_2} \widetilde{x}_2)$$

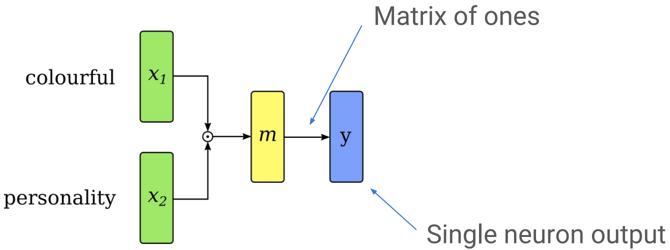


## Weighted similarity

If the input vectors $x_1$ and $x_2$ are normalised to unit length, the cosine similarity between them is equal to their dot product:

$$cos(x_1, x_2) \propto \sum_i x_{1,i} x_{2,i}$$

We can formulate this as a small neural network:

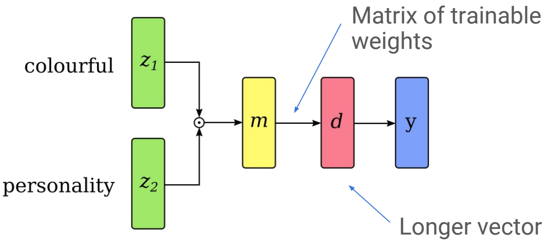


# Weighted similarity

We can instead create a version where vector *m* is passed through another layer, with weights that are optimised during training.

$$m_i = z_{1,i} z_{2,i}$$

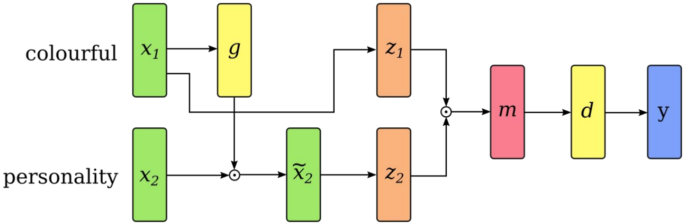
$$d = \gamma(W_d m)$$



## Supervised similarity network

The final network architecture, using:

- Word representation gating
- Specialisation
- Vector combination based on weighted cosine



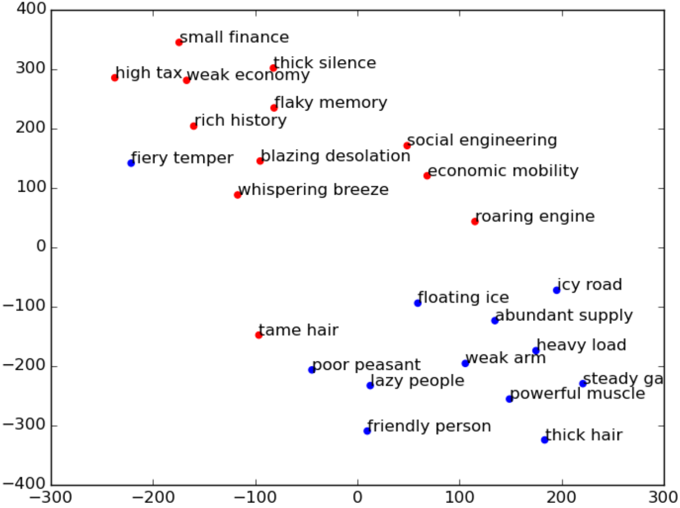
# Results: Adjectives

| | Acc | P | R | F1 |
|---|---|---|---|---|
| Tsvetkov et al. (2014) | - | - | - | **85** |
| Shutova et al. (2016) | | | | |
| linguistic | - | 73 | 80 | 76 |
| multimodal | - | 67 | **96** | 79 |
| Bulat et al. (2017) | - | 85 | 71 | 77 |
| FFN skip-gram | 77.6 | 86.6 | 65.4 | 74.4 |
| SSN skip-gram | **82.2** | **91.1** | **71.6** | **80.1** |

# Results: Verbs

| | Acc | P | R | F1 |
|---|---|---|---|---|
| Shutova et al. (2016) | | | | |
| linguistic | - | 67 | 76 | 71 |
| multimodal | - | 65 | **87** | **75** |
| FFN skip-gram | 71.2 | 70.4 | 71.8 | 70.5 |
| SSN skip-gram | **74.8** | **73.6** | **76.1** | **74.2** |

# Qualitative analysis

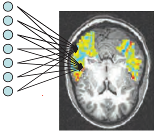


# Applications in social science (and beyond)

- Metaphor as a predictor of influence / popularity of politicians
- Vinod, Dan and I
- Facebook dataset
- The number of metaphors used can serve as a predictor of the number of shares, likes etc.
- Looking at the identity of the metaphors next

# Decoding literal and metaphorical sentences in the brain

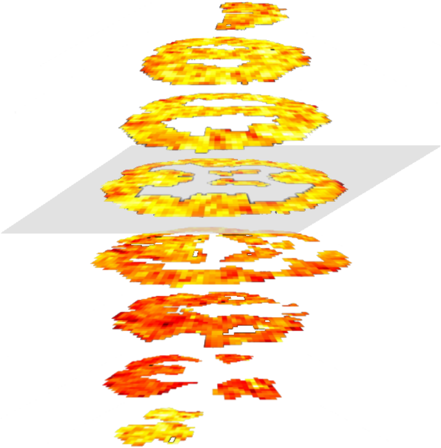
*Can we use semantic models to better understand metaphor processing in the brain?*
Gamez-Djokic, Maillard, Bulat and Shutova.



Experiments with brain imaging data

- Data: fMRI neural activation patterns associated with the meaning of literal and metaphorical sentences (Gamez-Djokic et al, forthcoming)
- Verbs in their metaphorical and literal contexts
- Task: decode patterns of brain activity
- Using data-driven semantic models

## functional Magnetic Resonance Imaging (fMRI)



- **Voxel**: a 3x3x6mm$^3$ cube of brain tissue
- **Voxel value**: intensity of brain activity in that voxel
- **fMRI image**: vector of voxel values (represents brain activation pattern)

## Our brain imaging dataset

- 15 participants
- 31 unique hand-action verbs
- 200 sentences
- 5 conditions

| **Condition** | **Sentence** |
|---|---|
| Affirmative Literal | She's *grasping* the cookie |
| Affirmative Metaphor | She's *grasping* the lecture |
| Negated Literal | He's not *grasping* the bill |
| Negated Metaphor | He's not *grasping* the problem |
| Affirmative Paraphrase | She's understanding the lecture |

## Stimuli presentation

- Disambiguation – object:

  *The physics lecture (2 seconds)*

- Interval:

  *(0.5 seconds)*

- Stimulus:

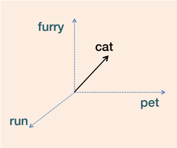
  *She is grasping the lecture (6 seconds)*

- Rest:

  *(8 seconds)*

# Semantic models

1. Linguistic models
    - Word representations
    - Compositional models




2. Visually grounded models
    - word and phrase representations
    - learned from images

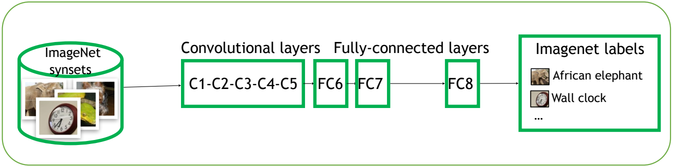
3. Multimodal models
    - combining linguistic and visual information

# Linguistic models

- Individual words: VERB and OBJECT
  - GloVe word embeddings (Pennington et al. 2014)
- VERBOBJECT: concatenation of verb and object embeddings
- ADDITION: addition of verb and object embeddings
- LSTM: learn representations for verb-object phrases
  - trained on the natural language inference task
  - taking Glove word embeddings as input

# Visual representations

1. retrieve images for a word or phrase using Google Search
2. transfer learning to extract image embeddings:
    - convolutional neural network trained on the ImageNet classification task (Kiela and Bottou, 2014)




    - forward pass
    - use penultimate layer (FC7) as image embedding

## Visual and multimodal models

Visual models:

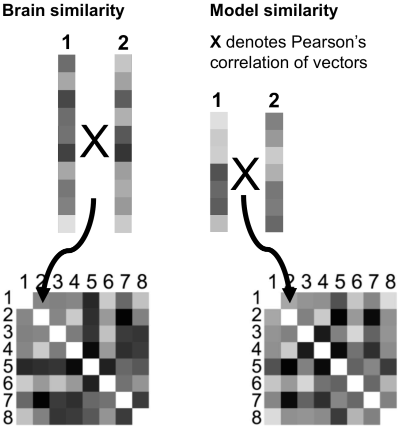
- Individual words: VERB and OBJECT
- VERBOBJECT: concatenation of verb and object embeddings
- ADDITION: addition of verb and object embeddings
- PHRASE: visual representation for the whole phrase

Multimodal models:

- Concatenation of the respective linguistic and visual models
- with the exception of LSTM

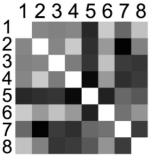
# Decoding brain activity

Similarity-based decoding (Anderson et al., 2016)

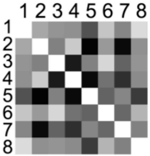



*Slide credit: Andrew Anderson*

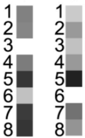
# Similarity-based decoding



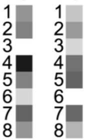

Brain similarity matrix




Model similarity matrix



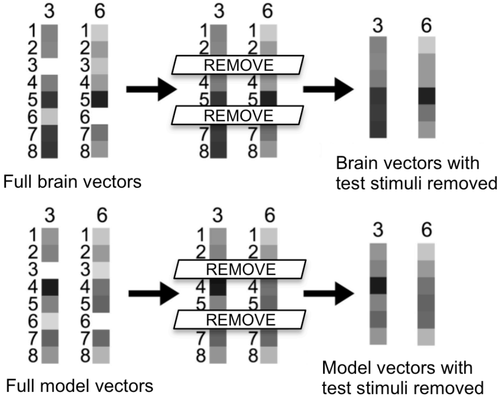

Brain similarity vectors




Model similarity vectors

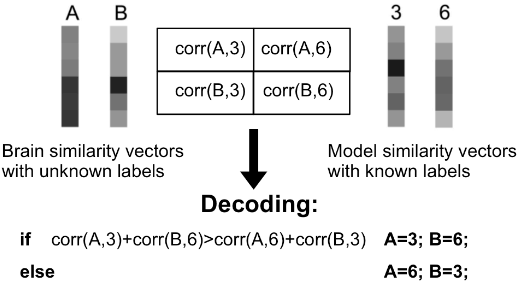
*Slide credit: Andrew Anderson*

# Similarity-based decoding




*Slide credit: Andrew Anderson*

# Similarity-based decoding




*Slide credit: Andrew Anderson*

# Results: Linguistic models

- The models were evaluated in terms of decoding accuracy
- Significance was determined via permutation testing

| | Literal | Metaphor |
|---|---|---|
| OBJECT | 0.51 | **0.67** |
| VERB | **0.71** | 0.54 |
| VERBOBJECT | 0.48 | 0.55 |
| ADDITION | **0.68** | **0.71** |
| LSTM | 0.6 | **0.62** |

## Results: Visual and multimodal models

| | Literal | Metaphor |
|---|---|---|
| VISUAL OBJECT | 0.58 | 0.44 |
| VISUAL VERB | 0.47 | **0.66** |
| VISUAL VERBOBJECT | 0.49 | 0.49 |
| VISUAL ADDITION | 0.47 | **0.68** |
| VISUAL PHRASE | 0.52 | 0.52 |

## Results: Visual and multimodal models

| | Literal | Metaphor |
|---|---|---|
| VISUAL OBJECT | 0.58 | 0.44 |
| VISUAL VERB | 0.47 | **0.66** |
| VISUAL VERBOBJECT | 0.49 | 0.49 |
| VISUAL ADDITION | 0.47 | **0.68** |
| VISUAL PHRASE | 0.52 | 0.52 |

| | Literal | Metaphor |
|---|---|---|
| MULTIMODAL OBJECT | 0.62 | 0.58 |
| MULTIMODAL VERB | 0.52 | **0.67** |
| MULTIMODAL VERBOBJECT | 0.48 | 0.54 |
| MULTIMODAL ADDITION | 0.55 | **0.72** |

## What can we learn from this?

1. The verb embedding is successful in decoding brain activity in the literal case
2. The object embedding and compositional models are more successful in the metaphor case

   *This may suggest that humans pay more attention to the object when interpreting metaphor (speculation)*

3. Visual representations yield significant decoding accuracies in the metaphor case, but not literal

   *This suggests that the visual information plays a role in metaphor processing (speculation)*

# I would like to thank...

- Collaborators: Marek Rei, Luana Bulat, Douwe Kiela, Jean Maillard and Vesna Gamez-Djokic

- Stanford NLP group for the invitation!